2014 IEEE Intelligent Vehicles Symposium (IV)
June 8-11, 2014. Dearborn, Michigan, USA

Robust and Continuous Estimation of Driver Gaze Zone by Dynamic Analysis of Multiple Face Videos

Ashish Tawari and Mohan M. Trivedi

Abstract— **Analysis of driver's head behavior is an integral part of driver monitoring system. Driver's coarse gaze direction or gaze zone is a very important cue in understanding driver-state. Many existing gaze zone estimators are, however, limited to single camera perspectives, which are vulnerable to occlusions of facial features from spatially large head movements away from the frontal pose. Non-frontal glances away from the driving direction, though, are of special interest as interesting events, critical to driver safety, occur during those times. In this paper, we present a distributed camera framework for gaze zone estimation using head pose dynamics to operate robustly and continuously even during large head movements. For experimental evaluations, we collected a dataset from naturalistic on-road driving in urban streets and freeways. A human expert provided the gaze zone ground truth using all vision information including eyes and surround context. Our emphasis is to understand the efficacy of the head pose dynamic information in predicting eye-gaze-based zone ground truth. We conducted several experiments in designing the dynamic features and compared the performance against static head pose based approach. Analyses show that dynamic information significantly improves the results.**

I. INTRODUCTION

Driver distraction and inattention is a leading cause of vehicular accidents [1]. With more and more devices whether as parts of in-vehicle infotainment systems or other nomadic devices such as smartphones, GPS navigation system etc., occupying space in car cockpit, drivers are presented with increasing opportunities to be distracted. Although automakers are very careful in designing in-vehicle systems keeping in mind safety concerns, they have no or very little control over other device usage by the driver. There exist laws banning certain distracting activities such as handheld cellphone usage, texting etc., but their enforcement are often not strict. Moreover, there are other naturally occurring distractions, e.g. checking speedometer at 'wrong' time, parent looking back to his/her child, minds wondering off, or sometimes looking in the 'wrong' direction etc. Hence, simply relying on mandatory laws is not sufficient, but it is important to assist driver with attention technology to avoid unpleasant consequences of distracting activities [2].

Surveys on automotive collisions [3], [4] demonstrated that drivers were less likely (30%-43%) to cause an injury-related collision when they had one or more passengers who could alert him to unseen hazards. Consequently, there is a great potential for human-centric intelligent driver assistance systems [5], [6], [7], [8] to alert or even guide [9] the driver briefly through the critical situation. Monitoring driver behavior is, hence, becoming an increasingly important component of an Intelligent Driver Assistance System (IDAS).

Authors are with the Laboratory of Intelligent and Safe Automobiles (http://cvrr.ucsd.edu/lisa/), University of California, San Diego, La Jolla, CA, USA {atawari, mtrivedi} @ucsd.edu

Driver gaze and head pose are linked to driver's current focus of attention [10]. Therefore, eye or head tracking technology has been extensively used for visual distraction detection. Driving environment, however, presents challenging conditions for a remote eye tracking technology to robustly and accurately estimate eye gaze. Often in research studies participants are asked to not wear glasses or use heavy mascara, or be bearded for reliable eye tracking [2]. Such requirement, however, renders such technology practically unusable for real-world deployment. Robustness requirement of IDAS has suggested the use of head dynamics. Even though precise eye gaze are desirable, coarse gaze direction is often sufficient in many applications. Head pose and dynamics can serve as a good proxy for the coarse gaze direction (hereafter called as gaze zone).

Recent studies predict a driver's intent to turn [11] and intent to change lanes [12] using head motion along with lane position and vehicle dynamics. In fact, head motion analysis provides early cues compared to eye gaze for lane change intent, 3s ahead of the intended event [13]. It is also argued that gaze away from road with detectable head deviation are, with respect to safety concern, more severe than without head deviation [14]. Another important observation made by Martin et al. [15] is that during a typical ride a driver spends 95% of the time facing forward. Then, a system may be able to perform reliably 95% of the time but it is those 5% non-frontal glances are of special interest since interesting events, critical to driver safety, occur during those times. It is when performance of monocular based systems degrades significantly due to decreased visibility of facial features and texture caused by self occlusion.

In this paper, we present head pose based driver gaze zone estimation framework with focus on accurate, robust and continuous performance in real-world driving conditions. We present a novel distributed camera based framework for continuous monitoring of driver head and face to determine the likelihood of a gaze zone. We quantitatively demonstrate the success of the proposed system on the road. For this, we collect naturalistic driving data and evaluated the system performance against zone ground-truth perceived by a human expert utilizing eye data. Furthermore, we show that incorporating dynamic information can improve the system performance significantly.

II. Related Work

Driver inattention study has been actively carried out for decades in multiple disciplines, including cognitive science, psychology, human factor, engineering, with the goal of precisely determining the driver's state of mind. Vision based systems are commonly used for this purpose as they provide non-contact and non-invasive solution. Eye gaze tracking methods have shown good performance under constrained environments. However, such systems are highly susceptible to illumination changes, particularly in real-world driving scenario. Eye-gaze tracking methods using corneal reflection with infrared illumination have been primarily used in an indoor setting [16] but are vulnerable to sunlight.

Due to the robustness requirement, we focus on head tracking based systems which have potentials to accurately and reliably determine gaze zone. We present select relevant head pose and gaze zone related studies. For a good overview of vision based head pose estimation systems, readers are encouraged to refer to a survey by Murphy-Chutorian and Trivedi [17], and for driver inattention monitoring system, a review by Dong et. el. [18].

To determine head orientation, Ji and Yang [19] used eigenspace algorithm to map pupil related feature to head pose space and further quantized the orientation into seven angles between -45^0 to $+45^0$. The pupil itself is obtained from bright and black pupil image acquired using specialized hardware setup. Gaze direction is then estimated using information about the head movement and relative position between pupil and glint, with the gaze direction quantized into nine zones. The work showed promising results under different lighting conditions in laboratory settings. Its evaluation on-road driving data is not available.

Facial feature based head pose estimation methods analyze geometric configuration of the features along with face model (e.g. cylindrical [20], ellipsoidal [21] or mean 3D face [15]) to recover head pose. Smith et al. analyzed global motion, and color and intensity statistics to track head and facial features such as eyes, lip corners, and the bounding box of the face [22]. From these tracked features, they estimated continuous head orientation and gaze direction. However, the method tracked these features separately and cannot always find them e.g. during partial occlusion due to hand, eye wears or during conversation. Kaminski et al. [23] detected pupils, pupil glints and nose bottom by analyzing the intensity, shape, and size properties. Using these properties, and a geometric model of human face and eye, continuous head orientations and gaze direction are estimated. However, the accuracy of the eye location significantly drops in the presence of large head movements, causing degradation in the performance for deviation from the frontal pose.

To address above concerns, we have used distributed camera setup as proposed in [24] to continuously and accurately track head pose over wide operating range. Unlike stereo setup, the system does not require large overlap between two cameras. We further divide the driver's filed of view to the zones related to driving tasks. We present a thorough analysis of the efficacy of head pose and dynamic based features to estimate gaze zone.

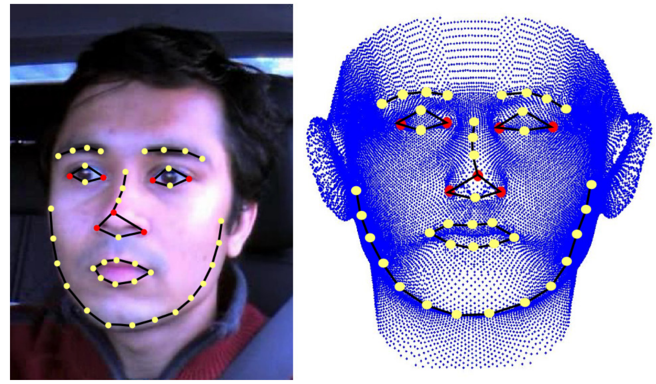

Fig. 1. Tracked facial feature/landmarks and their correspondence in 3D face image. Solid red circles are the points utilized for the head pose calculation.

III. Proposed Gaze Zone Estimation Framework

Improved operating range of the head pose tracking system is essential for continuous and accurate monitoring of driver's head movement even during large deviation from the frontal pose. For this, we utilize a distributed camera framework inside the car cockpit. Each camera perspective is processed independently in the framework, and hence does not require large overlap between two cameras. A perspective selection procedure at later stage provides the final head pose estimation. We present a brief description of the head pose estimation method and refer readers to [24] for algorithmic details including camera selection and 'optimal' camera position determination. We provide details about the features extracted from the head pose and its dynamic, which is used by a classifier to determine the zone-membership (a probability that the driver's gaze belongs to a given gaze zone).

Head pose estimation methods based on geometric approach utilizing facial landmark and its 3-D correspondences, are fast, simple and can provide a good estimate [17]. Facial feature based method has an advantage of providing tracked eye location which can be used for eye analysis too. The challenge, however, lies in detecting the landmarks accurately and reliably. With recent advancement in robust facial landmark tracking [25], [26], even during partial occlusion and presence of glasses, geometric approaches have shown great potentials [24]. In our system, we track landmarks like eye corners, nose corners and nose tip as shown in Figure 1. From the tracked landmarks and their relative 3D configurations, a weak perspective projection model, POS (Pose From Orthography and Scaling) [27] determines the rotation matrix and corresponding yaw, pitch and roll angles of the head pose.

A. Head Pose Feature Extraction

Our goal is to predict coarse eye gaze direction using head pose cues. Literature suggest that any visual search,

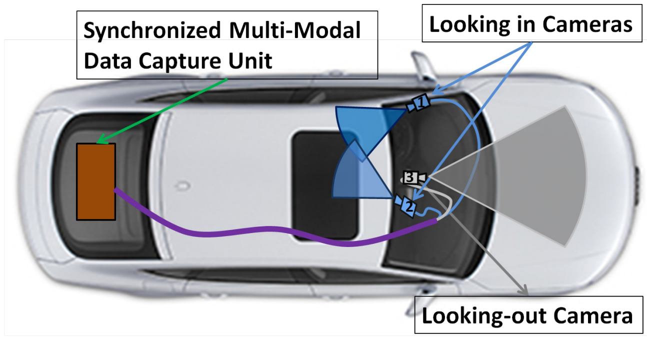

Fig. 2. LISA-A experimental testbed equipped with and capable of time synchronized capture of looking-in and looking-out cameras.

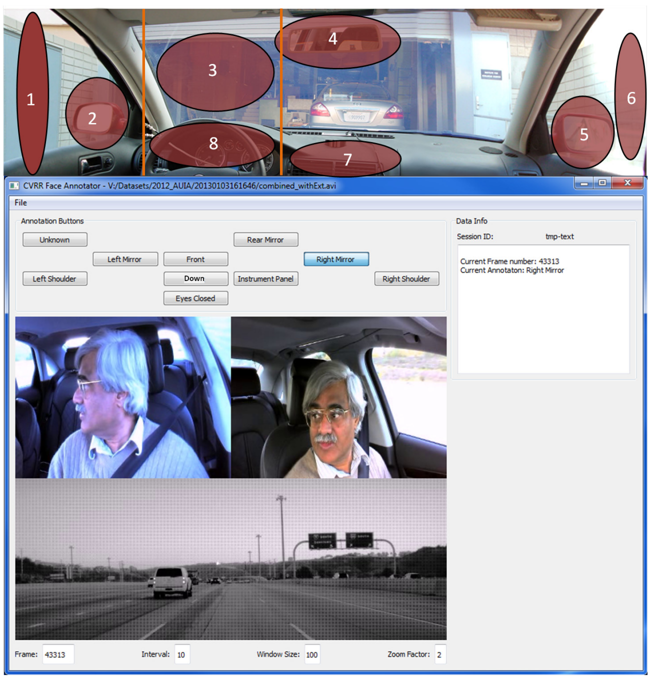

Fig. 3. Annotation tool box providing all the contextual information for ground truth generation. Top portion shows the approximate regions in vehicle frame used for gaze annotation.

for the most part, is presumed to be guided by some combination of a goal- (top-down) and stimulus-driven (bottom-up) approach, depending on the situation. In both of these approaches, head and eye gaze movements interact to each other in a coordinated manner to 'optimally' guide through the search space [28]. Such coordination between head and eye dynamics should be exploited in order to better predict the gaze zone using head pose cues.

It's our hypothesis that how one arrives to a particular head pose, can predict how eye gaze might have moved. For this investigation, we examine the role of both, static head pose- and head pose dynamic features, for gaze zone estimation.

1) Static features: Current head pose angles, i.e. yaw, pitch and roll, form a part of the static features. The raw signal, however, has high frequency component causing tiny fluctuations. Hence, we pass it through a low pass filter, a simple moving average as in Eq. 1.

$$y[n] = \frac{1}{L}\sum_{i=0}^{L-1} x[n-i] \qquad (1)$$

where, L is the filter length. Beside the current head pose, we also calculate center of the face, as centroid of the two eyes and mouth region. This is an informative cue specially in the constraint sitting position of the driver. Thus, total length of the static features is $3 + 2 = 5$.

2) Dynamic features: To capture the time trend, we extract several statistics of the time series. First, we smooth the raw data as in case of the static features. Then, from windowed time series (through W sec prior to the current time t), we extract: minimum value, position of the minimum, maximum value, position of the maximum, mean angle, mean angular velocity. The statistics are taken in all the three pose angles. Thus, total length of the dynamic features is $3 * 6 = 18$.

IV. Experimental Analysis and Evaluation

A. Testbed and Zone Dataset Description

Data is collected from naturalistic, on-road driving using the LISA-A testbed as shown in Figure 2. Two cameras are mounted facing the driver: one on the A-pillar and another near the rear view mirror. They capture face view in color video stream at 30fps and 640×360 pixel resolution. From the collected data, we mined sections where driver is making different maneuvers including right/left turns, right/left lane changes, stops at stop signs, and freeway merges for ground truth generation.

An annotation toolbox, as seen in Figure 3, is developed to assist a human expert in efficiently labeling the video frames. The expert is presented with both the driver looking cameras as well as the outside looking camera to provide full contextual information, and is asked to utilize all the information including eye data to annotate the 'best' perceived gaze zone region. The zone considered for annotations are shown in the Figure 3 with approximate zone boundaries. Total of over 15000 frames are annotated with multiple drivers, with eye-glass and no eye-glass.

B. On-Road Evaluations

For zone estimation, we use random forest classifier in conjunction with proposed feature set. Random forest has shown promising results in many machine learning applications. We also choose this classifier because of the ease of the ability to interpret the learned parameters, and low number of tuning parameters. We used the random forest library available in Matlab. The only parameters that we tune is the number of trees, as shown in Figure 4. In all our experiments, we fix the number of trees as 60.

Next, we present series of experiments to compare different time windows for dynamic feature extraction, static alone and dynamic included, classification performance, and different number of zones. For each experiment, we provide performance results for randomized tenfold cross validation,

TABLE I

STATIC FEATURE: CONFUSION MATRIX FOR 8 GAZE ZONES NUMBERED AS SHOWN IN FIGURE 3.

True Gaze Zone	Recognized Gaze Zone							
	1	2	3	4	5	6	7	8
1	85.9	13.2	0.9	0.0	0.0	0.0	0.0	0.0
2	1.1	87.5	11.2	0.1	0.0	0.0	0.0	0.1
3	0.4	5.3	88.3	1.9	1.8	0.3	0.6	1.3
4	0.0	0.0	12.4	80.6	2.8	0.0	4.1	0.0
5	0.0	0.0	2.6	6.4	87.1	1.9	2.0	0.0
6	0.0	0.0	0.0	1.1	25.8	73.0	0.0	0.0
7	0.0	0.0	9.3	5.8	1.3	0.0	83.3	0.3
8	0.0	0.8	42.4	4.0	0.0	0.0	3.2	49.6

Unweighted Accuracy = 79.4%
Weighted Accuracy = 85.7%

TABLE II

DYNAMIC FEATURE: CONFUSION MATRIX FOR 8 GAZE ZONES NUMBERED AS SHOWN IN FIGURE 3.

True Gaze Zone	Recognized Gaze Zone							
	1	2	3	4	5	6	7	8
1	94.4	5.2	0.5	0.0	0.0	0.0	0.0	0.0
2	0.8	95.7	3.3	0.1	0.0	0.0	0.0	0.2
3	0.2	2.3	93.9	1.1	0.5	0.0	0.5	1.4
4	0.0	0.0	5.7	90.9	1.1	0.0	2.3	0.0
5	0.0	0.0	1.4	1.9	96.1	0.3	0.3	0.0
6	0.0	0.0	0.0	2.2	1.1	96.6	0.0	0.0
7	0.0	0.0	6.1	3.3	0.3	0.0	90.4	0.0
8	0.0	0.8	45.6	2.4	0.0	0.0	3.2	48.0

Unweighted Accuracy = 88.2%
Weighted Accuracy = 93%

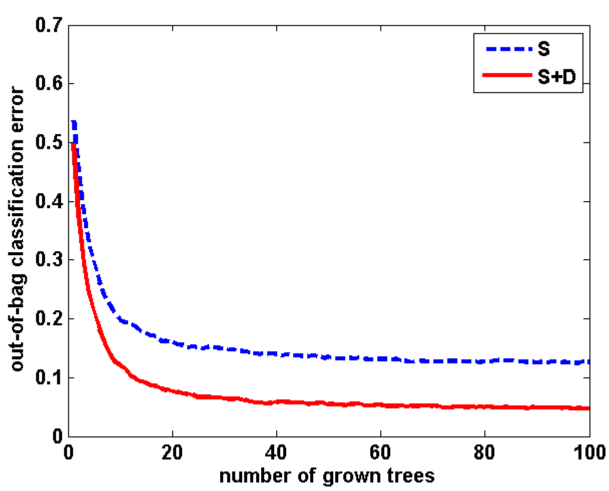

Fig. 4. Out-of-bag classification error as a function of number of trees.

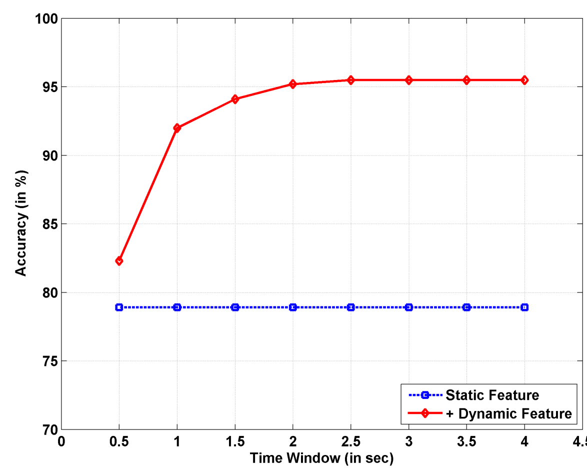

Fig. 5. Gaze estimation accuracy Vs time window (W sec)for dynamic feature extraction. After 2 sec time window the improvement in accuracy is negligible.

that is, the database is divided into ten folds in stratified manner so that they contain approximately the same proportions of labels as the original database, and the system is trained on nine folds and tested on the left out fold. This is repeated ten times, each time leaving out a different fold. We report weighted- (overall), unweighted- (per class) accuracy and confusion matrix for performance evaluation.

First, we examine the effects of time window size chosen for the dynamic features on the performance of the classifier. Figure 5 shows that as time window increases the accuracy improves. Window size greater than 2 seconds does not provide any further improvement. Hence, we chose and fixed this window size for next set of experiments.

TABLE III

STATIC FEATURE: CONFUSION MATRIX FOR 7 GAZE ZONES (FRONT AND DOWN COMBINED) NUMBERED AS SHOWN IN FIGURE 3.

True Gaze Zone	Recognized Gaze Zone						
	1	2	3 + 8	4	5	6	7
1	85.9	13.2	0.9	0.0	0.0	0.0	0.0
2	1.1	87.5	11.3	0.1	0.0	0.0	0.0
3+8	0.4	5.1	89.7	2.0	1.7	0.3	0.8
4	0.0	0.0	12.4	80.6	2.8	0.0	4.1
5	0.0	0.0	2.6	6.4	87.1	1.9	2.0
6	0.0	0.0	0.0	1.1	25.8	73.0	0.0
7	0.0	0.0	9.6	5.8	1.3	0.0	83.3

Unweighted Accuracy = 83.9%
Weighted Accuracy = 87.0%

TABLE IV

STATIC FEATURE: CONFUSION MATRIX FOR 7 GAZE ZONES (FRONT AND DOWN COMBINED) NUMBERED AS SHOWN IN FIGURE 3.

True Gaze Zone	Recognized Gaze Zone						
	1	2	3 + 8	4	5	6	7
1	94.4	5.2	0.5	0.0	0.0	0.0	0.0
2	0.8	95.7	3.5	0.1	0.0	0.0	0.0
3+8	0.2	2.3	95.2	1.2	0.5	0.0	0.6
4	0.0	0.0	5.7	90.9	1.1	0.0	2.3
5	0.0	0.0	1.4	1.9	96.1	0.3	0.3
6	0.0	0.0	0.0	2.2	1.1	96.6	0.0
7	0.0	0.0	6.1	3.3	0.3	0.0	90.4

Unweighted Accuracy = 94.2%
Weighted Accuracy = 94.7%

We compare the classification performance using static-alone and static-cum-dynamic features. Table I and Table II show the confusion matrices and the classification accuracies for static and static-cum-dynamic features respectively. It can be seen that dynamic features significantly improve the classification accuracies. A closer look to the confusion matrices suggests that most confusions exist in adjacent gaze zones, and front (zone #3) and down zones (zone #8) are the most confused ones. We run another experiment with down and front zone merged. Table III and Table IV shows similar trend for the static and the static-cum-dynamic features as earlier. The best unweighted accuracy (per class), however, is jumped to over 94% for 7 gaze zone regions. Figure 6 shows working of the overall system with probabilistic output for the three zones (Right, Front and Left) prediction, using dynamic features. Notice that static head pose (bottom right), without eye information, can be misleading, i.e. head pose alone (without dynamic information) would suggest non-frontal gaze zone.

Finally, to investigate the importance of the different dynamic features, we perform permutation test. The idea is that if the variable is not important, then randomly permuting the values of variable in the out-of-bag cases will not degrade prediction accuracy. Our findings are, first, the dynamic information of the yaw angle is found the most important and that of roll angle the least important. This is expected since the yaw angle shows the largest variability for the different gaze zones. Second, among the statistics, we find the position of the extrema (min and max) and mean angular velocity of yaw and pitch are more important than other statistics. This suggests that the direction of arrival to a particular pose can better indicate where one might be looking.

V. CONCLUDING REMARKS

Driver's coarse gaze direction can provide a lot of useful information for the driver behavior studies. In this paper, we presented a head pose and its dynamic based gaze zone estimation system. We performed our evaluation in naturalistic driving conditions. The gaze zone ground truth is obtained by a human experts utilizing all available vision information including eye data and surround context. We showed that dynamic features can significantly improve the system performance. However, certain classes are inherently confusing such as Front and Down gaze zones, since subtle eye movement is sufficient to guide the eye gaze between them. In future, we will integrate eye information, when available, to further disambiguate confusion between adjacent gaze zones. We would also like to incorporate other body parts analysis such as hand [29] directly with gaze zones for the driver activity analysis.

ACKNOWLEDGMENT

The authors would like to thank UC Discovery and industry partners especially Audi and VW Electronic Research Laboratory for their support and collaboration. We also thank LISA colleagues for their valuable assistance.

REFERENCES

[1] S. G. Klauer, F. Guo, J. Sudweeks, and T. A. Dingus, "An analysis of driver inattention using a case-crossover approach on 100-car data: Final report," *National Highway Traffic Safety Administration*, 2010.

[2] C. Ahlstrom, K. Kircher, and A. Kircher, "A gaze-based driver distraction warning system and its effect on visual behavior," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 965–973, June 2013.

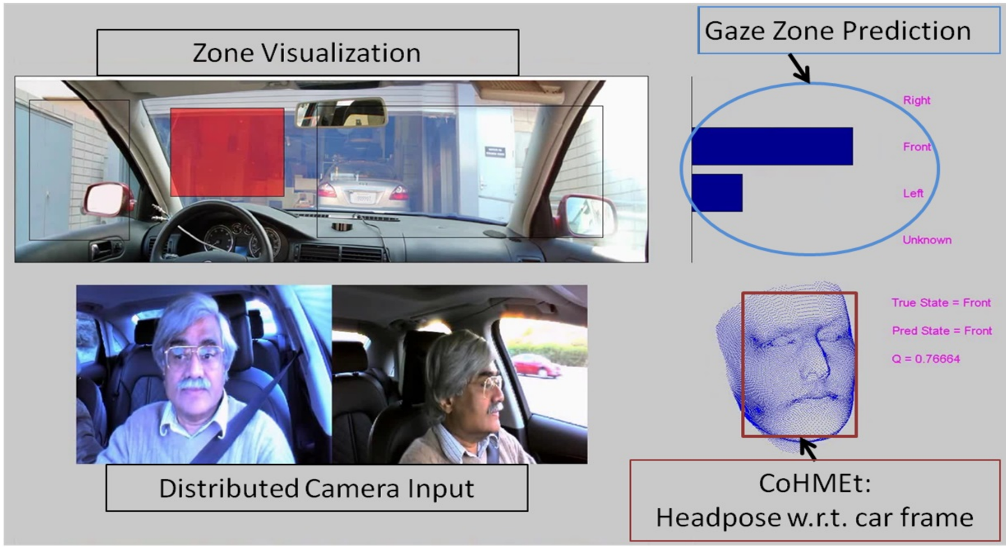

Fig. 6. Visualization of output by CoHMEt and Zone Estimation: Bottom left: input video feeds with multi-camera perspective; bottom right: Tracked head pose visualization; top left: 3 zones (left, front and right) visualization; top right: zone membership. Notice, static head pose alone (without eye and dynamic information) can be misleading in determining gaze zone.

[3] T. Rueda-Domingo, P. Lardelli-Claret, J. d. D. Luna-del Castillo, J. J. Jiménez-Moleón, M. Garcıa-Martın, and A. Bueno-Cavanillas, "The influence of passengers on the risk of the driver causing a car collision in spain: Analysis of collisions from 1990 to 1999," *Accident Analysis & Prevention*, vol. 36, no. 3, pp. 481–489, 2004.

[4] K. A. Braitman, N. K. Chaudhary, and A. T. McCartt, "Effect of passenger presence on older drivers risk of fatal crash involvement," *Traffic Injury Prevention*, vol. 15, no. 5, pp. 451–456, 2013.

[5] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 1, pp. 108–120, 2007.

[6] M. M. Trivedi and S. Y. Cheng, "Holistic sensing and active displays for intelligent driver support systems," *Computer*, vol. 40, no. 5, pp. 60–68, 2007.

[7] K. S. Huang, M. M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," in *IEEE Intelligent Vehicles Symposium*, 2003, pp. 444–449.

[8] A. Doshi, S. Y. Cheng, and M. M. Trivedi, "A novel active heads-up display for driver assistance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 1, pp. 85–93, 2009.

[9] A. Tawari, S. Sivaraman, M. M. Trivedi, T. Shannon, and M. Tippelhofer, "Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking," in *IEEE Intelligent Vehicles Symposium*, June 2014.

[10] E. Murphy-Chutorian and M. Trivedi, "Hyhope: Hybrid head orientation and position estimation for vision-based driver head tracking," in *Intelligent Vehicles Symposium, 2008 IEEE*, june 2008, pp. 512 –517.

[11] S. Y. Cheng and M. M. Trivedi, "Turn-intent analysis using body pose for intelligent driver assistance," *IEEE Pervasive Computing*, vol. 5, no. 4, pp. 28–37, 2006.

[12] A. Doshi, B. Morris, and M. M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues," *IEEE Pervasive Computing*, vol. 10, no. 3, pp. 22–34, 2011.

[13] 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*, vol. 10, no. 3, pp. 453–462, 2009.

[14] H. Zhang, M. Smith, and R. Dufour. A final report of safety vehicles using adaptive interface technology: Visual distraction. [Online]. Available: http://www.volpe.dot.gov/coi/hfrsa/work/roadway/saveit/docs.html

[15] S. Martin, A. Tawari, E.-M. Chutorian, S. Y. Cheng, and M. M. Trivedi, "On the design and evaluation of robust head pose for visual user interfaces: Algorithms, databases, and comparisons," in *4th ACM SIGCHI International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AUTO-UI)*, 2012.

[16] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections." *IEEE Trans. Biomed. Engineering*, vol. 53, no. 6, pp. 1124–1133, 2006.

[17] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607 –626, April 2009.

[18] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 596–614, June 2011.

[19] 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, 2002.

[20] R. Valenti, N. Sebe, and T. Gevers, "Combining head pose and eye location information for gaze estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 802–815, 2012.

[21] 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," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 1, pp. 254–267, 2011.

[22] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 4, no. 4, pp. 205–218, Dec 2003.

[23] J. Y. Kaminski, D. Knaan, and A. Shavit, "Single image face orientation and gaze detection," *Machine Vision and Application*, vol. 21, no. 1, pp. 85–98, 2009.

[24] A. Tawari, S. Martin, and M. M. Trivedi, "Continuous head movement estimator (cohmet) for driver assistance: Issues, algorithms and on-road evaluations," *IEEE Transactions on Intelligent Transportation Systems*, 2014.

[25] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2879–2886.

[26] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013, pp. 532–539.

[27] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *International journal of computer vision*, vol. 15, no. 1-2, pp. 123–141, 1995.

[28] A. Doshi and M. M. Trivedi, "Head and eye gaze dynamics during visual attention shifts in complex environments," *Journal of Vision*, vol. 12(2):9, 2012.

[29] S. Martin, E. Ohn-Bar, A. Tawari, and M. M. Trivedi, "Understanding head and hand activities and coordination in naturalistic driving videos," in *IEEE Intelligent Vehicles Symposium*, June 2014.