Where is the driver looking: Analysis of Head, Eye and Iris for Robust Gaze Zone Estimation

Ashish Tawari, Kuo Hao Chen and Mohan M. Trivedi

***Abstract*— Driver's gaze direction is a critical information in understanding driver state. In this paper, we present a distributed camera framework to estimate driver's coarse gaze direction using both head and eye cues. Coarse gaze direction is often sufficient in a number of applications, however, the challenge is to estimate gaze direction robustly in naturalistic real-world driving. Towards this end, we propose gaze-surrogate features estimated from eye region via eyelid and iris analysis. We present a novel iris detection computational framework. We are able to extract proposed features robustly and determine driver's gaze zone effectively. We evaluated the proposed system on a dataset, collected from naturalistic on-road driving in urban streets and freeways. A human expert annotated driver's gaze zone ground truth using information from the driver's eyes and the surrounding context. We conducted two experiments to compare the performance of the gaze zone estimation with and without eye cues. The head-alone experiment has a reasonably good result for most of the gaze zones with an overall 79.8% of weighted accuracy. By adding eye cues, the experimental result shows that the overall weighted accuracy is boosted to 94.9%, and all the individual gaze zones have a better true detection rate especially between the adjacent zones. Therefore, our experimental evaluations show efficacy of the proposed features and very promising results for robust gaze zone estimation.**

I. INTRODUCTION

Driver distraction and inattention are one of the major factors that lead to vehicular accidents [1]. In 2012 alone, there were 3,328 fatalities and 421,000 reported injuries in crashes involving distracted driver [2]. Cellphone or other electronic devices usage including in-vehicle infotainment are some of the most common causes of inattention. Distraction can also be caused by momentary lapse of attention e.g. looking into 'wrong direction', checking speedometer at 'wrong' time, etc. There exists an opportunity for an Advanced Driver Assistance System (ADAS) with ability to monitor driver's focus of attention to provide alerts or even guide through dangerous situations and improve road safety.

Driver gaze and head pose are linked to driver's current focus of attention [3]. Therefore, eye and head tracking technologies have been extensively used for visual distraction detection. Precise eye-gaze tracking, however, is still a very challenging task. Eye-gaze tracking methods using corneal reflection with infrared illumination have been primarily used in an indoor setting [4] but are vulnerable to sunlight. Besides technical difficulties, often times, these systems require specialized hardware setup with lengthy calibration procedure. With constant jolts and shakes during driving, maintenance of such systems can be prohibitively expensive.

Authors are with Laboratory for Intelligent and Safe Automobiles, University of California, San Diego, USA atawari@ucsd.edu, khc030@eng.ucsd.edu mtrivedi@ucsd.edu

On the other hand, coarse gaze direction is often sufficient for higher level semantic analysis in a number of applications [5], [6], [7], [8]. Coarse gaze direction is usually approximated by head pose [9]. A distraction detection study [5] suggests that even though precise eye gaze is a better predictor, head pose still provides a good estimate of the driver distraction. Tawari and Trivedi [10] showed that head pose dynamics provide better cues than static head pose alone to estimate coarse gaze directions or gaze zones such as front, left/right side mirrors, speedometer, center console, etc. The system, however, fails to distinguish between adjacent zones separated by subtle eye movement e.g. between front and speedometer.

In this paper, we present a framework to estimate driver's gaze zone using both head and eye cues. Our emphasis is on simple hardware setup and robust performance. Towards this end, we propose a distributed monocular camera based system and a *robust* eye feature estimation algorithm. Note that, we are not interested in precise gaze, but in coarse gaze direction, hence we compute eye features which may not provide precise gaze information, but can be estimated robustly and, hopefully, are informative of the driver's gaze zones. We quantitatively demonstrate the success of the proposed system on the road. For this, we collect naturalistic driving data and evaluate the system performance against zone ground-truth as perceived by a human expert using all vision information including eyes and surround context.

II. RELATED STUDIES

Driver distraction and inattention detection require monitoring driver's head and eyes. Vision based systems are commonly used for this purpose as they provide non-contact and non-invasive solution. In recent years, a number of efforts [10], [11] have been made towards robust and continuous monitoring of the driver. Such systems are often focused on head pose alone due to robustness requirement. It is argued that the drivers tend to use both head and eye combination to direct their gaze to the target [5]. It is shown [10] that incorporating dynamic information of the head can provide better zone estimation accuracy. Perhaps, because head and eye co-ordination during driving [12] is exploited by head pose dynamics, revealing eye movement information indirectly. However, ignoring eye related cues, when available, inherently limits the system performance.

Remote eye tracking, on the other hand, is still a very challenging task, especially in outdoor setting like driving. Many of the existing methods use active IR illumination technique to obtain bright and dark pupil images which

Fig. 1: The examples of non-ellipse iris shape.

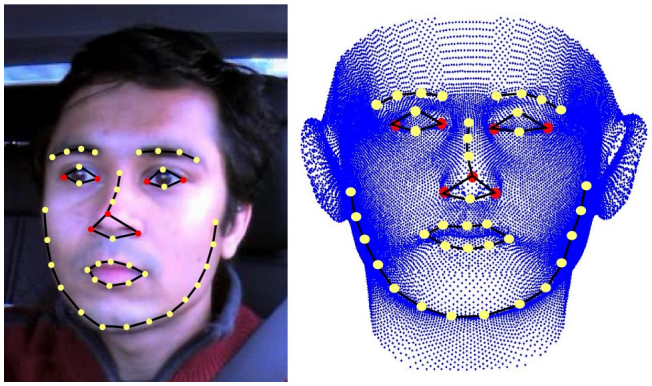

Fig. 2: Tracked facial feature/landmarks and their correspondence in 3D face model.

are further processed to detect and track iris and pupil. The success of such a technique strongly depends on the brightness and size of the pupils which are affected by several factors, including eye closure, eye occlusion due to the face rotation, external illumination interference, the distance of the subject to the camera, and the intrinsic properties of the eyes (i.e., the bright pupil reflection tends to be darker for older people) [13], [14]. Due to these reasons, such systems tend to perform well in indoor setting and are susceptible to outdoor lighting conditions.

Passive light approaches relying on visible light are potentially better suited for outdoors. Shape- and appearance-based are two prominent approaches for eye detection. Shape based methods are based on the premise that the open eye is well described by its shape, which includes the iris and pupil contours and the exterior shape of the eye (eyelids). Simple ellipse [15] to more complex models e.g. incorporating eyelids shape parameters [16] have been proposed to model the eye shape. However, in naturalistic setting the shape of the eye can change significantly due to free head movement, facial expression, eye state (open/closing) etc., see Figure 1. Appearance models or image-template based models [17] detect and track eyes directly based on the photometric appearance as characterized by the color/intensity statistics of the eye and its surroundings. These methods usually demand large amount of training data with eye images of different subjects, under different orientations and illumination for reasonable performance. The existing methods, both shape- and appearance-based, are to a large extent applicable to near frontal view angles, fully open eyes, and under relatively constrained light conditions. For further details on different approaches of eye tracking and gaze estimation, we encourage readers to refer to a survey by Hansen and Ji [18].

Recent times, local patch based approaches have shown promising results for object detection, recognition and categorization. The local-feature based methods have also been applied to the detection of facial landmarks such as eye, nose and mouth corners [19], [20], [21], [22]. We formulate iris tracking problem in this framework (Section III-A) without geometric shape-model and potentially provide better accuracy. Since features are calculated from local neighborhood, it provides more robust performance to face pose and illumination changes than the holistic appearance based eye detection approaches. Furthermore, from tracked eye landmarks and iris center, we propose 'robust' gaze-surrogate (Section III-B) measurements and show its efficacy for the driver's zone estimation. Our analyses show promising results on the naturalistic driving dataset (Section IV-A). Also important to note is that the models for head pose through facial landmarks and iris tracking are learnt using separate datasets [23], [24]. This shows ability of the framework to generalize to different subjects.

III. HEIDy Cue Extraction

To infer driver's gaze zone, we extract cues related to Head, Eye and Iris Dynamics (HEIDy). First, we automatically detect facial landmarks - eye corners and contour, mouth corners, nose corners and nose tip as shown in the Figure 2. From the corner points, we estimate head pose utilizing a generic 3-D face model using Pose from Orthography and Scaling (POS) algorithm [25]. For more details on multi-camera head pose estimation including facial landmark detection algorithms, camera placement and selection procedure, we encourage readers to refer to CoHMEt framework by Tawari et. al. [11]. Here, we discuss local patch-based regression framework for iris tracking. We detail the training procedure and the iterative refinement steps to detect the iris centers jointly for both the eyes. Next, we discuss gaze-surrogate features estimated using eye landmarks and the iris centers. Finally, a gaze zone classifier trained on the proposed features is discussed in the next section.

A. Iris Center Detection and Tracking

The iris centers are detected by training a sequence of regression matrix $S = (R_1, ..., R_K)$. This sequence of regression matrix S together with an initial estimate of the positions p_0 with respect to Image I can be used in Algorithm 1 to obtain the positions of iris centers p_K after desired K iterations. At each iteration, the estimated positions of iris centers p_k are computed by:

$$p_k = p_{k-1} + R_{k-1} feat(p_{k-1}, I) \quad (1)$$

where $feat$ gives a feature vector (e.g., HoG descriptor [26] in our case) extracted at p_{k-1} in image I. This can be viewed as coarse to fine adjustment of the center locations as the successive model attempts to reduce the residual errors.

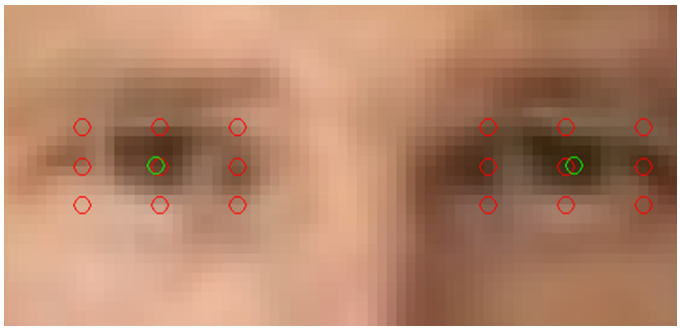

Fig. 3: The 10 initial estimates of iris centers.

In Algorithm 2, the pseudo code describes the training procedure for obtaining a sequence of regression matrix $S = (R_1, ..., R_K)$. Assume that we are given a set of images $\{I^i\}$, their corresponding ground-truth iris centers $\{p_*^i\}$, and their corresponding initial estimate of iris centers $\{p_0^i\}$. When training a regression matrix R_k at iteration k, we want to minimize the distance between the true iris center positions p_*^i and the positions p_k^i in every image. p_k^i represents the positions at iteration k starting from the initial estimate of iris center positions p_0^i. The loss function at each iteration can be written as in Eq.2, and our goal is to minimize this loss function.

$$L = \sum_i \sum_{p_k^i} \|p_*^i - p_k^i - R_k feat(p_k^i, I^i)\|^2 \qquad (2)$$

The regression matrix R_k which minimizes the loss function can be obtained by solving the linear least squares problem shown in Eq.3.

$$R_k = \arg\min_{R_k} \quad L \qquad (3)$$

Once the regression matrix R_k is found, it is plugged into Eq.1 to compute the new p_k^i which can be used in the next iteration. By repeating the same process for K iterations, we obtain the sequence of regression matrix $S = (R_1, ..., R_K)$.

Before training an iris center detection model (a sequence of regression matrix), there are important image processing steps needed to be done to ensure training correctness and quality. They are scaling, rotation, and histogram equalization. For scaling, an image is scaled to a size such that the distance between both eye centers equals to a fixed length. For rotation, an image is rotated in a way that the line going through both eye centers is parallel to the x-axis. Scaling and rotation ensure the size and the orientation of eyes to be approximately the same for every image. Also, histogram equalization is applied only in the eye region of an image to enhance image contrast for a better low-level feature extraction.

Once images are registered through the above steps, we train each regression matrix R_k. For $k = 1$, we need initial guess for the iris centers. It's important to provide enough perturbation to the system to learn 'good' model. We assign 10 initial estimates of the iris center to each eye in an image during training. The 10 initial positions are illustrated in Figure 3. They consist of eye centers (estimated from the tracked eye corners) and eight positions around the centers, and a ground-truth position. Inclusion of ground truth is to ensure regressor, R_1, does not deviate from actual position at the first step.

In order to improve iris center detection rate in videos, iris tracking is implemented. The tracking is done by using the detected positions from previous frame as the initial estimates of current frame. Since there is usually not much movement between two consecutive frames, it's a very effective way for tracking and increasing iris center detection rate.

Algorithm 1 Testing

Input: Image I, initial estimates p_0 and regression matrices $S = (R_1, ..., R_K)$

1: **for** $k = 1$ `to` K **do**
2: $\quad h = feat(p_{k-1}, I)$
3: $\quad p_k = p_{k-1} + R_{k-1}h$
4: **end for**

Output: p_K

Algorithm 2 Training

Input: Number of iterations K, data(I^i, p_0^i, p_*^i for $i = 1...N$)

1: **for** $k = 1$ `to` K **do**
2: $\quad h^i = feat(p_{k-1}^i, I^i)$
3: $\quad L = \sum_i \sum_{p_{k-1}^i} \|p_*^i - p_{k-1}^i - R_{k-1}h^i\|^2$
4: $\quad R_{k-1} = \arg\min_{R_{k-1}} L$
5: $\quad p_k^i = p_{k-1}^i + R_{k-1}h^i$
6: **end for**

Output: $S = (R_1, ..., R_K)$

B. Gaze-Surrogate Estimation

Given eye landmarks and detected iris centers, we compute two gaze-surrogate measurements: one to account for horizontal gaze movement and another for vertical gaze movement.

Vertical gaze movement with respect to head is inferred from the eye area. The eye area is computed using only upper-eyelid contour. This is inspired from physiology of normal eye, where upper eyelid naturally moves in the same direction as the vertical-gaze. Lower eyelid, however, does not provide any gaze relevant information. In fact, it may introduce noise e.g. upward movement of lower eyelid due to squinting or a facial expression. We believe this subtle difference is important to avoid unnecessary noise and hence improve the gaze estimate.

Horizontal gaze-direction is explained in our contemporary work by Tawari et al. [27]. We describe mathematical details here again for the sake of completion and the reader's convenience. The horizontal gaze-direction β with respect to head, see Figure 4, is estimated as a function of α, angle subtended by an eye in horizontal direction, head-pose (yaw)

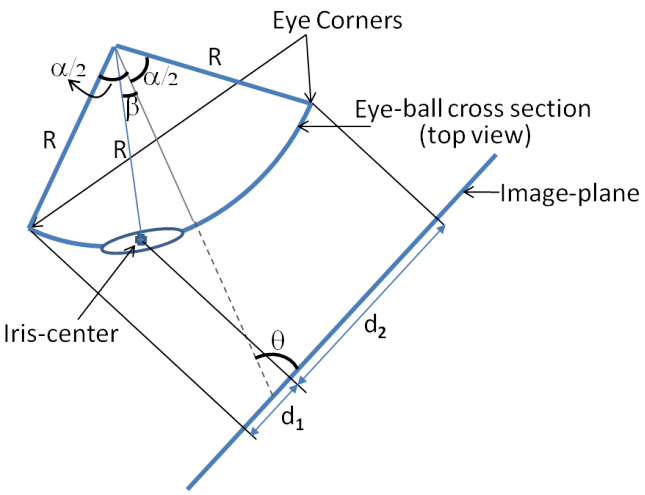

Fig. 4: Eye ball image formulation: estimating β, gaze-angle with respect to head, from α, θ, d_1 and d_2

angle θ, and the ratio of the distances of iris center from the detected corner of the eyes in the image plane. Equation 4-5 show the calculation steps.

$$\frac{d_1}{d_2} = \frac{\cos(\theta - \alpha/2) - \cos(\theta - \beta)}{\cos(\theta - \beta) + \cos(180 - \theta - \alpha/2)} \quad (4)$$

$$\beta = \theta - \arccos\left(\frac{2}{d_1/d_2 + 1}\sin(\theta)\sin(\alpha/2) + \cos(\alpha/2 + \theta)\right) \quad (5)$$

Since the raw eye-tracking data is noisy (due to blinking and tracking errors), we smooth angle β with a median filter.

IV. Experimental Evaluation

A. Datasets and Annotation

We trained the iris center detection models using the images from the LFW dataset [24]. LFW is a face recognition database which contains more than 13,000 images of faces collected from the web, and 1,680 of people have two or more distinct images in the dataset. We randomly picked a total of 330 images from the dataset and manually annotated the iris centers for all the images. 270 of the images are used for training, and the rest of the 60 images are used as testing images.

The iris center detection models are also applied to our naturalistic driving dataset to evaluate their efficacy for gaze zone estimation in real-world data. The driving data is collected from naturalistic, on-road driving using our LISA-A testbed [28]. There are two looking-in cameras facing the driver for capturing the driver's movements, and one looking-out camera for capturing the front road condition. The two looking-in cameras capture face view in color video stream at 30 fps and 640x360 pixel resolution.

For gaze zone estimation, an annotation effort is done on the naturalistic driving data. This annotation effort focuses

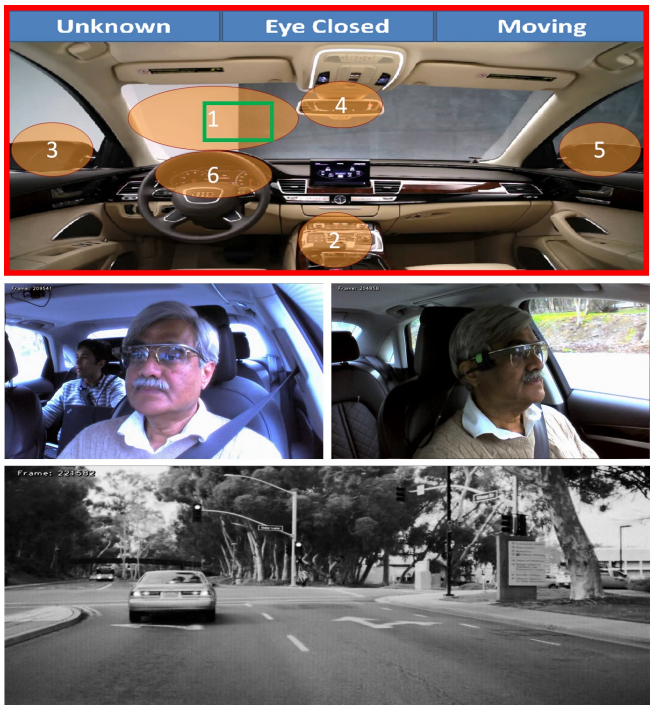

Fig. 5: Graphical user interface for annotating gaze zone.

on annotating the driver's gaze zone at a given moment. An annotation toolbox is developed to assist a human expert to efficiently label the video frames. The annotation toolbox displays a vehicle interior front-view image, two looking-in camera views, and one looking-out camera view (see Figure 5). The looking-in and looking-out camera views are for assisting a human expert to determine where the driver is looking, and the vehicle interior image is an annotating plane to perform labeling. A human expert would look at the three available camera views to determine which zone the driver is looking at and draw a box within the zone on the vehicle interior image, such as the green box seen in Figure 5. We collected about 2,000 annotations from this naturalistic driving data, and these annotations are used in the gaze zone experiments.

B. Results: Iris Tracking

For determining a good patch size to extract features around each position p_k, a total of six different iris center detection models are trained for six different patch sizes, which are 8, 16, 24, 32, 40, and 64. Figure 6 presents normalized error in training set as a function of iterations. The normalized error of the models at each iteration was calculated by computing the difference between the detected iris center positions p_K and the annotated ground-truth positions p_*. Patch size 40 produced the best results. No significant improvement beyond iteration 2 is observed, hence we choose $K = 2$. Next, we evaluated the model performance in a testing set of 60 images. The average normalized error obtained on testing set is 0.0532.

After verifying the iris center detection model on the LFW dataset, the next step of the evaluation is in applying the

TABLE I: Head pose alone: Confusion matrix for 6 gaze zones numbered as shown in Figure 5. Notice that speedometer (zone 6) is confused with front (zone 0)

True Gaze Zone	Recognized Gaze Zone					
	1	2	3	4	5	6
1	80.7	5.3	1.8	5.3	0.0	7.0
2	2.4	94.0	0.0	3.6	0.0	0.0
3	6.5	0.0	93.5	0.0	0.0	0.0
4	1.2	16.7	0.0	76.2	4.8	1.2
5	0.0	0.0	0.0	3.0	97.0	0.0
6	100.0	0.0	0.0	0.0	0.0	0.0

Unweighted Accuracy = 73.6%
Weighted Accuracy = 79.8%

TABLE II: With gaze-surrogate feature: Confusion matrix for 6 gaze zones numbered as shown in Figure 5. Notice the performance improvement in speedometer and other zones.

True Gaze Zone	Recognized Gaze Zone					
	1	2	3	4	5	6
1	98.2	0.0	0.0	1.8	0.0	0.0
2	0.0	98.8	0.0	0.0	0.0	1.2
3	3.2	0.0	96.8	0.0	0.0	0.0
4	3.6	0.0	0.0	91.7	4.8	0.0
5	0.0	0.0	0.0	3.0	97.0	0.0
6	20.8	0.0	0.0	0.0	0.0	79.2

Unweighted Accuracy = 93.6%
Weighted Accuracy = 94.9%

model on our naturalistic driving data to determine how well it performs in real-world conditions. First, a qualitative evaluation is performed by human experts manually inspecting over 1,000 frames with different driver head and eye poses. The analysis shows promising results. Figure 8 shows the images of the 5 good and the 5 bad examples from the naturalistic driving data based on human inspection. Second, the quantitative evaluation of the iris detection is performed based on gaze experiment as discussed in the following section.

C. Results: Gaze Zone Estimation

In this section, we conduct two experiments to estimate driver's gaze zone. First experiment is with head pose information alone, and second experiment is with both head and eye cues. We have divided driver's view into six different gaze zones related to driving tasks. The zones are labeled as following, 1 - front windshield, 2 - center console (infotainment panel), 3 - left window rear-view mirror, 4 - center rear-view mirror, 5 - right window rear-view mirror, and 6 - speedometer.

We use a random forest classifier to estimate driver's gaze zone. In many machine learning applications, random forest has shown promising results. The reason for choosing random forest is because of the ease of the ability to interpret the learned parameters as well as the low number of tuning parameters. In fact, the only parameter that we tune is the number of trees which is 60.

For each experiment, the naturalistic driving data is divided into several short time segments where driver possibly gazed into different zones. We call these segments as events. Each event is randomly assigned to one of ten folds. Each fold contains multiple events, but there is no duplicated events among folds. This is to ensure that evaluation is performed on temporally separated frames. We provide performance results for randomized tenfold cross validation. The system is trained on nine folds at a time and tested on the remaining fold. This procedure is repeated ten times, each time leaving out a different fold. For performance evaluation, we calculate weighted- (overall), unweighted- (per class) accuracy and confusion matrix which are shown in Table I & II.

In the first experiment, we examine the performance of the classifier with head pose information alone i.e. yaw, pitch and roll. Table I shows that the classification result is fairly good for zone 1 to 5 with only head pose information. However, zone 6, the speedometer area, has a very poor result using only the head pose and it is classified as front looking (zone 1). This is expected since drivers usually only move their eyes but do not make any head movement when checking on speedometer. Similarly, zone 4 (center-rear view mirror) is often confused with zone 2 (center consol). This is again because the driver can direct his/her gaze using eye movement alone between these zones.

In the second experiment, we include eye cues i.e. proposed gaze-surrogate measurements. The result is presented

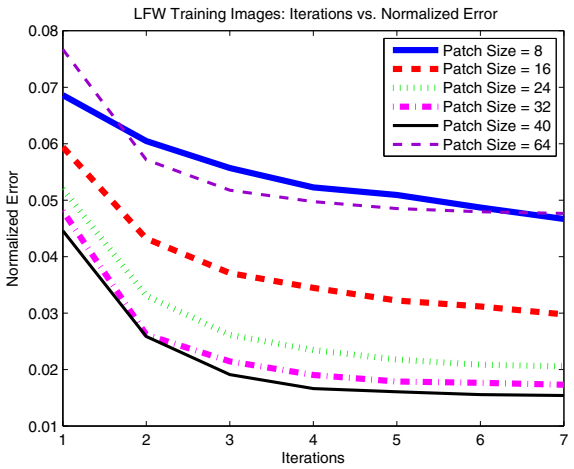

Fig. 6: Normalized error as a function of iterations for models with different patch sizes. Patch size 40 converges fast and provides better error performance.

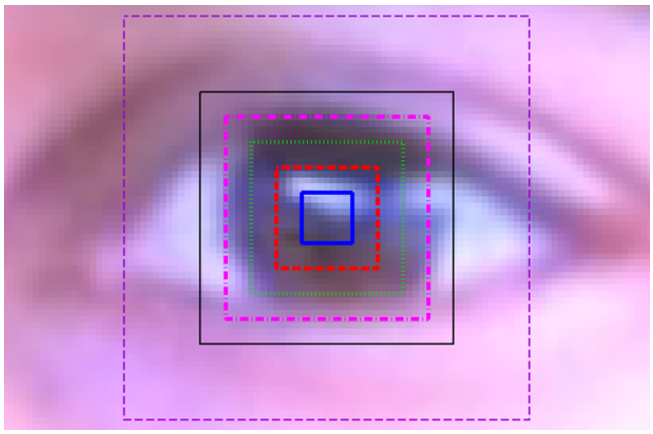

Fig. 7: Illustration of patch size 8 (blue), 16 (red), 24 (green), 32 (pink), 40 (black), and 64 (purple).

in Table II. Clearly, the system with eye features significantly outperforms the head pose alone system by 20% absolute improvement in per-class classification accuracy for the six zones. Notice, the center-console (zone 2) and the center rear-view mirror (zone 4) are well separated with no false-detection between them. This can be attributed to the effectiveness of the vertical-gaze-surrogate feature. Also, note that the zone 1 (front windshield) performance for this system is much better than that for the previous system, where the front zone is confused with the adjacent zones. This can be attributed to the effectiveness of the horizontal-gaze surrogate feature. The greatest improvement is noticed for the zone 6, the speedometer, as expected. Yet, this is the zone with the lowest accuracy and the most confusion with the front zone. This suggests the inherent difficulty associated with separating these two zones. A closer inspection on frame by frame results, though, are inspiring as the adjacent frames are often classified correctly. Hence, a video segment level analysis may further improve the result, which we plan to conduct in the future work.

V. Concluding Remarks

Driver's gaze estimation is an important component of the driver monitoring systems. We motivated that the coarse gaze direction is sufficient in a number of applications. We presented a computational framework to estimate coarse gaze direction and infer gaze zones using head and eye cues. Our emphasis is on the practical realization of a robust and reliable system. Towards this end, we presented a novel iris tracking algorithm with no geometric shape assumptions. Since the features are computed in the local neighborhood, the method enjoys the benefits as in other local patch-based methods in other applications over holistic appearance based approaches. We further proposed gaze surrogate measurements to estimate horizontal and vertical eye movements. Our analyses for gaze zone estimation show very promising results with significant improvement over head pose alone system.

In current implementation, we only use video sequence for iris tracking. In future, we will extent proposed framework to utilize video sequence to also incorporate gaze-dynamics in inferring gaze zones as well as other driver activities [29]. With improved driver's viewing direction using proposed framework and dynamic vehicle surround analysis [30], we will be able to better understand driver intentions.

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] (2014, April) Traffic safety facts: Distracted driving 2012. [Online]. Available: http://www-nrd.nhtsa.dot.gov/Pubs/812012.pdf

[3] 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.

[4] 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.

[5] 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

[6] A. Doshi and M. M. Trivedi, "Tactical driver behavior prediction and intent inference: A review," in *IEEE International Conference on Intelligent Transportation Systems*, Oct 2011, pp. 1892–1897.

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

[8] A. Doshi, B. T. Morris, and M. M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues," *Pervasive Computing, IEEE*, vol. 10, no. 3, pp. 22–34, 2011.

[9] 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*, June 2003, pp. 444–449.

[10] A. Tawari and M. M. Trivedi, "Robust and continuous driver gaze estimation by dynamic analysis of multiple face videos," *IEEE Intelligent Vehicles Symposium*, June 2014.

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

[12] A. Doshi and M. M. Trivedi, "Head and gaze dynamics in visual attention and context learning," in *IEEE CVPR Joint Workshop for Visual and Contextual Learning and Visual Scene Understanding*, June 2009.

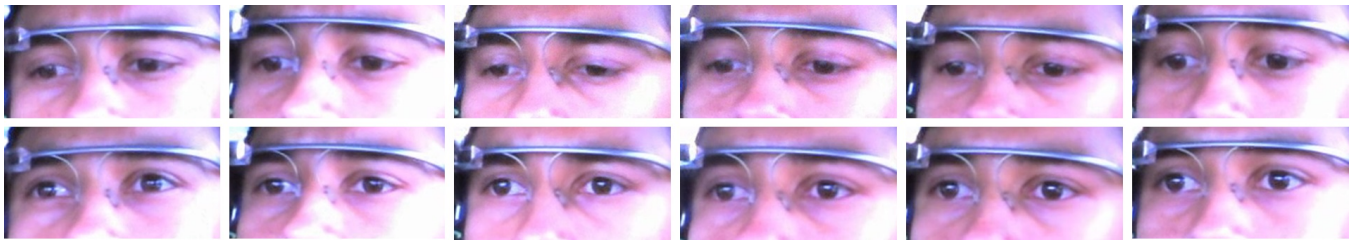

Fig. 8: Challenges faced in estimating gaze direction when the driver looks at the speedometer. The top row is driver looking at speedometer. The bottom row is right before driver looks at speedometer.

Fig. 9: Example results on the real-world driver dataset. The top row has five correct results, and the bottom row has five incorrect results.

[13] J. S. Agustin, A. Villanueva, and R. Cabeza, "Pupil brightness variation as a function of gaze direction," in *Proceedings of the 2006 Symposium on Eye Tracking Research and Applications*, ser. ETRA '06. New York, NY, USA: ACM, 2006, pp. 49–49. [Online]. Available: http://doi.acm.org/10.1145/1117309.1117334

[14] K. Nguyen, C. Wagner, D. Koons, and M. Flickner, "Differences in the infrared bright pupil response of human eyes," in *Proceedings of the 2002 Symposium on Eye Tracking Research and Applications*, ser. ETRA '02. New York, NY, USA: ACM, 2002, pp. 133–138. [Online]. Available: http://doi.acm.org/10.1145/507072.507099

[15] D. Li, D. Winfield, and D. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *Computer Vision and Pattern Recognition - Workshops*, June 2005, pp. 79–79.

[16] A. Yuille, P. Hallinan, and D. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–111, 1992.

[17] P. Wang and Q. Ji, "Learning discriminant features for multi-view face and eye detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, June 2005, pp. 373–379 vol. 1.

[18] D. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, March 2010.

[19] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models." in *British Machine Vision Conference*, vol. 2, no. 5, 2006, p. 6.

[20] X. Xiong and F. De la Torre, "Supervised Descent Method and its Applications to Face Alignment." *In CVPR*, 2013.

[21] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2011.

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

[23] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[24] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[25] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *International Journal of Computer Vision*, vol. 15, pp. 123–141, 1995.

[26] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection." *In CVPR*, March 2005.

[27] A. Tawari, A. Møgelmose, S. Martin, T. Moeslund, and M. M. Trivedi, "Attention estimation by simultaneous analysis of viewer and view," in *IEEE International Conference on Intelligent Transportation Systems*, Oct 2014, p. pp.

[28] A. Tawari, S. Sivaraman, 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 Proceedings*, June 2014, pp. 115–120.

[29] E. Ohn-Bar, S. Martin, A. Tawari, and M. M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *IEEE International Conference on Pattern Recognition*, 2014.

[30] A. Doshi and M. M. Trivedi, "Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions," in *IEEE Intelligent Vehicles Symposium*, June 2009, pp. 887–892.