Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Integration of Head & Eye-Tracking for Driver Monitoring in Vehicular Environment

Sai Charan Addanki[a]*, Rolf Assfalg[b], Hrishikesh Venkataraman[a]

[a]*Indian Institute of Information Technology Sri City, Andhra Pradesh 517646, India*
[b]*DHBW Heidenheim, 89518 Heidenheim, Germany*

Abstract

In the context of Driver Assistance systems, estimating the state of the driver (e.g. distracted, fatigue) is crucial for safe driving. In this paper, state-of-the-art techniques with the subject on Face & Eye-tracking to develop a hybrid model for the real-time driving conditions in both day and night time driving is explored. This paper surveys the advantages and disadvantages of the existing eye and face tracking mechanisms and their integration with the driving performance measures (drivability). A theoretical analysis of Eye-Tracking patterns under a controlled environment is presented and it could provide a reliable solution for safe driving especially at night-time driving.

Keywords: Face-tracking; Eye Gaze Patterns; Driver State; State-of-the-art

1. Introduction

Over the last 10 years, driver assistance systems in the vehicular environment have evolved considerably, starting from an auto air-bag release in case of accident to automatic gear change while driving. At the same time, there have been several approaches made to develop systems for monitoring the on-board alertness level in drivers and his/her driving pattern. Advanced Driver Assistance Systems (ADAS) is a promising technology in which additional electronic devices in motor vehicles are used for supporting the driver in certain driving situations. This focuses on safety aspects and on enhancing driving comfort. Current cars are well equipped with ADAS features such as Attention Assist, Electronic Traction System (ETS), Curve Dynamic Assist, Crosswind Assist, Lane Keeping Assist, Active Blind Spot Assist, Brake Assistance System (BAS), and Active Parking Assist etc. In the context of ADAS, the work in [1] reports a system Driver Assisting System (DAISY) as a monitoring and warning

* Corresponding author
E-mail address: charan.a@iiits.in

aid for the driver in longitudinal and lateral control on German motorways. The warnings are generated based on the knowledge of the behavioural state and condition of the driver. However, this is a model-based approach and with no proper onboard testing and validation leaves a gap for further research. CMU has developed a video-based drowsy driver monitor system [2], which estimates PERCLOS. However, the system addressed can't suffice against illumination variation and with head rotation. The co-pilot system uses bright pupil dark pupil method for the detection of eyes, which requires active illumination method. Significantly, using a dashboard-mounted head-and-eye tracking system [3] developed a Driver Assistance System (DAS), in which the Distillation algorithm is used to monitor the driver's performance. In the paper [4], a drowsy driving detection algorithm is developed using the datasets from a driving simulator. The algorithm relies on steering angle signal only which is used an artificial neural network as a classifier to detect drowsiness. It achieved a high accuracy on classifying the drivers state whether drowsy or awake. Notably, [5] analysed the drowsy drivers steering wheel behaviour and proposed feature sets to capture drowsy steering patterns, and compared five machine learning methods (linear kernel Support Vector Machine (SVM), radial kernel SVM, k-nearest neighbour, decision tree and logistic regression) which reported a recognition rate of 86.1 percentage based on a simulator database. Along with these types of research, there is much other existing research on developing a drowsy driving detection system based on driving pattern, but most of them are conducted using databases acquired from a driving simulator. This makes it quite difficult to implement in a real-time practical environment. Although promising work has been carried out on onboard Driver Assistance Systems, still there are gaps to be filled and need to develop a reliable eye tracking mechanism and integrate with face/head tracking. The integration of remote eye tracking and face movement analysis allows a completely non-intrusive research setup for capturing both emotional arousal and valence. The block diagram of the integration mechanism of head pose and eye gaze estimation is shown in Fig.1. This paper is organized as follows. We instigated the driver state concept in section 2. Next, related work in the field of eye-tracking and head pose estimation is explained in section 3, followed by the state-of-the-art techniques using CNN and DCNN in section 4. In section 5, our proposal on eye patterns and the possible integration with head pose estimates is explained.

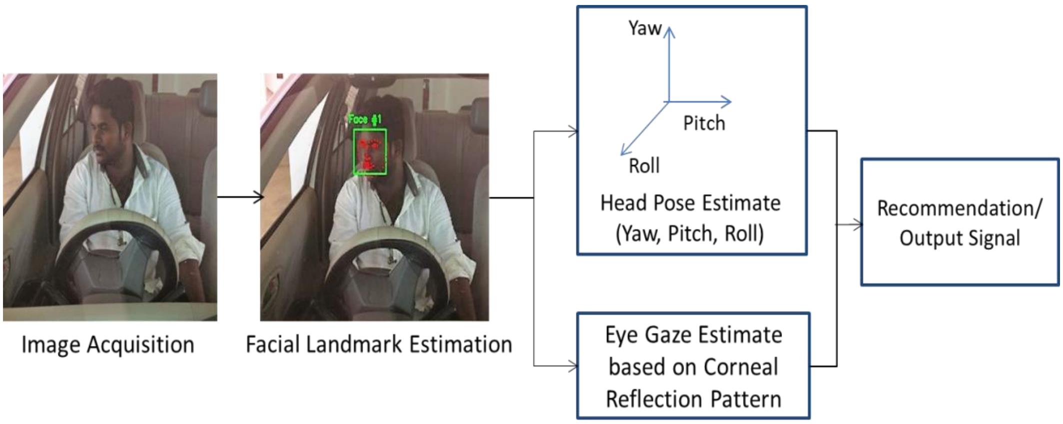

Fig.1. Block diagram - Integration of Face & Eye tracking

2. State of the driver

Extensive work has actively been carried out for decades, with the aim of determining the driver's state. Here, we classify it into two categories: Distraction and Fatigue.

2.1. Distraction

The causes of driver distraction are various and create huge risk factors and close to half of the on-road crashes involves inattention caused by driver distraction [6]. A general definition for the driver distraction has been introduced in [9] as "Driver attention is a diversion of attention away from activities critical for safe driving toward a competing activity". From the driver's functionality point of view, the National Highway Traffic Safety Administration (NHTSA) classified distractions into four categories [6] as detailed in Table 1.

Table.1. Types of Distraction

No.	Types of distraction	Example
1	Visual	Eyes off the road
2	Cognitive	Daydreaming or lost in thought
3	Auditory	Attending the ringing phone call
4	Bio-mechanical	Manual adjustment of in-vehicle components (dashboard)

In the report [10], Safety Vehicle Using Adaptive Interface Technology (SAVE-IT) program mentioned that the glance duration of eyes-off the road, head-off the road, deviation of lane position are significant measures for visual distraction. It is observed in [7] that the driving performance differs considerably with cognitive distraction from visual distraction; Cognitive distraction affects longitudinal vehicle control like car following, whereas Visual distraction affects lateral vehicle control. It is found that [8] cognitive distraction results in under compensation, whereas neglect steering and overcompensation is from visual distraction. A real-time vision-based system [11] is proposed to track the driver's facial features to precisely detect Eyes-Off-the-Road (EOR). A comprehensive experimental analysis is carried out under a variety of lighting conditions and facial expressions and achieved an accuracy of 90% EOR for all the experimented scenarios. It was reported [12] that there was a decrease of temperature at the driver's tip of the nose while performing a cognitive distraction task like thinking about something during driving.

2.2. Fatigue

The case of fatigue varies from the concept of distraction. By the definition in [13], fatigue refers to a combination of the instinctive feeling of drowsiness and symptoms like impaired performance. Studies [14] show that 25% to 30% are fatigue based driving accidents. By the definition [15] given by European Transport Safety Council (ETSC) states that "fatigue concerns the inability or declination to continue an activity, generally because the activity has been going on for too long". The effect of fatigue will change the driver behaviour pattern and in [16], several actions were found correlated with the fatigue; radical increase in eye-blinking, increase in the frequency of yawning, more relaxed hand position on the steering wheel etc. In this paper [17], a general observation was that the speed of the vehicle has not shown any promising correlation with drowsiness. In this report [18], depreciation in the steering reversal frequency was observed with the sleep-deprived drivers. For practical application, [19] developed a simple illumination compensation algorithm and a novel eyelid movement detection method for drowsiness detection systems using a single camera. The system achieved more than 98% of eye detection rate under various illumination conditions. The work in [20] analysed the relationships between drowsiness and lane departure events to figure out the effects of lane departure warning system (LDWS) on drowsy driving. According to this study, 85.3 per cent of the lane departure events that are caused by sleepiness could be prevented by LDWS. Further, [21]-[23] estimated driving fatigue based on analysis of steering wheel operation. Moreover [24], [25] and Mercedes-Benz's Attention Assist System adopted different vehicle-manipulation signals for better discrimination of drivers alertness state.

3. Head & eye-tracking mechanisms

Broadly speaking, gaze direction estimation approaches can be grouped in two types. One is using only the driver's head pose and other is to use head pose and gaze. In this section, the existing works on face and eye-tracking mechanisms are described; where most of the works are based on Machine Learning techniques.

3.1. Eye-tracking

The geometric and motion characteristics of the eyes are unique which makes gaze estimation and tracking important for many applications such as human attention analysis, human emotional state analysis, interactive user interfaces and human factors. There are many different approaches to implementing eye detection and tracking systems [26]. Gaze estimation methods can be divided into model-based or appearance-based [27]. Model-based approaches use a geometric model of an eye and can be subdivided into corneal-reflection-based and shape based methods. Corneal reflection- based methods [28], [29] rely on external light sources to detect eye features. On the other hand, shape-based methods [30], [31] infer gaze direction from observed eye shapes, such as pupil center and iris edges. These approaches tend to suffer from low image quality and variable lighting conditions, as in our scenario. Appearance-based methods [32]-[34] directly use eyes as input and can potentially work on low-resolution images. Appearance-based methods are believed [35] to require larger amounts of user-specific training data as compared to model-based methods. Recently, several gaze estimation systems have been introduced, and decent advancements have been made by both industry and academic community [36]-[37]. However, still, there is a gap for further improvement in the robustness and accuracy of the systems. Developed a complete gaze estimation method [39] based on a weighted linear regression technique, which achieved a high estimation accuracy (less than 1-degree visual angle error). In this, Supervised Decent Method (SDM) is used to localize and track the eyes. Some commercial products are coming to the market such as Smart Eye and Eye Alert Fatigue Warning System [40], [41]. A robust system is proposed for the recognition of facial expressions in different illuminating conditions [42]. An array of NIR camera and a Visible light camera is equipped for experimentation. It is observed that sorting rate using the NIR camera is better when compared with the results of visible light. The proposed NIR camera setup can be used for reliable recognition of facial expression in changing lighting conditions. A camera-based system [43] with fixed parameters is proposed to combine the computation of head pose and gaze detection from an image. No user calibration is required during the process. Both head pose and eye gaze detection are based on their Geometric models. Optical flow based Head Movement and Gesture Analyzer (OHMeGA) [44] presented and analysed on an available dataset, which is robust to large head movement, occlusions from eyewear, varying lighting conditions etc. The results observed in the paper [44] show a decent accuracy of 97.4% in a laboratory environment, but only 86% accuracy in an on-road experiment. In this article [46], drivers gaze estimation method is proposed using visible spectrum eye tracking provided with the failure of conventional gaze detection approach using IR spectrum. This eye-tracking method doesn’t need calibration of the head movement, as it uses automated facial landmark detection for head pose calculation. Driver’s Attentive distraction and Cognitive distraction are classified by evaluating Laplacian support vector machine (LSVM) & Semi-supervised extreme learning machine using eye and head movements in real driving conditions [47]. In considering driver visual attention, [48] proposed a regression model for eye gaze estimation based on head orientation. The system achieved high accuracy in horizontal direction, but low performance in vertical direction.

3.2. Head tracking

Head pose estimation can be linked with visual gaze estimation i.e., estimation of direction and focus of a person’s eyes. Physiological investigation case studies have concluded that a person’s gaze prediction comes from a combination of both head pose and eye gaze direction [49]. There are several literature surveys considered general human motion [50]-[51], face detection [52]-[53], face recognition [54] etc. The work in this paper [55] presents a state-of-art survey paper on eye-tracking and head pose estimation methods. Despite the amount of research carried in these fields, still, there is a gap for further improvement in the robustness and accuracy of the systems for other applications as well. In this paper [45], extensive experimental evaluations are carried out for head movement estimation using a unique and novel head pose dataset of naturalistic driving scenario on freeways. Tracking and analysing of the geometrical facial features is done for head pose estimation using a 3D model along with different camera configurations. There has been several head pose estimation approaches considering different functionalities [56] like Appearance template methods, Detector array methods, Nonlinear regression methods, Manifold embedding methods, Flexible methods, Geometric methods, and Tracking methods. In the Appearance template method, initially a set of Support Vector Machines (SVMs) is trained to localize the face, and eventually, the support vectors are used as templates to approximate the head pose [57]-[58]. In the Detector array methods, image evaluation is carried out with a detector trained on several images using a supervised learning algorithm. For pose estimation [59], a classifier called router classifier is used to choose a single consequent detector. In the case of nonlinear regression methods, pose estimation is done by nonlinear mapping of image space to pose directions, and neural networks have been widely used for head pose estimation. If we consider nonlinear regression tools for head pose estimation, neural networks have been the most used in the literature. For example, Multi-Layer Perceptron (MLP), comprises many feed-forward cells in multiple layers [60]-[61]. An MLP can also be trained for precise head pose estimation over a continuous pose range. In this method, the network has a single output for each degree of freedom (DOF) and the process of output initiation is proportional to its respective orientation [62]-[63]. In the Manifold embedding methods, head pose estimation is carried out by considering allowable pose variations to define a rigid model of a head, and an embedded technique to provide a new sample into the manifold. Some of the Manifold embedding methods include; KLDA [64] for coarse head pose estimation, Isometric feature mapping [65] for nonlinear dimensionality reduction, Locally Linear Embedding [66] an Eigen vector based dimensionality reduction, etc. An example of the Flexible model is Elastic Bunch Graph [67], created for every discrete pose, and compared the subsequent new perception of the head. Active Appearance Model (AAM) [68], is one of the evolved flexible models for head pose estimation. In Geometric methods, five facial points are used as a facial symmetry axis to estimate the head pose. Using the same five facial points, and the coarse estimate of the nose position [64], the head pose can be estimated. In tracking methods, usually, head pose estimation is based on a rigid 3D model of the head, in which the movement of the head is observed as a transformation model. Tracking approaches can be made automatically initialized, using templates whenever the original view of the head is compared to the head pose estimate [69].

4. Convolutional Neural Network (CNN)

With the high GPU computational power and access to huge datasets, Deep learning has been into active play in recent years. Importantly, CNN [70] has been used in various domains like Medical, Automotive, Satellite imaging and Manufacturing for different applications. However, DNNs has been used in human face related applications for better precision and classification accuracy. In this paper [71], a three-stage cascade structure is proposed using two shallow neural networks and one DNN for coarse-to-fine feature localization of the face. They also analysed the absolute value rectification and local weight sharing techniques for feature localization. A CNN cascade structure is proposed for face detection, robust to different face pose, illumination variation and facial expression [72]. Using DNN, a learning method is proposed to map the relation between 3D head movement and its visual appearance. In this method, with the assumption of accurate face detection, the network model estimates the head movement with a relative error of 3 degrees. Since this method doesn’t consider the face detection part which is crucial in head pose estimation; and its challenging to obtain the mentioned performance in real-time. In these papers [73]-[74], they considered head pose estimation as a nonlinear regression problem which computes 3D pose parameters continuously. Many works on continuous head pose estimation use depth information with Machine Learning techniques. A random forest based framework [75]-[76] for the localization of 3D facial landmarks has been proposed for real-time head pose estimation. They also provided the Biwi Kinect Head Pose database; consists of depth maps, colour images, and ground truth data. With the depth information, the advantage is these methods obtain better results, especially at night time. However, the disadvantage is the requirement of such a device, say Kinect and the application is meant mainly for the indoor environment only. A new synthetic dataset was built based on a public 3D face dataset [77]. It comprises a large number of facial expressions and large head pose variations, making it one of the best facial landmark localization datasets to date. Face shape consistency for landmark localization with minimal texture can be a challenge to be addressed for other applications. A deep CNN based face detection method [78] introduced for discrete head pose estimation overcoming the challenges like varying lighting conditions, occlusions, different viewpoints etc. VIVA-Face dataset, a publicly available dataset is used for training the CNN. In the Multi-Task Learning (MTL) approach, the paper [79] describes the novel method for addressing the tasks of face detection, head pose estimation, and facial landmark localization and the same was later extended in [80]. In this method, each facial landmark is modeled as a separate part and uses a mixture of trees to estimate the topological changes with viewpoint variations. For simultaneous face and landmark point detection in an image, a joint-cascade method [81] is proposed which gives a better detection performance. In this paper [82], features only from the last layer is used for training a CNN for pose estimation but the use of MTL is limited for face analysis. In the method [83], a CNN model using MTL is provided for facial landmark localization with the discrete head pose yaw angle estimation, smile and gender detection. In real-time driving conditions, a novel dataset Drive A Head [84] developed to evaluate head pose estimation algorithms. It is stated that the proposed head pose dataset is the largest publicly available dataset providing both 2D and 3D data at the pixel level. Notably, a deep learning model for head pose estimation is presented that achieved better performance than existing algorithms by analysing its performance with their dataset. A reliable deep learning based head pose estimation method [85] is proposed and is evaluated on CAS-PEAL data set. The results of the experimentation showed that the method can be effective for more accurate and precise pose estimation. With little changes on existing CNN structures, a new model [86] is introduced to perform face detection and landmark-based head pose estimation. Alex Net structure and Stacked Hourglass Network is used for refined face detection and localization. The proposed work can provide a solid foundation for advanced facial analysis of the driver. For mobile devices, an end-to-end platform [87] for eye-tracking is proposed and a Deep Convolution Neural Network (DCNN) is used for robust prediction of gaze and achieved a low error of 1.04 cm on mobile phones.

Reference	Technique/Model/Approach	Objective
27, 28, 29, 32, 33, 34	Appearance based	Eye tracking
30, 31	Shape based	Eye tracking
39	Linear Regression (SDM)	Eye tracking
43	3D Cascade Regression	Head tracking
59, 60, 61	Appearance Template (SVMs)	Head tracking
61	Detector Arrays (Router Networks)	Head tracking
64, 65	Nonlinear Regression (MLP)	Head tracking
66, 67, 68	Manifold Embedding (KLDA, Isomap, LLE)	Head tracking
69, 70	Flexible Models (Elastic Graph Matching, AAM)	Head tracking
66, 44, 48	Geometric (Planar and 3D)	Head & Eye tracking
46	Optical Flow	Head tracking
50	LSVM	Head & Eye tracking
73, 74	CNN	Head Pose Estimation
81, 82, 83, 84, 85	MTL using CNN	Head Pose Estimation
80, 86, 87, 88	DCNN	Head Pose Estimation
77, 78	Random Forest	Head Pose Estimation

Fig.2. Summary of the sections (3 and 4)

5. Beyond State-of-the-art: A Proposal

Most of the vision-based eye trackers use IR light source for illumination of the eye, wherein the light produces a reflection on the cornea, named as corneal reflection. Many works on the gaze estimation are based on the corneal reflection and importantly serves as a reference point. Typically, with the direction of gaze, the positional relationship between corneal reflection and pupil will change. From the literature [88]-[90], it is found that there are a variety of eye tracking systems to detect the gaze point. However, the conceptual base looks to be same as the eye captured by the camera changes with the rotation of the eye or translation in 3D space. There are few eye trackers available in the market. For instance, eye tracker [91] technology is based on a single camera and one IR source. By using multiple IR sources to the system will give better results compared with the single light source. Several works have been done by changing the camera setup and light sources. [92] developed a single camera system with two light sources and [93] introduced a 3D position computation method for an eye using one camera and two light sources. [94] Used four light sources placed on the corners of the computer screen and one camera positioned slightly below the center of the screen to compare and improve cross-ratio based eye trackers for improving head movement tolerance. The challenge with a fixed camera system is the limited view of angle needed to acquire sufficient high-resolution eye images for reliable gaze estimation. To overcome the limitation, research has been done using multiple camera systems with a wide-angle lens or movable narrow-angle lens for gaze tracking. The work in [95] introduced a 3D gaze tracking mechanism using two stereo cameras and two light sources. [96] Presented a two wide-angle cameras system, wherein using one camera, the coarse eye region is detected by a pre-defined eye detector based on shape features. Then another camera is used to fine focus the eye region. By considering the concept of multiple cameras as the base idea, two different methods are proposed for head tracking and are explained as follows:

5.1. Static & Active camera

In the Static method, three cameras can be used for face/head position estimation, in which one camera is a reference camera; mounted on the driver's dashboard to detect and monitor the frontal face of the driver. The other two cameras are mounted to the left and right side of the drivers face; provided with the input of face/head pose angle with reference to the reference camera. The reference camera detects and keeps tracking the face of the driver. If there is a change in face position to the left, then a signal is sent to the left camera resulting in the (followed by) tracking of the face/eyes by the left camera. The same process continues for the right camera also. The main advantage of this method is to track the eyes, even with the movement of the head. The disadvantage is the delay in processing of the signal from one camera to another and the fusion of the information of each camera is critical. The position of the three static cameras is depicted in Fig. 3.

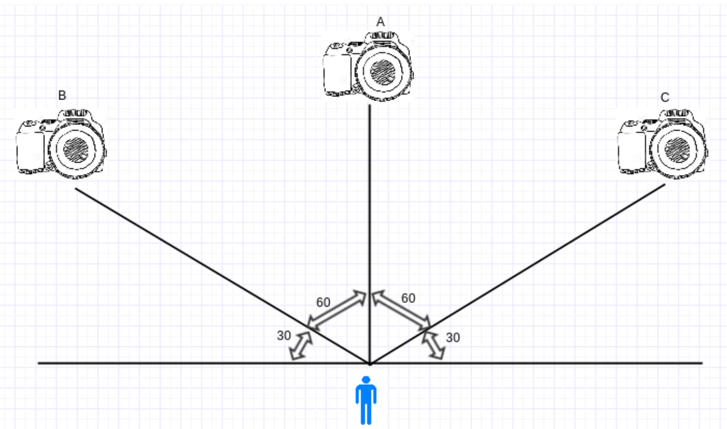

Fig.3. The position of the three cameras from the driver's point of view

In the Active method, a single customized camera with actuator/motor attached could be used for the purpose of tracking. The actuator works in such a way that the camera moves (0-180) degrees with the drivers face as the reference plane. With this, the camera keeps moving in turns, tracking the face along with the movement of the face in both horizontal and vertical direction.

5.2. Lighting conditions

In this work, the light source is considered into two categories. One is sunlight, and the other is the external light source (headlights of the opposite vehicle). In the process of image detection of eyes, the image in the eyes of the driver may not only be created by the IR illumination inside the car, but also by the illumination of outside vehicle or sunlight. So there is a need to eliminate the unwanted illuminations which distract the eye detection. As sunlight has IR rays [97], one possible solution to stop the penetration is to have a windshield screen with IR elimination filters [98] (which is not feasible for car manufacturers). Another solution could be an adjustable (automatic) screen/shield according to the angle of penetration of sun light through the wind shield screen on drivers face/eyes. The two major light sources in particular to a real-time driving environment are described below.

- Case 1: Sunlight: While driving, Sunlight is one of the predominant obstacles for the driver's eyesight. The effect black out the view of oncoming vehicles or pedestrians or traffic control devices by eliminating the image of the retina with a bright spot or pattern. Evaluation of sunlight effect requires many factors to be considered like latitude and longitude, weather conditions, driver position, time of the day, road direction, whether the driver is wearing sunglasses, and any other parameters that could affect the line of sight.
- Case2: Oncoming vehicles high/low beam: In recent times, the automobile companies has come up with intelligent head light technologies to overcome the problem of driver blinding because of oncoming/preceding vehicles head lights/tail lights at night time. For example, BMW cars have Intelligent Head Light Technology [99] providing significantly improved visibility at night. The process could be explained as follows:
  - A camera mounted inside the car for outside view automatically detects the traffic situation at night and passes the information on to the headlights through a control unit.
  - The headlights adapt the light distribution according to the traffic situation.
  - High beam remains active without glaring the vision of any road traffic participants.

The road traffic situation significantly varies from region to region. For example, consider the India scenario, where most of the cars are not equipped with the latest technology components and it is still a major problem for the drivers to drive at night time. So there is a need to investigate to address such issues for safe driving. Mainly the car manufacturers looked at the optimization of outside lighting illumination [99] from headlights, but there has been to a solution to prevent/limit the light beam penetration through the windshield onto the drivers face. Since our goal is to develop a real-time eye-tracker for vehicular application, many parameters need to be considered including sunlight and other (Oncoming and Preceding) vehicle light beam. Studies state that even at a constant light, the size of the pupil varies with the emotions of the person. So, pupil size needs to investigated by taking multiple readings at different emotion levels [100] for different persons. The reflection pattern on the eyes with glasses, and without glasses makes a huge difference, and could depend:

- On the distance between the eyes and glasses: The reflection pattern [101] might change with the distance between the eyes and the wearable glasses.
- On eyesight (positive/plus) [102]: Because of the thickness of the lens glass, the brightness of the image will be less, and the image size looks bigger.
- On eyesight (negative/minus) [102]: In this case, brightness will be less, image size looks smaller; which makes the detection of eyes difficult. In addition, Noise will be more, and it will be less accurate.
- On the viewing angle of the driver wearing glasses whether through glasses or naked eye: The reflection pattern on the glasses will be very different with the reflection pattern on a naked eye. Some people look at the objects through the wearable glasses, and sometimes through the naked eye. In that case, it will be difficult to capture and analyse the eye moving pattern considering both the cases for a single person.
- On bifocal lens [103]: There are two categories of bifocal lenses: Translating and Simultaneous. Each lens has multiple designs. Translating or alternating bifocal lenses allow your pupil to alternate between two power segments when you look either upward or downward. Usually, the distance power is the upper segment and the near power is the lower segment. For example, if you glance downward to read the newspaper, the lower segment will make the words larger and clearer. Simultaneous bifocal lenses enable you to look at both distance and near powers at the same time.

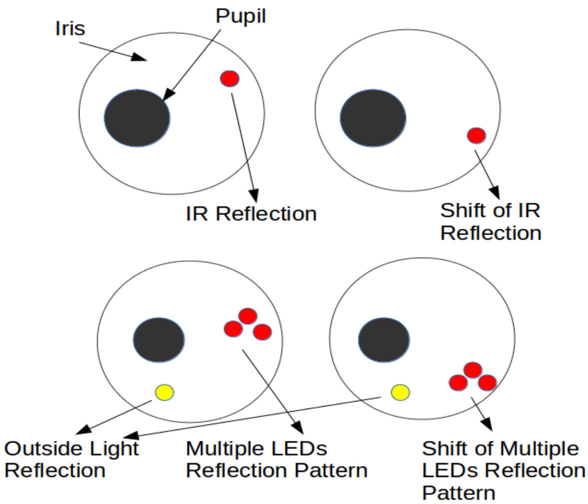

Fig.4. Reflection Patterns of Light on Eye

For example, consider an eye-gaze position on any object (dashboard, display). Relative deviation of gaze can be analysed, but absolute position of the gaze is unknown (derived by the position of the reflection by IR Led on eye) depends on different parameters like driver position, distance between the windshield etc., In order to distinguish between inside light and outside light(disturbance), the IR reflection pattern of both should be different(maybe in size or in pattern of reflection), but there is a constraint at which there is a possibility of overlapping of both inside and outside reflections on one another, which is depicted in Fig. 4. With this, the desired reflection pattern of light is observed and can be used to estimate the gaze direction. The head pose estimation is carried out simultaneously with the eye-tracking for continuous driver monitoring. The Block diagram for the development of a stand-alone eye-tracker is shown in Fig.5.

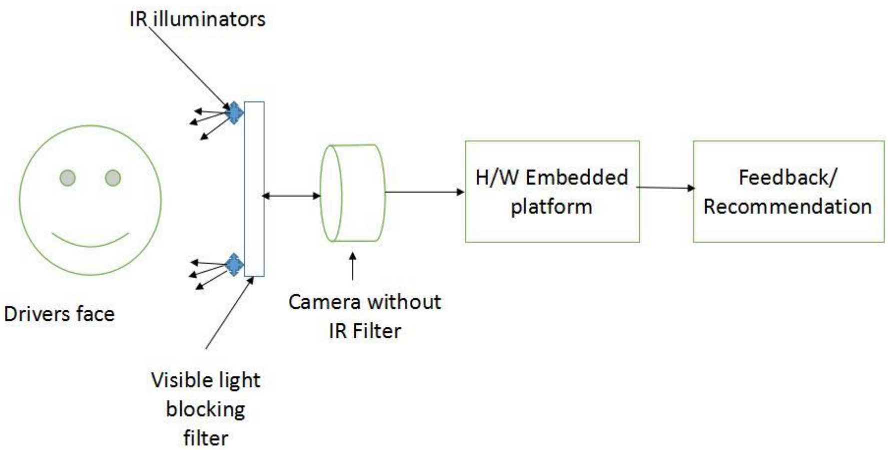

Fig.5. Block diagram of a stand-alone Eye-tracker

The proposed approach is applicable for both the day and night time driving scenarios. Driving at night time is a challenging task for a new or an experienced driver and the driver attention monitoring is more important as it comes with various challenges like limited vision, glare from headlights of the outside vehicles causing eye dizziness etc. As discussed earlier, estimation of driver's visual attention is a crucial part for driver assistance systems and eye tracking can provide good gaze estimation. In practice, eye-tracking difficulties involve occlusion of the eye in presence of glasses, rapid changes in illuminating conditions, vehicle vibration, low-quality video, motion blur etc. Computational hardware/software requirement and good resolution recording equipment further increase the difficulty of making practical and deployable solutions.

6. Conclusion

In this paper, we presented a detailed survey of the existing and state-of-the-art mechanisms of head pose estimation along with eye-tracking. A theoretical analysis has been proposed on the integration of head pose with eye-tracking for driver monitoring especially at night time driving to overcome the outside light reflections on driver's eyes.

7. References

[1] R. Onken., 1994. DAISY, an adaptive, knowledge-based driver monitoring and warning system. In Proceedings of the Intelligent Vehicles Symposium.
[2] E. M. Ayoob, A. Steinfeld, and R. Grace., 2003. Identification of an 'appropriate' drowsy driver detection interface for commercial vehicle operations. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
[3] L. Fletcher, L. Petersson, and A. Zelinsky., 2003. Driver assistance systems based on vision in and out of vehicles. In Proceedings of the Intelligent Vehicles Symposium.
[4] R. Sayed and A. Eskandarian., 2001. Unobtrusive drowsiness detection by neural network learning of driver steering. In Proceedings of the Institution of Mechanical Engineers. Part D, Journal of Automobile Engineering, 215-969-975-5.
[5] J. Krajewski, D. Sommer, U. Trutschel, D. Edwards, and M. Golz., 2009. Steering wheel behaviour based estimation of fatigue. 5th International driving symposium on human factors in driver assessment, Training and vehicle design, 118-124.
[6] T. A. Ranney, E. Mazzae, R. Garrott, and M. J. Goodman., 2000. NHTSA driver distraction research: Past, present, and future. Transp. Res. Center Inc., East Liberty, OH, Technical Report.
[7] O. M. J. Carsten and K. Brookhuis., 2005. Issues arising from the haste experiments. Transp. Res. Part F: Traffic Psychol. Behaviour, vol. 8, no. 2.
[8] Y. Liang and J. D. Lee., 2010. Combining cognitive and visual distraction: Less than the sum of its parts. Accid. Anal. Prev., vol. 42, no. 3, pp. 881-890.
[9] J. D. Lee, K. L. Young, and M. A. Regon., 2009. Driver Distraction: Theory, Effects and Mitigation. Boca Raton, FL-CRC.
[10] H. Zhang, M. Smith, and R. Dufour., 2008. A final report of safety vehicles using adaptive interface technology (Phase II: Task 7c): Visual distraction. Delphi Electron. Safety, Kokomo, IN, Tech. Report.
[11] F. Vicente, Z. Huang, X. Xiong, F. D. Torre, W. Zhang, and D. Levi., 2015. Driver Gaze Tracking and Eyes Off the Road Detection System. IEEE Trans. on Intelligent Transportation Systems.
[12] M. Itoh., 2009. Individual differences in effects of secondary cognitive activity during driving on temperature at the nose tip. In Proc. IEEE Int. Conf. Mechatronics Automation, pp. 7-11.
[13] STRID., 2006. Sub-Group on Fatigue, Canadian operational definition of driver fatigue, Ottawa, ON, Canada.
[14] T. R. S. for the Prevention of Accidents., 2001. Driver fatigue and road accidents: A literature review and position paper, Birmingham, U.K.
[15] H. D. Croo, M. Bandmann, G. M. Mackay, K. Rumar, and P. Vollenhoven., 2001. The role of driver fatigue in commercial road transport crashes. European. Transp. Safety Council, Brussels, Belgium, Tech. Report.
[16] A. Eskandarian, R. Sayed, P. Delaigue, J. Blum, and A. Mortazavi., 2007. Advanced driver fatigue research. U.S. Dept. Transp., Fed. Motor Carrier Safety Admin., Washington, DC, Tech. Rep., Rep. FMCSARRR-07- 001.
[17] T. A. Dingus, L. Hardee, and W. W. Wierwille., 1985. Development of impaired driver detection measures. Dept. Ind. Eng. Oper. Res.,Virginia Polytechnic Inst. State Univ., Blacksburg, VA, Tech. Rep., Dept. Rep. 8504.
[18] T. Mast, H. Jones, and N. Heimstra., 1989. Effects of fatigue on performance in a driving device: Highway research record. In Driver Fatigue Research: Development of Methodology, Haworth, Vulcan, Triggs, and Fildes, Eds. Victoria, Australia: Accid. Res. Center, Monash Univ. Australia.
[19] P. Ilkwon, A Jung-Ho, and B Hyeran., 2006. Efficient Measurement of Eye Blinking under Various

Illumination Conditions for Drowsiness Detection Systems. Pattern Recognition, ICPR.
[20] D. Rimini, T. Altmueller, L. Ladstaetter, and M. Rossmeier., 2005. Effects of lane departure warning on drowsy drivers performance and state in a simulator. In Proceeding of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 88-95.
[21] Y. Takei and Y Furukawa., 2005. Estimate of drivers fatigue through steering Motion. In IEEE International Conference on Man and Cybernetics, pp. 1765-1770.
[22] J Krajewski et al., 2009. Steering Wheel Behaviour Based Estimation of Fatigue. In Proceeding International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.
[23] J Krajewski et al., 2010. Detecting fatigue from steering behaviour applying continuous wavelet transform. In Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research, pp.326-329.
[24] M. Farid et al., 2006. Methods to develop a driver observation system used in an active safety system. VDI Berichte, vol. 1960, pp. 639-650.
[25] Y. J. Zhong et al., 2007. Localized energy study for analyzing driver fatigue state based on wavelet analysis. In International Conference on Wavelet Analysis and Pattern Recognition, pp. 1843-1846
[26] D. W. Hansen and Q. Ji., 2010. In the eye of the beholder: A survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell.,vol. 32, no. 3, pp. 478-500.
[27] D. H. Yoo and M. J. Chung., 2005. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. CVIU.
[28] Z. Zhu and Q. Ji., 2005. Eye gaze tracking under natural head movements. CVPR.
[29] Z. Zhu, Q. Ji, and K. P. Bennett., 2006. Nonlinear eye gaze mapping function estimation via support vector regression. In Pattern Recognition Letters.
[30] J. Chen and Q. Ji., 2008. 3D gaze estimation with a single camera without IR illumination. ICPR.
[31] R. Valenti, N. Sebe, and T. Gevers., 2012. Combining head pose and eye location information for gaze estimation. TIP.
[32] W. Sewell and O. Komogortsev., 2010. Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. SIGCHI.
[33] F. Lu, Y. Sugano, T. Okabe, and Y. Sato., 2014. Adaptive linear regression for appearance-based gaze estimation. IEEE Trans. PAMI.
[34] F. Lu, T. Okabe, Y. Sugano, and Y. Sato., 2014. Learning gaze biases with head motion for head pose-free gaze estimation. Image and Vision Computing.
[35] Z. Zhu and Q. Ji., 2004. Eye and gaze tracking for interactive graphic display. Machine Vision and Applications, vol. 15, no. 3, pp. 139-148.
[36] M. Reale, S. Canavan, L. Yin, K. Hu, and T. Hung., 2011. A multi-gesture interaction system using a 3-d iris disk model for gaze estimation and an active appearance model for 3-d hand pointing. IEEE Transactions on Multimedia, vol. 13, pp. 474-486.
[37] R. Valenti, N. Sebe, and T. Gevers., 2012. What are you looking at?. IJCV, vol. 98, pp. 324-334.
[38] C. Topal, S. Gunal, O. Kocdeviren, A. Dogan, and O. Gerek., 2014. A low computational approach on gaze estimation with eye touch system. IEEE Transactions on Cybernetics, vol. 44, pp. 228-239.
[39] N. M. Arar, H. Gao, and J. P. Thiran., 2016. Regression-based User Calibration Framework for Real-time Gaze Estimation. IEEE Trasns. On Circuits and Systems for Video Technology.
[40] http://www.smarteye.se - Accessed on 27th April 2018
[41] http://www.eyealert.com - Accessed on 27th April 2018
[42] L. A. Jeni, H. Hashimoto, and T. Kubota., 2012. Robust Facial Expression Recognition using Near Infrared Cameras. Journal of Advanced Computational Intelligence.
[43] J. Y. Kaminski, D. Knaan, and A. Shavit., 2007. Single image face orientation and gaze detection. Machine Vision and Applications, pg.85-98.
[44] S. Martin, C. Tran, A. Tawari, J. Kwan, and M. M. Trivedi., 2012. Optical flow based Head Movement and Gesture Analysis in Automotive Environment. 15th International IEEE Conference on Intelligent Transportation Systems.
[45] A. Tiwari, S. Martin, and M. M. Trivedi., 2014. Continuous Head Movement Estimator for Driver Assistance: Issues, Algorithms, and On-Road Evaluations. IEEE Transactions on Intelligent Transportation Systems, Vol. 15, No.2.
[46] T. Imabuchi, O. D. A. Prima, and H. Ito., 2016. Visible Spectrum Eye Tracking for Safety Driving Assistance.

Springer International Publishing Switzerland, pp. 428-434.
[47] T. Liu, Y. Yang, G. B. Huang, Y. K. Yeo, and Z. Lin., 2015. Driver Distraction Detection Using Semi-Supervised Machine Learning. IEEE Transactions on Intelligent Transportation Systems.
[48] S. Jha and C. Busso., 2014. Analyzing the Relationship between Head Pose and Gaze to Model Driver Visual Attention. IEEE 19th International Conference on Intelligent Transportation Systems.
[49] A. L. Yuille, P.W. Hallinan, and D.S. Cohen., 1992. Feature extraction from faces using deformable templates. Int J Comput Vis 8(2):99-111.
[50] J.P Ivins and J Porrill., 1998. A deformable model of the human iris for measuring small 3-dimensional eye movements. Machine Vision Applications, 11(1):42-51.
[51] S Kawato and N Tetsutani., 2002. Real-time detection of between-the eyes with a circle frequency filter. In Asian conference on computer vision.
[52] G Sommer, M Michaelis, and R Herpers., 1998. The SVD approach for steerable filter design. In Proceedings of international symposium on circuits and systems 1998, Monterey, California, vol 5, pp 349-353.
[53] G Yang and A Waibel., 1996. A real-time face tracker. In workshop on applications of computer vision, pp 142-147.
[54] G Loy and A Zelinsky., 2003. Fast radial symmetry transform for detecting points of interest. IEEE Trans PAMI 25(8):959-973.
[55] A. A. Rahayfeh and M. Faezipour., 2013. Eye Tracking and Head Movement Detection: A State-of-Art Survey. IEEE Journal of Translational Engineering in Health and Medicine.
[56] T. Moeslund, A. Hilton, and V. Kruger., 2001. A Survey of Computer Vision-Based Human Motion Capture. Computer Vision and Image Understanding, vol. 81, no. 3, pp. 231-268.
[57] T. Moeslund, A. Hilton, and V. Kruger., 2006. A Survey of Advances in Vision-Based Human Motion Capture and Analysis. Computer Vision and Image Understanding, vol. 104, no. 2, pp. 90-126.
[58] E. Hjelmas and B. Low., 2001. Face Detection: A Survey", Computer Vision and Image Understanding, vol. 83, no. 3, pp. 236-274.
[59] T. Horprasert, Y. Yacoob, and L. Davis., 1997. An Anthropometric Shape Model for Estimating Head Orientation", Proc. Third Intl Workshop Visual Form, pp. 247-256.
[60] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld., 2003. Face Recognition: A Literature Survey. ACM Computing Surveys, vol. 35, no. 4, pp. 399-458.
[61] B. Fasel and J. Luettin., 2003. Automatic Facial Expression Analysis: A Survey. Pattern Recognition, vol. 36, no. 1, pp. 259-275.
[62] E.M. Chutorian and M.M.Trivedi., 2009. Head Pose Estimation in Computer Vision: A Survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 4.
[63] J. Ng and S. Gong., 2002. Composite Support Vector Machines for Detection of Faces across Views and Pose Estimation", Image and Vision Computing, vol. 20, nos. 5-6, pp. 359-368.
[64] R. Duda, P. Hart, and D. Stork., 2001. Pattern Classification, second ed. John Wiley & Sons.
[65] E. Seemann, K. Nickel, and R. Stiefelhagen., 2004. Head Pose Estimation Using Stereo Vision for Human-Robot Interaction. Proc. IEEE Intl Conf. Automatic Face and Gesture Recognition, pp. 626-631.
[66] R. Stiefelhagen., 2004. Estimating Head Pose with Neural Networks-Results on the Pointing04 ICPR Workshop Evaluation Data. Proc. ICPR Workshop Visual Observation of Deictic Gestures.
[67] M. Voit, K. Nickel, and R. Stiefelhagen., 2007. Neural Network-Based Head Pose Estimation and Multi-View Fusion. Multimodal Technologies for Perception of Humans: Proc. First Intl Workshop Classification of Events, Activities and Relationships, pp. 291-298.
[68] M. Voit, K. Nickel, and R. Stiefelhagen., Head Pose Estimation in Single- and Multi-View Environments Results on the CLEAR07 Benchmarks. Proc. Intl Workshop Classification of Events, Activities and Relationships.
[69] Raytchev, I. Yoda, and K. Sakaue., 2004. Head Pose Estimation by Nonlinear Manifold Learning", Proc. 17th Intl Conf. Pattern Recognition, pp. 462-466.
[70] Y. Lecun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel., 1989. Backpropagation applied to handwritten zip code recognition. Neural Computation 1 (4), 541-551.
[71] Y. Sun, X. Wang, and X. Tang., 2013. Deep convolutional network cascade for facial point detection. CVPR. pp. 3476-3483.
[72] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua., 2015. A convolutional neural network cascade for face detection. CVPR. pp.5325-5334.

[73] C. B. Abdelkader., 2010. Robust head pose estimation using supervised manifold learning. ECCV. pp. 518-531.
[74] H. Ji, R. Liu, F. Su, Z. Su, and Y. Tian., 2011. Robust head pose estimation via convex regularized sparse regression. ICIP, pp. 3617-3620.
[75] G. Fanelli, T. Weise, J. Gall, and L. V. Gool., 2011. Real time head pose estimation from consumer depth cameras. ICPR. pp. 101-110.
[76] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. V. Gool., 2013. Random forests for real time 3d face analysis. IJCV.101, 437-458.
[77] M. Oliu, C. Comeanu, L. A. Jeni, and J. F, Cohn., 2016. Continuous Supervised Descent Method for Facial Landmark Localisation. ACCV.
[78] K. Yuen, S. Martin, and M. M. Trivedi., 2016. On looking at Faces in an Auomobile: Issues, Algorithms and Evaluation on Naturalistic Driving Dataset. ICPR.
[79] X. Zhu and D. Ramanan., 2012. Face detection, pose estimation, and landmark localization in the wild. CVPR. pp.2879-2886.
[80] X. Zhu and D. Ramanan., 2015. FaceDPL: Detection, pose estimation, and landmark localization in the wild. preprint.
[81] D. Chen, S. Ren, Y. Wei, X. Cao, and J. Sun., 2014. Joint cascade face detection and alignment. In European Conference on Computer Vision, volume 8694, pages 109-122.
[82] G. Gkioxari, B. Hariharan, R. B. Girshick, and J. Malik., 2014. R-CNNs for pose estimation and action detection", CoRR-abs/1406.5212.
[83] Z. Zhang, P. Luo, C. Loy, and X. Tang., 2014. Facial landmark detection by deep multi-task learning.In European Conference on Computer Vision, pages 94-108.
[84] A. Schwarz, M. Haurilet, M. Martinez, and R. Stiefelhagen., 2017. DriveAHead A Large-Scale Driver Head Pose Dataset. CVPRW.
[85] Y. Cao and Y. Liu., 2017. Head Pose Estimation Algorithm Based on Deep Learning. Materials Sciences, Energy Technology, and Power Engineering I, AIP Conf. Proc. 1839.
[86] K. Yuen, S. Martin, and M. M. Trivedi., 2016. Looking at Faces in a Vehicle: A Deep CNN Based Approach and Evaluation. ITSC.
[87] K. Krafka et.al., 2016. Eye Tracking for Everyone. CVPR.
[88] D. Hansen and Q. Ji., 2010. In the eye of the beholder: A survey of models for eyes and gaze. IEEE Trans. PAMI, 32(3):478-500.
[89] M. Bohme, A. Meyer, T. Martinetz and E.Barth., 2006. Remote eye tracking: State of the art and directions for future development. In the 2nd Conf. on Communication by Gaze Interaction - COGAIN,Turin, Italy, pp. 10-15.
[90] Z. Orman, A. Battal and E. Kemer., 2011. A Study on Face, Eye Detection and Gaze Estimation", Intl. Journal of Computer Science and Engineering Survey, 2(3): 29-46.
[91] http://www.eyegaze.com/ - accesed on June 4, 2018
[92] T. Ohno., 2006. One-Point Calibration Gaze Tracking Method. Proc. Symp. Eye Tracking Research and Applications, pp. 34-34.
[93] C.H. Morimoto, A. Amir, and M. Flickner., 2002. Detecting eye position and gaze from a single camera and 2 light sources. In Proc. of the International Conference on Pattern Recognition.
[94] F. L. Coutinho and C. H. Morimoto., 2012. Improving head movement tolerance of cross-ratio based eye trackers. International Journal of Computer Vision.
[95] K. Zhang, X. Zhao, Z. Ma and Y. Man., 2010. A simplified 3D gaze tracking technology with stereo vision. International Conference on Optoelectronics and Image Processing, pp. 131-134.
[96] R. M. Hung and L. Yin., 2010. Pointing with the eyes: Gaze estimation using a static/active camera system and 3D iris disk model. IEEE Intl. Conf. on Multimedia and Expo, pp. 280-285.
[97]https://www.fondriest.com/environmentalmeasurements/parameters/weather/photosynthetically-active-radiation/- Accessed on June 4, 2018
[98] https://www.creativeplanetnetwork.com/news-features/dv101-makinginvisible-visible-understanding-infrared-filtration-423128 - Accessed on June4, 2018.
[99] https://spectrum.ieee.org/transportation/advanced-cars/bmw-laserheadlights-slice-through-the-dark - Accessed on June 4, 2018.
[100] https://www.nature.com/articles/s41598-018-23265-x Accessed on June 4, 2018.
[101] http://www.phys.uconn.edu/gibson/Notes/Section5-3/Sec5-3 - Accessed on June 4, 2018.

[102] http://hyperphysics.phy-astr.gsu.edu/hbase/geoopt/lenseq.html - Accessed on June 4, 2018.
[103] https://www.vsp.com/eyeglasses-lenses.html - Accessed on June 4, 2018.