Detection and Controlling of Drivers' Visual Focus of Attention

Partha Pratim Debnath
Lecturer, Dept. of Information and Communication Engineering
Bangladesh Army University of Engineering and Technology
Natore, Bangladesh
parthapratim.ice10@gmail.com

A. F. M. Rashidul Hasan
Associate Professor, Dept. of Information and Communication Engineering
University of Rajshahi
Rajshahi, Bangladesh
mirza_iu@yahoo.com

Dipankar Das
Professor, Dept. of Information and Communication Engineering
University of Rajshahi
Rajshahi, Bangladesh
dipankar_ru@yahoo.com

Abstract*— The paper presents an approach to detect and control the focus of attention of the driver using his/her eye gaze and head movement direction. If the driver changes his/her eye gaze, the corresponding coordinate of the pupil also changes. To determine the eye gaze direction, we first divide the detected eye area into different regions corresponding to different target objects. Then the gaze direction is detected based on in which region the coordinate of pupil is located in the eye area. We also determine and classify the head movement direction into three areas: Central Field of View (CFV), Near Peripheral Field of View (NPFV), and Far Peripheral Field of View (FPFV). Finally, the combination of eye gazes and head movement areas are used to determine the driver's focus of attention. We detect both transient and sustained attention. If the driver changes his/her direction of attention abruptly, in such a situation we cannot always confirm that the driver has a new attention to another direction because it may be a transient attention. To detect a sustained attention we need to wait few moments (in experiment we set it to 8 seconds). We have tried to detect the transient focuses with different duration as well as the sustained focus of attention with optimal accuracy that arises while real driving. If the system detects the sustained focus of attention to other direction than driving, we generate a controlling signal to return his/her focus of attention to driving. We have implemented and tested the proposed system in both controlled and real environment. The experimental results reveal that the proposed system is able to detect and control the driver's focus of attention in real driving situation.

***Keywords-* Visual focus of attention, eye center localization, gaze detection, sustained attention.**

I. INTRODUCTION

Head and eye gaze behavior is a very interesting research topic for many years in psychology and in computer vision community. Especially, in the field of driver's visual focus of attention (VOFA) [1] [2] detection this study has become a part of active research. However, there are very little works by which driver's attention may be controlled accurately through human machine interface (HMI). Most of the existing methods of controlling driver's attention use sensors based head mounted devices to estimate the attention level [3] [4]. Although the sensor-based approaches are accurate, however these approaches require technology applicable to controlled environment. Thus we use vision based gaze fixation technique to estimate the attention of driver that requires a very low cost senor (such as a USB camera) situated in front of a driver (celling of the car). Basically the term "Fixation" refers to a process of maintaining the visual gaze at a single location. Visual fixation is never perfectly steady, fixational eye movement occurs involuntarily. Thus, we need an accurate method for gaze estimation based on object [5] detection. Most of the vision researchers use the bottom-up saliency detection technique [6]. In their approach, visual information such as color, intensity, and orientation from local image characteristics is used. Although there exists a close relation between visual saliency and target object, most of the existing approaches [7] [8] [9] do not consider a computational model to thoroughly explore the two concepts and their mutual effects. Thus, the researcher currently focused on combining head and eye for gaze estimation rather than using the only head movement [10] or only tracking the eye [11] [12]. However, no study is performed on the feasibility of an accurate appearance-only gaze estimator that considers both the head pose and eye location factors. Therefore, our goal is to develop a system capable of analyzing the visual gaze of a person starting from monocular video images. This allows studying the movement of the user's head and eyes in a more natural manner than traditional methods. Moreover, since Bangladesh is a developing country, we aim at building a system that can tract the gaze of driver at a very low cost.

II. VISUAL FOCUS OF ATTENTION DETECTION

A. Method Overview

At first we need a camera or sensor to detect the visual focus of attention of the driver. Then we shall feed the captured data frame by frame to an intelligent system. The system will convert the image into grey scale for easier processing and analyze the data pixel by pixel. These analyzed data will be used to detect the loss of attention if the driver changes his/her visual focus of attention (VFOA) to another direction. How

978-1-5090-5627-9/17/$31.00 ©2017 IEEE

head direction changes at the time of loss of attention for different tasks, will be observed from the captured videos. The minimum deviation of the head orientation can be used as a clue to detect the loss of attention because when loss of attention occurs, drivers change the tilt/pan angle of their head. Based on the head location and its area from a 3-D head tracker [13], we can roughly estimate the eye regions of the face. Then vector field of image gradient (VFIG) method is used to detect the iris center. Now the head position using the active shape model [14] (ASM) and the localized iris center are used together for VFOA estimation using different techniques. Eye gaze tracking provides precise results on the front view; while head pose tracking is more suitable for tracking areas of interest than for tracking points of interest on the side view. Finally, we shall apply the rules to detect the sustained or transient focus of the driver. No transients or too many transients mean inattention of the driver. In this way, if inattention is detected, it is controlled by providing an alarm or external signaling system so that the driver is alerted to pay proper attention to driving.

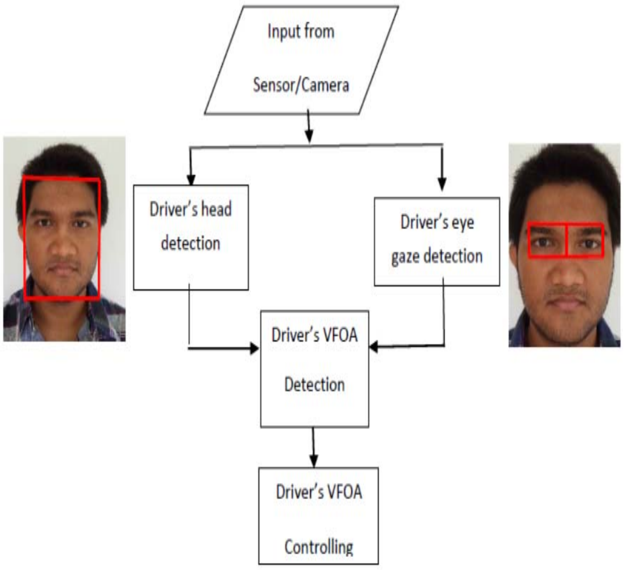

Figure 1: Overview of the proposed system

B. Field of View of the driver

The field of view (FOV) of the driver is divided into three regions-

- *Central Field of View (CFV):*

This FOV exists at the center of the human FOV. This zone is set to a 30° cone shaped area (75° to 105°).

- *Near Peripheral Field of View (NPFV):*

It is defined as the 45° fan shaped area on the both sides of CFV zones. At the right side of CFV (30° to 75°) it is defined as the right near peripheral field of view (RNPFV) and at the left right side of CFV (105° to 150°) it is defined as the left near peripheral field of view (LNPFV).

- *Far Peripheral Field of View (FPFO):*

This FOV exists on both sides at the edge of the human FOV. The right side of the RNPFV (-35° to 30°) is known as the right far peripheral field of view (RFPFV) and on the left side (-145° to 150°) is known as left far peripheral field of view (LFPFV)

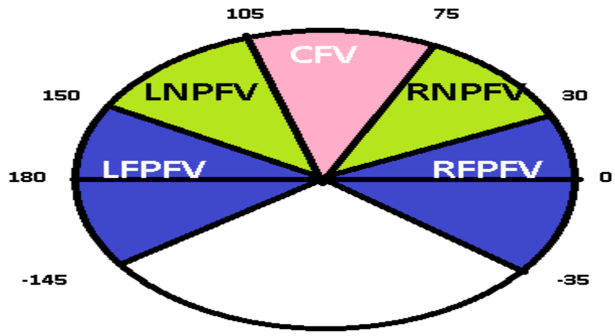

Figure 2: Field of View (FOV) of drivers

The drivers sitting in vehicle does not move his head if the focus is in CFV, LNPFV or RNPFV. So if the focus of attention of the drivers is in these three regions, we shall track the pupil to get the gaze detection without considering the head movement. However, while looking on the back view mirror and in the case of abrupt transients, the focus is in LFPFV and RFPFV. In this case we must have to consider the head movement for proper focus of attention detection.

C. Gaze Detection Using the Coordinate of the Pupil

If we can track the pupil of the eye efficiently, then from the variation of the coordinate of pupil of the driver we can detect on which direction he is focusing. From the face points eye rectangle is created around the eye region. For simplicity, we shall consider the coordinate of only left eye because we know that the pupil of the two eyes move simultaneously. The width of eye rectangle is divided into three regions: left, monitor and right. If the driver changes his focus from central field of view (CFV) to any of the near peripheral field of view (NPFV), the corresponding x coordinate of pupil also changes. Within the eye rectangle we shall detect on which region the x coordinate of eye falls and thus detect the gaze of the driver. The followings are different approaches to define the threshold for Left, Monitor and Right regions:

1) Equal Division of Eye Rectangle Width

In this approach the width of the eye rectangle is equally divided among three regions. For example, if the width of the eye rectangle is 60 pixel and the x coordinate of the left eye rectangle starts from 300 and ends at 360, the threshold of three regions will be defined as follows-

Left = {X: X < 320 and X > 300},

Monitor = {X: X < = 340 and X > = 320},

Right = {X: X < 360 and X > 340}

But this technique is not very efficient and results in a lot of error. Because there are many point in the eye rectangle not actually covered by the eye (showed in following figure 2.5). If we consider these coordinates where the x coordinate of pupil never goes, will produce a lot of error.

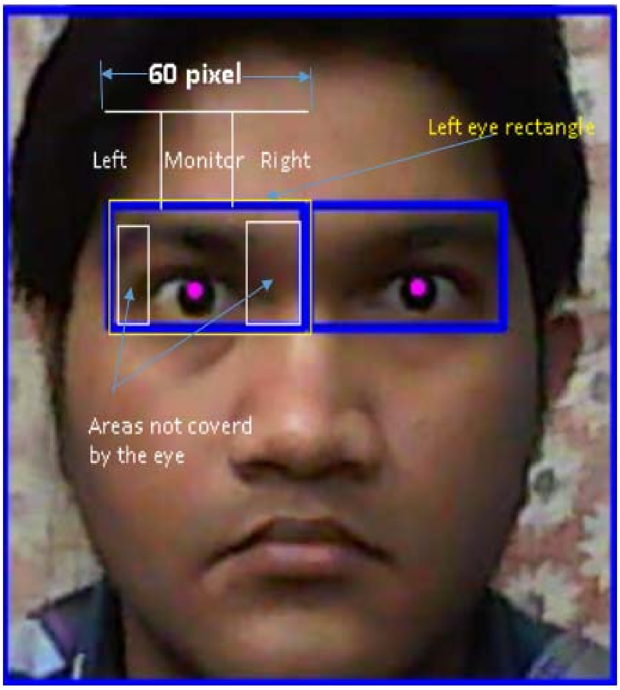

Figure 3: Problem of equal division eye rectangle

We see from the above figure that most of the eye pixels are covered by the monitor region, left very little amount of pixels for left and right region. So it can detect the monitor gaze, but not the left and the right

.

2) Partitioning Eye Rectangle Leaving the Uncovered Regions

In this approach the total width of the rectangle is divided into five equal regions: Left uncovered, Left, Monitor, Right and Right uncovered. Let us again consider that the left eye rectangle starts from X = 300. For 60 pixel width of eye rectangle, the threshold of these regions are defined as follows-

Left uncovered = {X: X > 300 and X < 312},

Left = {X: X >= 312 and X < 324},

Monitor = {X: X >= 324 and X <= 336},

Right= {X: X > 336 and X <= 348},

Right uncovered = {X: X > 348 and X < 360}

From figure 4, we see that the monitor region has been reduced and the left and the right region cover most of the left and right portion of the eye respectively. The "left uncovered" and "right uncovered" regions do not cover any portion of the actual eye. So while gaze estimation, these two areas are discarded. But in this technique the slight change in focus from Central Field of View (CFV) to the left or right cannot be detected very efficiently. Because a slight change in focus means a slight change in x coordinate of pupil and hence cannot be detected by the left or right region.

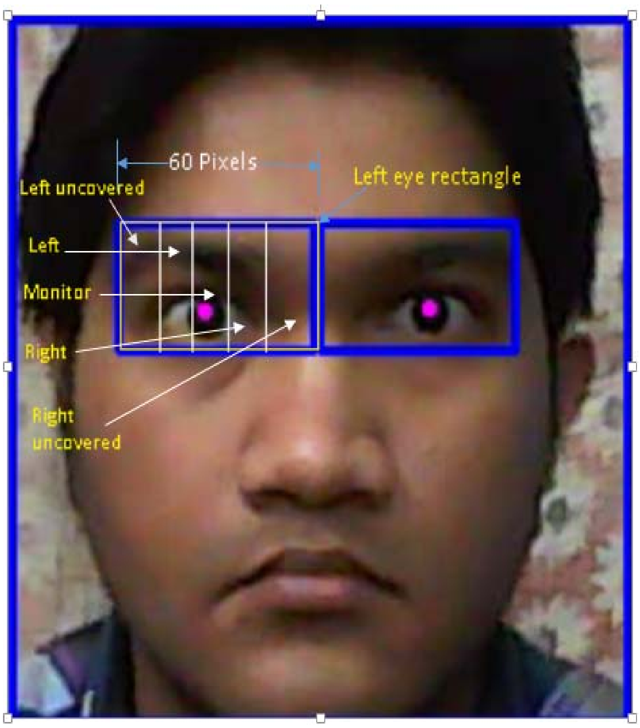

Figure 4: Partitioning eye rectangle leaving the uncovered regions

3) Unequal Partitioning of Eye Rectangle Leaving the Uncovered Regions

In this technique we leave the uncovered regions but the width of the left, monitor and right regions are not the same as we discussed before. To detect the slight change of focus from Central Field of View (CFV) to the left or right, the width of the monitor region is decreased and hence the width of the left as well as the right regions increases so that they can accommodate the slight change of focus in Near Peripheral Field of View (NPFV) and enhance the total accuracy. The threshold detection for left, monitor and right regions are discussed below-

Let the width of the left eye rectangle be 60 pixels and it is divided into five equal parts as before. A rough threshold for the five regions are defined as follows-

Left uncovered1 = {X: X > 300 and X < 312}

Left1 = {X: X >= 312 and X < 324}

Monitor 1= {X: X >= 324 and X <= 336}

Right1= {X: X > 336 and X <= 348}

Right uncovered 1= {X: X > 348 and X < 360}

Now the length of the monitor region is divided into 4 segments. If the length of the Monitor region is 12 pixels, the length of each of the segment, d = 12 / 4 = 3. Now the new threshold is estimated as-

Left uncovered = {X: X > 300 and X < 312}

Left = {X: X >= 312 and X < (324 + d)}

={X: X >= 312 and X < 327}

Monitor = {X: X >= (324 + d) and X <= (336 – d+1)}

={X: X >= 327 and X <= 334}

Right= {X: X > (336-d+1) and X <= (348+d-1)}

={X: X >= 334 and X <= 350}

Right uncovered = {X: X > (348+d-1) and X < 360}

={X: X >= 350 and X < 360}

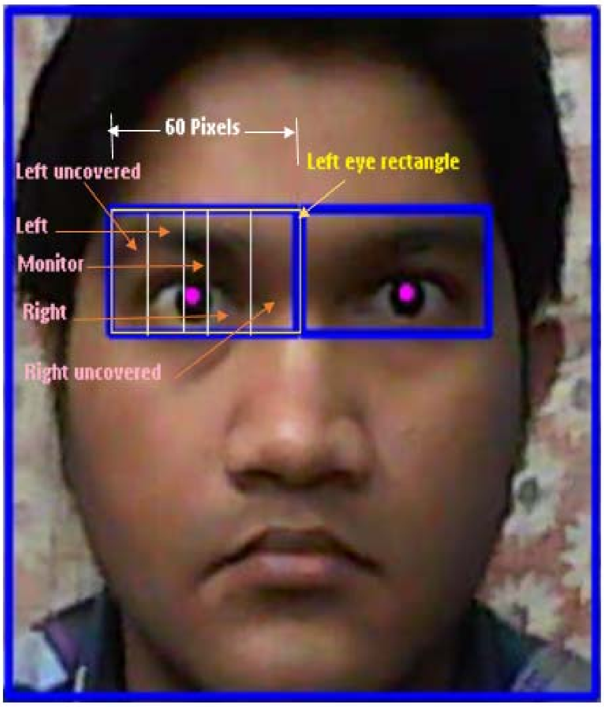

Figure 5: Unequal Partitioning of eye rectangle leaving the uncovered regions

D. *Gaze Detection Using the Head Rectangle*

If the driver's gaze is in far peripheral field of view (FPFV) or near this zone, head movement occurs. In this case we cannot detect the eyes for gaze estimation very accurately. So depending on the head movement, we decide on which object the driver is looking. To accomplish this, the width of the head rectangle is divided into three regions: Front, Left side and Right side. From the detected face points, the middle of the face is detected and it is basically the middle of the eye rectangle. Now if we move our head, this middle line of our face also moves with respect to the head rectangle. We decide whether the driver is looking on left, front or right based on in which head region the middle line of the face falls. For example let us consider that the width of the head rectangle is 180 pixels. It is divided into five equal segments. The length of each segment, s = 180 / 5 = 36 pixels. If the x coordinate of the head rectangle starts from 300, the threshold values for the three head regions are defined as follows-

Left side = {X: X > 300 and X < (300 + 2 * s)}

= {X: X > 300 and X < 372}

Front = {X: X >= (300 + 2*s) and X <= (300 + 3* s)}

= {X: X > 372 and X < 408}

Right side = {X: X > 408 and X < (300 + 5 * s)}

= {X: X > 408 and X < 480}

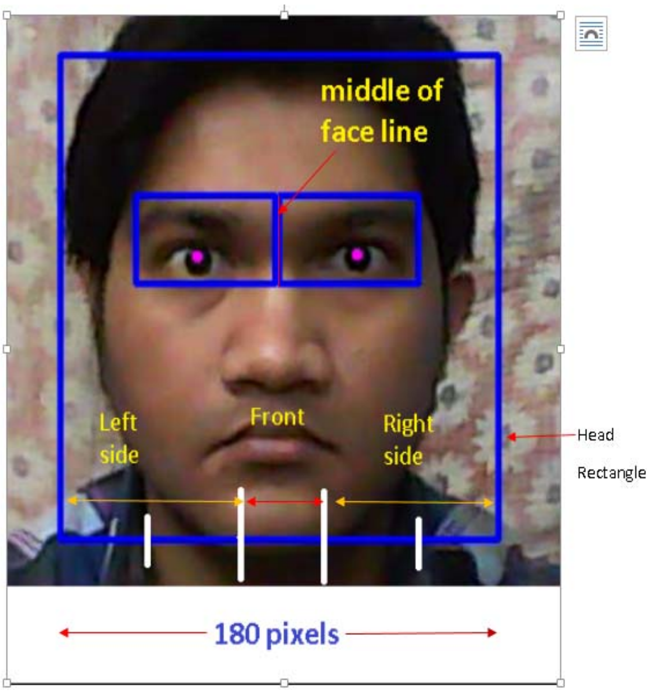

Figure 6: Gaze detection using the head rectangle

E. *Sustained and Transient Focus of Attention Detection*

Selective sustained attention, also known as focused attention, is the level of attention that produces the consistent results on a task over time. However, the time span of sustained attention varies with different type of tasks. On the other hand, transient attention is a short-term response to a stimulus that temporarily attracts/distracts attention. The minimum time for sustained detection is 8 seconds [15], because the maximum time for transient attention is 8 seconds. Detection of sustained and transients are very important to ensure a safe driving. For a given time span while driving, if we can detect that the transient attention of the driver are occurring very frequently, we can ensure the inattention of the driver and give an alarm to alert the driver. To ensure a safe driving most of the attention must be sustained. No transients also mean inattention. Because while driving a driver need to look at the back view mirrors at least twice per minute which are also transient attention. For one-minute time span during driving, we can ensure a safe driving if it belongs more than two transients. The threshold values of time for transient and sustained focus of attention detection are given as follows-

Transient = {T: T > 0 and T < 3}

Sustained = {T: T > 8}

III. DATA COLLECTION

A. Participants

There were total 3 male nonpaying participants with the age 25, 24 and 22 years respectively. For data collection of real time driving, the age of the participant was 40 years.

B. Procedure of Gaze Detection Based on Object Detection

In head and eye pupil tracking experiments, the participants were asked to seat at different distances from the camera under different lighting conditions. They sat at 30cm, 40cm, 50cm, 60cm, 70cm, 80cm, 90cm and 100cm away from the camera and look at different objects to create sustained and transient focus of attention. These objects are situated 0.5m apart from each other. One of the three objects is located in the central field of view (CFV) and the other two in the near peripheral field of view (NPFV).Different lighting conditions were provided by deploying different number of (1, 2, 3 and 4) 32 Watt energy bulbs. The area of our room was $4m^2$. The average video length was 2 minutes.

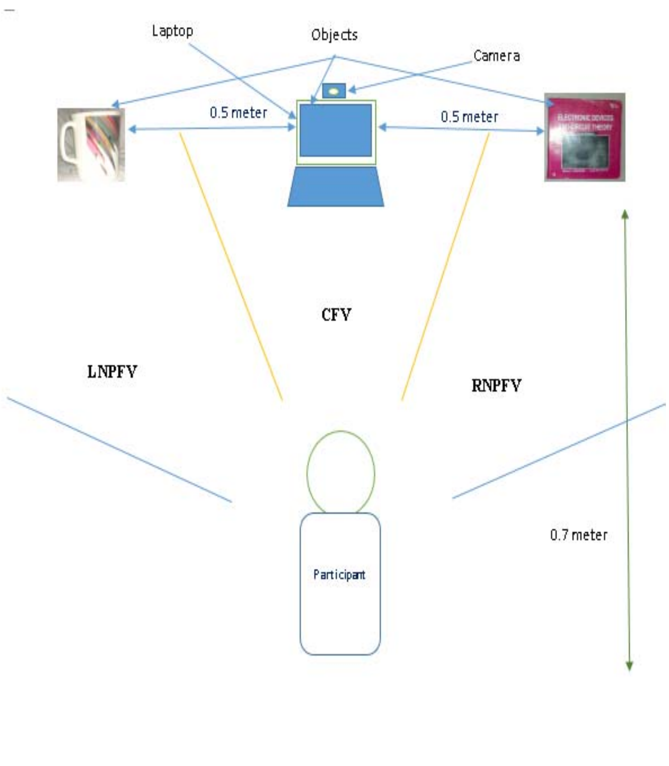

Figure 7: Gaze detection based on object detection

C. Procedure of Real Time Data Collection

We videotaped 3 experiments while driving a car. In the first experiment the participant (age 40 years) was asked to drive the car slowly (35 km/h). In the second experiment the driver (the participant) drove the car comparatively fast (40km/h) through a busy road. In the final experiment the driver drove at the speed of 55 km/h. The distance between the driver and the camera was 0.7 m. Since the day was sunny, proper illumination was provided. The length of recorded video was 3 minutes 33 seconds, 4 minutes 20 seconds and 6 minutes 17 seconds respectively. The frame rate was 30 fps (Frame per second).

Figure 8: Initial setup for real time experiments

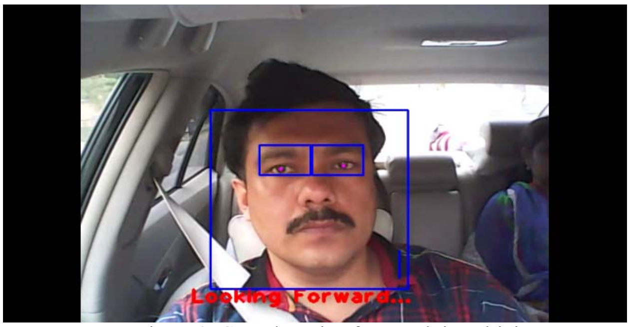

Figure 9: Gaze detection from real time driving

IV. EXPERIMENTAL RESULTS

Following figure shows the gaze detection accuracy based on object detection. To get the best performance the distance between the camera and the participant was kept 70cm and adequate illumination was provided.

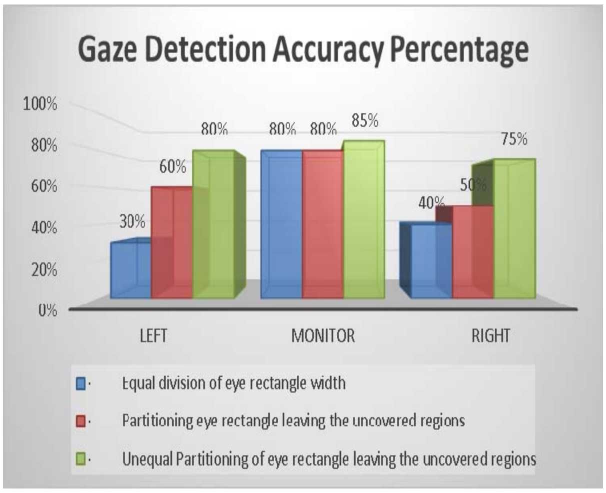

Figure 10: Gaze detection of the driver based on object detection

We conducted 3 separate experiments with varying duration and varying the speed of the car. The accuracy of head movement detection is defined as follows-

Accuracy= (Average number of head movement detection at a particular direction in 30 seconds) / (Average number of head movement occurred at that direction in 30 seconds) × 100%

The accuracy of transient focus of attention detection is defined as follows-

Accuracy= Average number of transients detection at a particular direction in 30 seconds) / (Average number of transients occurred at that direction in 30 seconds) × 100%

Figure 11 shows the accuracy of head movement detection and figure 12 shows the accuracy for transient focus of attention detection respectively.

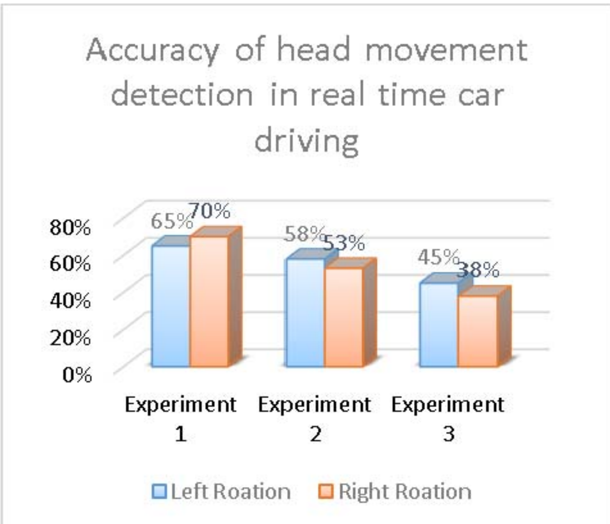

Figure 11: Accuracy of head movement detection in real time car driving

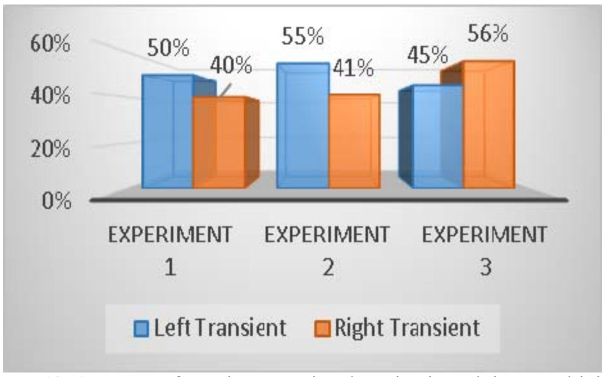

Figure 12: Accuracy of transient attention detection in real time car driving

V. CONCLUSION

For gaze detection from the coordinate of pupil, we applied different techniques. However, the method of "Unequal Partitioning of Eye Rectangle Leaving the Uncovered Regions" shows the best performance in tracking the gaze. In gaze detection combining with head pose we conducted different experiments with variable head rotation time. It is seen that, if the head rotation time is 3 seconds then the gaze tracking performance is the best. In the performance evaluation of real time driving, we see that the accuracy of gaze detection reduces. The accuracy of head detection depends on the speed of the car. For slower speed of the car, the accuracy of head detection increases. The reason behind this fact is that if we increase the speed of the car, the car shakes more. Moreover, the road is not smooth everywhere. The accuracy of transient attention detection is relatively poor. It is because for transient attention detection, the head must be stable. But the head and the eye move very frequently while real time driving and as a result we cannot track the transients very efficiently. To control the attention of the driver, an external alarming system was deployed. In our system, we provided with a beep sound if transient focus of attention are detected. Thus we alerted the driver successfully, if inattention is tracked.

ACKNOWLEDGEMENT

The authors would like to thank their colleagues and all the lab assistants for their cooperation.

REFERENCES

[1] Judd T., Ehinger K., Durand F., and Torralba A., "Learning to Predict Where Humans Look," in IEEE 12th international conference on Computer Vision, pp. 2106-2113, IEEE, September 2009.

.

[2] Cal H., Lin Y., and Mourant R., "Evaluation of Drivers' Visual Behavior and Road Signs in Virtual Environment," in proceeding of HFES 51st annual meeting, Baltmore, vol. 5, USA, pp. 1645–1649, 2007.

[3] Zhang H., Smith, M. R., and Witt, G. J, "Identification of Real-Time Diagnostic Measures of Visual Distraction with An Automatic Eye-Tracking System," Human Factors: The Journal of the Human Factors and Ergonomics Society, 48(4), pp. 805-821, 2006.

[4] Goferman S., Zelnik-Manor L., and Tal A., “Context-Aware Saliency Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence,” 34(10), pp. 1915-1926, ctors 48(4): 805–821, 2006.

[5] Alexe B., Deselaers T., and Ferrari, “What is an object,” V. 2010. CVPR, pp. 73-80.

[6] Hou X., and Zhang L, “Saliency detection: A Spectral Residual Approach,” CVPR'07, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8. IEEE, June 2007.

[7] Koch C., and Ullman S., “Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry,” in Matters of intelligence, pp. 115-141.

[8] Itti L., Koch C., and Niebur E., “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” IEEE Transactions on Pattern Analysis & Machine Intelligence (11), pp. 1254-1259, 1998.

[9] Harel J., Koch C., and Perona P, “Graph-Based Visual Saliency,” in Advances in neural information processing systems, pp. 545 -552, 2006.

[10] Murphy-Chutorian, E., & Trivedi, M. M. (2009). Head Pose Estimation in Computer Vision: A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(4), pp. 607- 626.

[11] Kim S., Chung S. T., Jung S., Kim, J., and Cho S. “Multi-Scale Gabor Feature Based Eye Localization,” World Academy of Science, Engineering and Technology 21, pp. 483-487, 2007.

[12] Kroon B., Hanjalic A., and Maas S. M., “Eye Localization for Face Matching: Is It Always Useful and Under What Conditions?” in Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 379-388, ACM, July 2008.

[13] An K. H., and Chung, M. J., “3d Head Tracking and Pose-Robust 2d Texture Map-Based Face Recognition Using A Simple Ellipsoid Model,” in 2008, IROS 2008, IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 307-312, IEEE, September 2008.

[14] Cootes T. F., Taylor C. J., Cooper D. H., and Graham J., “Active Shape Models-Their Training and Application,” in Computer vision and image understanding 61(1), pp. 38-59, 1995.

[15] David Cornish M., and Dukette D, “The essential 20: twenty components of an excellent Health Care Team,” Dorrance Publishing.