Assessing Driver Hazard Attention
Toward an Intelligent, Artificial Co-Driver

1Siti Nor Hafizah binti Mohd Zaid, 2Abdel-Hamid Soliman
Faculty of Computing, Engineering and Sciences, Staffordshire University
Stafford, United Kingdom

Abstract- This paper introduces an eye-gesture template approach for assessing driver hazard observations. The approach divides the vehicle windscreen into 24 uniform cells and matches the driver’s eye with eye templates drawn from the driver looking at the centre of each windscreen cell under natural head movements so as to estimate the driver’s point of regard on the windscreen. The windscreen cells are then mapped onto the video image captured from a road-facing scene camera producing 24 corresponding mapped cells on the camera image. Once a hazard is detected within a mapped cell in the camera image, a template matching result showing that the driver is looking at the corresponding windscreen cell allows the system to assume an increased probability that the driver has seen the hazard. Experimental results have shown that the template matching algorithm correctly identifies if the driver is looking at a particular windscreen cell with an 88.3% success rate. Analysis of the results has also indicated algorithm ways to improve the detection accuracy using an extension of the eye-gesture template approach.

Keywords- advanced driver assistance systems, adas, driver attention, driver hazard awareness, driver eye-tracking

I. INTRODUCTION

Road traffic accidents are a major cause of death with over one million people losing their lives and a further fifty million seriously injured each year worldwide. However, recent research into the causation of road accidents have found that momentary lack of attention featured in as much as 78% of road accidents (Neale et al, 2005). Some researchers claim that lack of attention is the main cause of accidents as factors such as fatigue, alcohol or drug use, distraction and speeding all impair the driver’s capacity to pay attention to the vehicle and road conditions[7]. These factors have motivated research efforts that aim to improve driver performance and thus help to reduce accidents. These research efforts has led to the development of Advanced Driver Assistance Systems (ADAS). ADAS systems are on-board computer systems that attempt to reduce the risk of accidents by monitoring the driver, vehicle and environmental conditions and taking some action when a risk is identified.

However, there is comparatively little published work that tackles the problem of driver attention with much of the work focusing upon detecting and responding to vehicle and environmental state. For this reason, this paper presents continued work on driver attention monitoring by measuring driver point of regard on potential roadside hazards. The aim is to link driver eye-tracking techniques to environmental sensors able to detect and locate potential roadside hazards so as to determine the level of the driver hazard awareness. The main idea underpinning this work is the assumption that a roadside hazard becomes less of a hazard if the driver is giving it due visual attention.

Although, much work has been done on gaze estimation techniques and object detecting and tracking using sensors, there is still much scope for deriving algorithmic techniques optimized for use within vehicles. For this reason, this paper presents work toward an efficient and robust hazard awareness detection system.

II. LITERATURE REVIEW

Recently research efforts targeted at developing ADAS systems have focused on three main areas: Driver State, Vehicle State and Environmental State. For this reason, the following review is broadly based upon these categories.

A. Driver State Approach

Some research efforts aim to develop ADAS systems based on the driver’s state. These approaches focus on using sensors to detect the driver’s behaviour while driving. Some research efforts attempted to detect driver state from cues obtained from facial features, eye movement and head pose [8]. These parameters are used to identify if the driver is drowsy or not paying attention while executing any driving tasks. Some researchers have considered physiological state by measuring the driver heart rate using electrocardiography (ECG). However this technique is intrusive and uncomfortable for the driver as they have to attach electrodes to their body especially in the facial area.

Smith et al. [21] have analyzed human driver alertness based upon the driver’s head pose using facial feature detection, the eye blinking rate and eye gaze to help identify the driver’s level of alertness. Head pose was tracked using parameters based upon detecting the corners of the lips, the center of the eyes and the side of the face. They use a color predication technique to find the features of the face in color images that was originally proposed by Kjedlsen et al. [11]. These features were used to track the head pose. The authors detect prolonged rotation of the head (10 frames out of the last 20 frames) as a sign of low vigilance to the road and they detect high eye closure rates (eyes closed in more than 40 frames out of the last 60 frames) as a sign of low visual attention. This was achieved using a single camera located on the dashboard.
which was used to detect head rotation and blinking rate. The results have shown that the system is able to determine if the driver is not paying attention to an object on the road, for example if he/she has been looking away for a long time. The system is also capable of detecting whether the driver has seen the left/right blind spots, looking at rear view mirror, checking side mirrors, looking at the speedometer as well as looking ahead for an appropriate time.

B. Environment State Approach

The environmental state approach is based upon determining the driver’s state corresponding to any autonomous condition within the environment. For instance, a potential hazard (such as a pedestrian crossing the road), a road sign or any other environmental factor which influences driver state. Doshi and Trivedi et al. [5] have developed a system which correlates eye gaze and head postetopredict driver intent to perform lane changes. To monitor eye gaze, they position a monocular camera in the middle of the dashboard trained at the driver’s face. Due to the difficulty of accurately assessing eye gaze, video images are manually processed to obtain data relating to changes in gaze direction. Head motion was estimated using optical flow and block matching techniques whereby optical flow vectors are calculated for different face regions over each frame in the time window being considered (one or two second moving windows). The vectors are then input as features into a classifier aimed at capturing rapid head movements. They track lane changes using the VioLET lane tracker proposed by McCall and Trivedi [13]. They utilise the ‘Driver Intent Inference System’ proposed by McCall et al [14] to make predictions about driver intent based upon selected cues. Results from this study showed that driver head motion when combined with lane position and vehicle dynamics is a reliable cue for lane change intent. They say that the use of eye gaze changes is not as effective as head movements when predicting lane changes as head movements tend to occur before changes in eye gaze.

C. Vehicle State Approach

Much work has been done on detecting vehicle state and environment state within ADASresearch. This work has focused on vehicle assessment based on sensors like GPS, or accelerometers. Some researchers have used image processing techniques to detect lane curvature and sensors to measure the distance between cars or the angle of steering wheel using an potentiometer in order to determine if the vehicle is within the lane.

D. Driver’s, Vehicle and Environment’s State

Some researchers have fused factors relating to driver, vehicle and environmental state to develop ADAS systems. Petersson et al [20] have used driver’s eye gaze as cues for driver intent to change lane and monitor whether the driver is paying full attention to speed signs before executing any driving task. The Face lab eye and head tracking system from Seeing Machines is used to track the vehicle driver’s eyes and sensors are used to monitor the vehicle state: a three axis accelerometer (gas pedal detection), three axis gyro-meter (lane change detection), GPS (to monitor the position of the car) and a potentiometer to measure the angle of steering wheel. They use a camera oriented to capture the driver’s face and another camera facing the road to detect and classify speed signs and obstacles on the road. These factors were then used to identify if the driver is not responding to a road sign or hazard (obstacle) on the road. Atouch screen monitor was used to alert the driver if a hazard has been identified and an audio signal is given if the system detects that the driver is driving too fast or too slow for the road conditions based upon the detection of roadside speed signs.

E. Hazard Awareness Review

Within the driver attention approaches in ADAS, there is a subcategory that focuses on hazard awareness which aims to assess if the driver has perceived a hazard on the road. According to Nagayama [12], more than 50% of all collisions in road traffic arise from a missing or delayed hazard perception. In a study of road safety, Treat et al (1977) found that human error was the sole cause in 57% of all accidents and was a contributing factor in over 90% of all accidents. Poor hazard awareness is one such driver error.

Velichkovsky et al. [22] considered fixation duration and the amplitude of related saccades in order to determine if a vehicle driver has observed a hazard on the road. The authors employed a driving simulator called SIRCA to obtain the hazard image such as a pedestrians crossing the street. The test subject’s eye gaze was recorded during the driving simulation using the Eye-Link head-mounted eye-tracking system. The simulation environment was a built up area with a 50km/h speed limit or less.

The work presented in this paper is also focused on driver hazard awareness and is an extension of our earlier work on the Non-intrusive Intelligent Driver Assistance and Safety System (Ni-DASS) [15]-[17] which focused on the eye-gesture template approach for low resolution driver point of regard determinations and under natural head movement [18].

III. HAZARD ATTENTION SYSTEM

The proposed hazard awareness system involves dividing the windscreen into 24 uniform cells consisting of four rows of six columns with each cell approximately 22cm wide and 14cm high. Figure 1 shows the test car with the windscreen cells mapped out with masking tape. A camera is positioned on the dashboard and oriented to capture the driver’s face. Eye-gesture templates are captured from the driver camera as the driver rotates his/her head to face each windscreen cell in turn and then fixes on the center of each cell while keeping his/her head pose fixed. This has the effect of producing 24 head rotations with each rotation having 24 eye-gesture templates with each template comprising an image of the driver’s left eye looking at the center one of the windscreen cells. For each head rotation, the head pose is estimated in terms of the three Euler angles.
of yaw, pitch and roll using the 3D approach proposed by Gee [9].

A scene camera is attached to the windscreen and aligned to capture the road in front of the car. The windscreen cells are mapped onto the scene camera image plane by asking the driver to sit in the normal driving position and look through the corners of each windscreen cell onto a scene gaze point. This projection of the gaze vector through the corner of a windscreen cell onto a gaze point within the scene is mapped onto the camera image plane by determining the co-ordinates of the gaze point in the camera image. This process is repeated for all 35 windscreen cell corner points so as to determine the projection of all 24 windscreen cells onto the camera image. In this way, if the driver is looking through windscreen cell (2, 3) (e.g. row 2 and column 3) then this will be mapped onto a corresponding mapped cell (2, 3) in the camera image.

Once the image coordinates of the cell corners in the camera image co-ordinate system were recorded, vertical lines of best fit were plotted through the row co-ordinates producing a mapping of windscreen cells through the driver’s line of sight onto the screen camera image (Fig. 6).

When performing the cell mapping, the driver was allowed to make natural head movements while looking through the cell corner points. For instance, the driver was allowed to rotate her head in a natural fashion. When the driver fixates on a roadside hazard, the co-ordinates of the intersection of the driver’s gaze vector with the corresponding windscreen cell is assumed to be invariant to moderate head movements. Although this is not true in actuality, this approximation is possible due to the fact that the displacement of the intersection of the gaze vector on the windscreen resulting from head movements is proportional to (but less than) the displacement of the eye in Cartesian space. Figure 2 illustrates this by showing the position for the driver’s left eye located at co-ordinates \((x_l, y_l, z_l)\) and the displacement of the left eye resulting from a horizontal head movement of \(\Delta x\) to co-ordinates \((x_l + \Delta x, y_l, z_l)\).

The arrows represent the gaze vector from the two eye positions to a roadside hazard (star). The distance of the windscreen from the driver’s left eye is represented by \(d\). The displacement of the intersection of the gaze vector with the windscreen resulting from the change in eye position is represented by \(\Delta w\). The value of \(\Delta w\) will depend upon the co-ordinates of the hazard \((x, y, z)\), the co-ordinates of the left eye \((x_l, y_l, z_l)\), the size of the eye displacement \(\Delta x\) and the distance of the windscreen from the left eye \(d\).

A little thought will reveal that, as these parameters change, the value of \(\Delta w\) will always be some fraction of \(\Delta x\). As the value of the left eye displacements \(\Delta x\) resulting from horizontal head movements while driving is normally in the order of a few centimeters, the value of \(\Delta w\) will always be less than this. As a result, the small head movements will have a limited effect on the efficacy of the eye-gesture template eye-tracking algorithm. The same reasoning is true for vertical displacements along the y-axis and forward and backward movements along the z-axis.

![Fig. 1: Windscreen cells comprising 4 rows of 6 columns marked out with masking tape.](image)

With 24 cells arranged in 4 rows of six columns, there are 35 windscreen cell corner points. However, due to the position of the camera and driver and the camera’s 140 degree field of view, the projection of the leftmost and rightmost cell corners onto scene points lay outside the camera field of view.

![Fig. 2: Displacement (\(\Delta w\)) of the intersection of the driver’s gaze vector through a point on the windscreen toward a hazard located at position \((x, y, z)\) when the driver’s eye moves by \(\Delta x\) from position \((x_1, y_1, z_1)\) to position \((x_2, y_1, z_1)\) where \(d\) is the distance of the windscreen from the driver’s eye.](image)

IV. EXPERIMENTAL PROCEDURE

The aim of the experiment is to use eye–gesture templates to determine the driver’s point of regard on the windscreen of a Hyundai Matrix 1.6cc car. The driver’s point of regard is then matched with the projection of a hazard in the scene camera image onto the mapped windscreen cells. In this way, if a hazard is detected within mapped cell \((i, j)\) in the scene camera image and the driver is looking at windscreen cell \((i, j)\) then there is an increased probability that the driver will have seen the hazard because his/her eye gaze is in the vicinity of the hazard.
The scene camera, an R300 Synchronous Recording Vehicle Mounted DVR, was mounted on the windscreen 10cm above the center of the dashboard and 8cm to the left of the center of the dashboard to capture the road image. Another camera, a Thorlab’s DCC1545M CMOS camera was positioned 5cm to the left of the center of the dashboard and orientated to capture the driver’s upper body and head.

To create the eye-gesture templates, the driver rotated her head to look at the center of each windscreen cell in turn. Each time the driver rotated her head to face a windscreen cell, she looked at each of the 24 windscreen cells while keeping her head pose fixed.

Eye-gesture templates were captured of the driver’s left eye as she looked at each cell. The eye templates were cropped to contain a small amount of skin surrounding the eye.

The process was repeated for all 24 head rotations. With 24 windscreen cells, there were 24 corresponding head rotations with each head rotation having 24 eye-gesture templates giving a total of 576 templates.

The experiment was conducted in daylight conditions in UK summertime where ambient lighting conditions were dull. The eye-tracking procedure was performed without distraction to the driver with a passenger making the video recordings and operating the eye-tracker software on a laptop computer.

The scene camera was used to capture a video of a pedestrian hazard in 20 different positions in front of the car at a distance of less than 20 meters. The pedestrian hazard held a bright green fixation object in his hand. At each hazard position, the driver fixated on the green fixation object through a windscreen cell for a period of 20 seconds.

The co-ordinates of this windscreen cell were then recorded. The location of the hazard in the scene video was also tracked so as to create a record of the location of the green fixation object within the mapped windscreen cells in the scene camera image.

The driver’s head pose was estimated in terms of the three Euler angles of yaw, pitch and roll based upon the a heuristic model for the forward facing face proposed by Gee [9].

To achieve this, rectangular face feature templates were captured of the driver’s eyes, nose and mouth when the driver was looking left, right and forward.

These facial feature templates were matched in real-time with the driver video using the MatchTemplate function within OpenCV 4.1. Once detected, the position of eyes, nose and mouth was assumed to be the center of the matched template rectangle. These center points were then used as the feature co-ordinates within the head tracking algorithm [9].

In each frame of the driver video, an eye-gesture template set was selected from the 24 sets available based upon the closest matching yaw and pitch values stored with each template set with those recorded in real-time form the driver video. To do this, all templates sets with a yaw value within a +5 or −5 degree range from the yaw value recorded from the driver video were selected and, from those sets, the set with the smallest sum of absolute differences between template yaw and pitch values and driver yaw and pitch values was selected for template matching.

During this selection process the match between yaw values was given precedence over the match between pitch values because yaw rotation (left-right) is typically more pronounced than pitch rotation (up-down) during natural head movements in normal driving conditions. Roll rotation was not used as head tilting rotations are not as common as yaw and pitch rotations during routine driving.

Once the correct template set was selected, the 24 eye-gesture templates within the template set were matched with the driver’s left eye and the highest matching template was selected. During matching, a region of interest was used for the left eye based upon the position of the nose using equation 1 to 4 below:-

\[
\text{Eyes.X} = \text{Nose.X} - 2 \times \text{Nose.Width} \\
\text{Eyes.Y} = \text{Nose.Y} - 2 \times \text{Nose.Height} \\
\text{Eyes.Width} = \text{Nose.X} + 3 \times \text{Nose.Width} - \text{Eyes.X} \\
\text{Eyes.Height} = \text{Nose.Y} + \text{Nose.Height} - \text{Eyes.Y}
\]

The highest matching template from the selected template set was recorded as an indicator of the driver’s point of regard on the windscreen. For example, if the driver’s left eye was matched with eye gesture template \((y, x)\), then it was assumed that the driver was looking at windscreen cell \((y, x)\) where \(x\) is the windscreen column and \(y\) is the windscreen row.

V. RESULTS

Figure 3 below shows the Receiver Operating Characteristic (ROC) curve for the simple matching case where a positive test result is indicated by the gesture template algorithm correctly identifying that the driver is looking at the windscreen cell containing the green fixation object. The Youden index \(J\) [23] was calculated to estimate the optimal criterion value (in this case the template matching threshold) where the index \(J\) is defined as:

\[
J = \max \left\{ \text{sensitivity} c + \text{specificity} c - 1 \right\}
\]

where \(c\) ranges over all possible criterion values.

The result was 0.82 for the template matching threshold which gives a sensitivity of 0.883 and a specificity of 0.448.

Interpreting these results, it is possible to say that if the driver is looking at a cell, the eye-gesture template
approach is able to correctly identify that the driver is looking at this cell with an 88.3% success rate. If the driver is not looking at a cell then the eye-gesture template approach is able to determine that the driver is not looking at this cell with a 44.8% success rate.

Fig. 3: ROC curve showing results of the gesture template matching for windscreen cells.

In order to understand why the specificity is low, it should be borne in mind that a number of factors contribute to an increased false positive result. Of the 736 match results, 223 were negative. Within these negative results there were 129 misclassification for the cell directly above or below the correct cell. There were a further 34 misclassifications for the cell directly to the left or right of the correct cell. There were 60 misclassifications for other cells (including those located within the immediate neighborhood of the correct cell and located toward the top left, top right, bottom left or bottom right of the correct cell). The high occurrence of misclassifications for cells directly above or below the correct cell is primarily due to three factors. Firstly, the height of the windscreen cells is lower than their width leading to a reduced vertical variance in eye appearance when comparing eye patterns for the driver looking at two adjacent cells located within the same windscreen cell column. Secondly, occlusion of the iris by the eye-lids is known to make eye tracking based upon the location of the iris less accurate in the vertical direction than in the horizontal direction.

Thirdly, and importantly, during the experimental procedure the pedestrian hazard often spanned multiple cells within the driver’s field of view which would sometimes lead the driver to observe the green hazard fixation object with a line of sight that lay close to the boundary between windscreen cells.

VI. DISCUSSION

There are several things that can be done to reduce the occurrence of false positives within the eye tracking results. Firstly, it may be desirable to reduce the number of rows of windscreen cell from four to three allowing for the height of the cells to be increased. Alternatively, the height of the top row of cells could be reduced as the driver’s line of sight through the top row often rises above the road scene toward the sky.

The height of the bottom row of cells could also be reduced as the driver’s line of sight through the bottom row of cells partially intersects the car bonnet. The use of an infra-red illuminator coaxial to the driver camera could produce the bright pupil effect where the light reflects off the back of the retina and causes the pupil to be bright white within the camera image.

This would have the advantage of making the pupil distinct from the iris (something that is seldom the case with ambient sunlight). As the eye-lids only partially occludes the iris, introducing the bright pupil effect will likely increase the variance between eye-patterns when the driver is looking at adjacent cells located within the same windscreen cell column. This is true because the pupil is normally not distinct from the iris in visible light spectrum images.

Another approach that could be used is to include additional overlapping boundary cells. This is illustrated in Fig. 4 and Fig. 5 below which show the use of two sets of overlapping windscreen cells. The first set (Fig. 4) contains 4 rows of five cells centered on the horizontal boundary between original neighboring cells. Figure 5 shows the corresponding overlapping cells position to be centered on the vertical boundary between original neighbouring cells.

Using this approach, the driver would rotate her head to face each of the original 24 cells, eye-gesture templates would then be captured of the driver looking at the center of each of the original 24 cells and each of the 20 cells in the horizontal and vertical overlapping cell sets. This would result in 24 sets containing 64 eye-gesture templates. However, as the template set is selected based upon the driver’s head rotation, the runtime requirements would consist of performing template matching using 64 small eye-gesture templates.

Fig. 4: 24 original windscreen cells (gray) overlayed with 20 boundary cells (orange) positioned to be centred on the horizontal boundary of the original windscreen cells.
Fig. 5: 24 original windscreen cells (gray) overlayed with 20 boundary cells (green) positioned to be centred on the vertical boundary of the original windscreen cells.

VII. CONCLUSIONS

Experimental results indicate that the eye-gesture approach is an efficient way of determining coarse driver point of regard with natural head movements for driver hazard awareness determinations. The ROC analysis indicates that the eye-gesture template approach is able to correctly identify that the driver is looking at a given cell with an 88.3% success rate. However, if the driver is not looking at a cell then the eye-gesture template approach is only able to determine that the driver is not looking at this cell with a 44.8% success rate. This relatively low specificity indicates the occurrence of false positives within the eye eye-tracking result. Analysis of the results showed that 17.5% of the matching results were for cells directly above or below the correct cell. However, this is likely to be a consequence of the driver’s eye-gaze being positioned near the boundary between two cells which will reduce the variance between the driver eyepatterns for both cells. Several approaches have been proposed within the discussion to alleviate this problem. Principle among them is to include two sets of overlapping cells so as to reduce the misclassification. Analysis of the efficacy of this approach is left for future work.

Fig. 6: 24 windscreen cells (yellow) overlayed on the scene camera image to show mapping between the driver’s point of view through the windscreen and the scene camera.

REFERENCES


Eye-Gesture Templates using Receiver Operating Characteristic”, World Academy of Science and Technology, 70, 921-927. (2012c)


