Viusal Cues Extraction for Monitoring Driver's Vigilance

Qiang Ji George Bebis
Department of Computer Science
University of Nevada
Reno, NV 89557
E-mail: {qiangji, bebis}@cs.unr.edu

Abstract

Results based on traffic accident analysis indicate that a leading cause for automobile accidents is due to a diminished vigilance on the part of the driver. Two major factors that may result in a reduction of a driver's alertness are drowsiness and intoxication. Developing effective technologies for measuring drowsiness/intoxication is therefore imperative for reducing automobile accidents.

The development of technologies for monitoring driver's vigilance is a major challenge in the field of accident avoidance systems. In order to accurately and robustly characterize a driver's vigilance level, information from different sources (e.g., driver's physiological conditions and physical reactions, sensing of vehicle behavior, response of driver, weather conditions, etc.) needs to be collected and processed. This task poses two major challenges: (i) how to extract appropriate information about the driver's condition in a non-intrusive way, and (ii) how to systematically integrate diverse sources of evidence in a uniform manner so that a consistent overall evaluation of the driver' vigilance can be achieved.

Towards the first challenge, we propose to use non-intrusive techniques based on an analysis of driver's facial images. Towards the second challenge, we propose using Bayesian networks which provide a mathematically coherent and sound basis for systematically aggregating visual evidences from different sources, augmented with relevant contextual information that may lead to a decline in a driver's vigilance level. This report describes our latest efforts in developing techniques for extracting various visual cues typically reflecting the state of vigilance of the driver.

1 Introduction

Many efforts [16] [21] [4] [6] [12] [7] [10] [24] have been reported in the literature for developing active safety systems intended for reducing the number of automobile accidents due to reduced vigilance. Among different techniques, the best detection accuracy is achieved with techniques that measure physiological conditions like brain waves, eye blinking, heart rate, and pulse rate [24][17]. These techniques, however, are intrusive, causing annoyance to drivers. A driver's state of vigilance can also be characterized by the behaviors of the vehicle he/she operates. Vehicle behaviors including speed, lateral position, turning angle, and moving course are good indicators of a driver's alertness level. While these techniques may be implemented non- intrusively, they are, nevertheless, subject to several limitations including the vehicle type and driving conditions [21]. For example, it may be necessary to design a separate logic for each different vehicle type.

To overcome these limitations, we propose to access a driver's vigilance level through visual observation of his/her physical conditions using a camera and state-of-the-art technologies in computer vision. Techniques using computer vision are aimed at extracting visual characteristics from the images of the driver. Several studies have shown the feasibility and promising of this approach [10], [16], [4], [21] For example, study by Ueno et al [21] showed that the performance of their system is comparable to those of techniques using physiological signals.

Despite the success of the existing approaches for extracting characteristics of a driver using computer vision technologies, current efforts in this area, however, focus on using only a single visual cue such as eye movement or line of sight for the characterization of a driver's state of alertness. The system relying on a single visual cue may encounter problems when the required visual features cannot be acquired accurately or reliably. For example, drivers with glasses could pose serious problem to those techniques based on detecting eye characteristics. Glasses can cause glare and may be total opaque to light, making it impossible for camera to monitor eye movement. Furthermore, the degree of eye openness may vary from people to people. Another potential problem with the use of a single visual cue is that the obtained visual feature may not always be indicative of one's mental conditions. For example, the irregular head movement or line of sight (like briefly look back or at the minor) may yield false alarms for such a system.

However uncertain these visual cues are individually, they all, to certain degree, reflect a driver's vigilance. It is our belief that simultaneous presence of several visual cues is a strong indication of a decline in driver's vigilance. It is therefore important to simultaneously use different visual cues to improve the detection accuracy and robustness.

To further improve the detection accuracy, we are also interested in circumstantial factors that may facilitate the onset of a decline in a driver's vigilance. Specific contextual information of interest include road condition, time, and weather condition.

The use of different visual cues and contextual information requires means to systematically integrate the diverse sources of evidence in a uniform manner so that a consistent overall evaluation of a driver's vigilance level can be obtained. We choose to use the Bayesian inference networks for information fusion. A Bayesian network offers an intuitively meaningful semantic network for knowledge and uncertainty representation and for integrating evidences at different levels of abstractions in a manner consistent with probability theory. An overview of the proposed system is illustrated in figure 1. This paper discusses our techniques for extracting various visual cues.

2 Visual Cues and Vigilance Levels

A driver with a reduced vigilance level often exhibits certain unique observable facial characteristics. These visual cues may include eye characteristics (e.g. blink frequency, eye closure duration, percent of eye closure), gaze direction, and head movement.

Eye characteristics appear to be the most relevant symptoms indicating a driver's level of vigilance [22]. In fact, based on a recent study by the Federal Highway Administration [23], of the many drowsiness-detection measures, PERCLOS was found to be the most reliable and valid measure of a driver's alert-

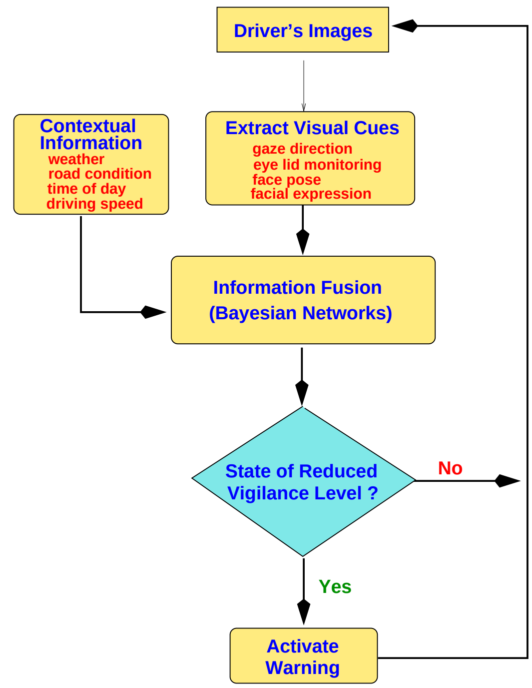


Figure 1. A flowchart of the proposed vigilance monitoring system

ness level. PERCLOS measures the percentage of eyelid closure over the pupil over time.

The movement of a driver's line of sight (gaze) may have the potential to indicate a driver's intention and his/her physical or mental conditions. For normal driving, line of sight is front. When people are drowsy or drunken, their visual awareness cannot cover a wide enough area. Consequently, the line of sight is concentrated on one direction. Head movement like nodding or inclination is a good sign of sleepiness or intoxication. In this research, we focus on investigating techniques for extracting those three types of visual cues.

3 Image Acquisition

For image acquisition, the driver's face will be illuminated using infrared LEDs to minimize the impact of different ambient light conditions. This ensures quality images under varying real-world conditions including poor illumination, day, and night. To further minimize interference from light sources beyond infrared light and to maintain uniform illumination under different climatic conditions, an infrared filter will be installed on the camera.

Two infrared sensitive cameras, mounted on the dashboard, will be used to obtain images of different aspects of the driver for accurately characterizing his/her visual features and head movement. One camera (the wide-angle camera) obtains a global view of the driver's face and head while the other camera (the narrow-angle) focuses on the eyes of the driver. The wide-angle camera monitors the driver's head movement and performs face tracking. Upon detecting any irregular head movement, the camera computes the pose of the driver's face and passes it to the narrow-angle camera, which then pans, tilts, and zooms to obtain a close-up image of the driver's eyes as shown in figure 2.

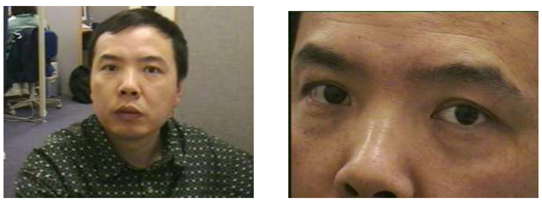


Figure 2. An example of two images obtained from the wide-angle (a) and close-up (b) cameras.

This allows to more accurately monitor driver's eye-lid movements and gaze direction. The acquired images will then be processed using techniques described in the next sections to extract the required visual cues.

4 Visual Cues Extraction

Visual cues that typically reflect a driver's level of alertness include eye movement (i.e., degree of openness and eye open/close frequency), head orientation, and gaze direction. Figure 3 depicts the proposed visual extraction system. Next, we will briefly explain each component.

5 Face Detection and Tracking

5.1 Face Detection

Locating faces in images is a prerequisite step for face tracking, eyes tracking and gaze monitoring. Common approaches for automatic face detection rely on color, texture, facial features, or motion cues. These techniques tend to be generic, require strong assumptions, and do not exploit any domain-specific constraints. In our application, there exist a number of constraints which can significantly simplify face detection. Specifically, the environment under which our system will work may be summarized as

- relatively simple background
- limited area of face movement

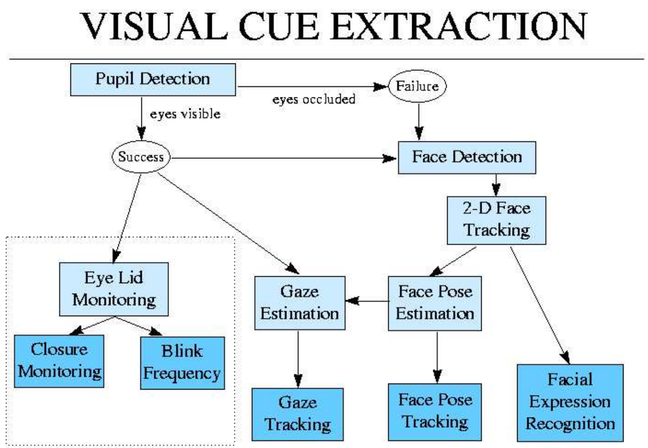


Figure 3. Components of the proposed visual cues extraction system

- single individual face
- infrared illumination

The use of infrared LEDs for illumination allows us to employ a different strategy for face detection. This approach first searches for certain easily identifiable facial features (e.g., eyes) and then uses the locations of these features to locate face. This approach suites very well with this research.

Morimoto et al [11] and Ebisawa [5] have proposed a simple yet effective approach for pupil detection based on a differential lighting scheme. Specifically, their technique obtains a dark and a bright pupil image by illuminating the eyes with IR LEDs located off and on the optical axis respectively. The pupil is detected from the difference of the two images as shown in Figure 4. Although special lighting and synchronization schemes are required for this approach to work, they, nevertheless, have shown that this approach works in a wide range of scales and illumination conditions.

If the driver's eyes are occluded due to glasses or if the driver is not facing the camera, we plan to use different techniques for face detection. One method that we are currently investigating uses an ellipse to model the projection of the head in the image [2]. The algorithm is quite simple: given an image, the image gradient is computed first. Then, a local search is performed to determine the position and size of the best ellipse by maximizing the normalized sum of image gradients around the perimeter of the ellipse. Despite its simplicity, the method is quite effective as Figure 5 illustrates. The above method can be further improved by requiring that the gra-

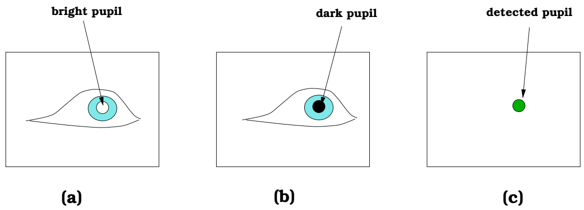


Figure 4. The proposed technique for pupil detection: (a) bright pupil image generated by IR light source along camera optical axis; b) dark pupil generated by IR light off the optical axis; c) detected pupil from the difference of the two images in (a) and (b)

dient vector and ellipse normal at a point on the perimeter of the ellipse have very close directions. This will increase the robustness of the algorithm, especially when strong image gradients due to background objects exist in the image. We are in the process of implementing and testing this idea.

Figure 5. The elliptical head detector and tracker. The ellipses drawn around the head, show the detected locations of the head over a sequence of consecutive frames.

5.2 Face Tracking

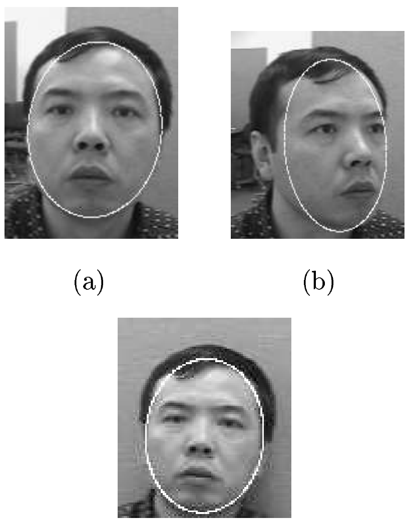
To continuously monitor the driver, it is important to track his face over different frames in real-time. This can be done efficiently by using the location of the face in previous frames to predict the location of the face in future frames. For our project, face tracking can be facilitated by taking into consideration that the driver's face will not undergo significant size changes and that face movement is limited to a restricted area. We have used these constraints to test the face detection technique described in the previous section. Specifically, to find the best ellipse in a new frame, the local search begins from a predicted position and considers only a small neighborhood around that position. The predicted position is obtained using the positions found in the previous two frames, assuming that the head's velocity is constant. The results shown in Figure 5 have been obtained using this simple model. The method has been tested on a number of different cases and it has shown to tolerate large degrees of head rotation as well as occlusion. In addition, it is capable of recovering from failure (i.e., when the face disappears from the view of the camera for some time). While this method will be sufficient when the driver is at a normal state, more robust techniques are required when the driver moves his head irregularly (e.g., intoxicated drivers). We plan to investigate prediction schemes based on Kalman filtering [3] for dealing with these cases effectively.

6 Face Pose Estimation

Head movement can be captured by estimating and tracking the driver's face pose (face orientation and position). Methods for face pose estimation can be classified into two main categories: *model-based* and *appearance-based*. Model-based approaches assume a 3D model of the face and typically recover the face pose by establishing 2D to 3D features correspondences or using shading ([20], [8], [9]). On the other hand, appearance-based approaches are based on view interpolation or eigenfaces and their goal is to construct an association between appearance and face orientation ([15], [13], [14]). Although appearance-based methods are simpler, they are expected to be less accurate than model-based approaches. Our application, however, does not require very accurate face pose estimation. Thus, we plan to investigate both model-based and appearance-based techniques in this project.

In order for a model-based approach to be practical, it should make minimal assumptions about the structure of the face. Currently, we are investigating two model-based approaches: the first one is a new approach which models the shape of the face with an ellipse. According to this approach, face pose estimation amounts to reconstructing the 3D ellipse (3D face) from the observed image ellipses (2D faces). The second approach has been proposed by Gee and Cipolla [8] and models the human face using lengths between five facial points (the far corners of the eyes, the far corners of the mouth and the tip of the nose). We have found this approach attractive since it employs a very simple and general model.

First, we discuss our proposed approach which models the face with an ellipse. Elliptical shapes appear distorted at different orientations and position as shown in figure 6. There are two sources contributing to face shape distortion: foreshortening and perspective projection. Face orientation changes introduce foreshortening distortion as shown in figure 6 (b). Perspective distortion renders a smaller face as the face is far from camera as shown in figure 6 (c). Perspective distortion may be quantified by the area of the face ellipse while foreshortening by the ratio of the ellipse semiaxes. In practice, the face shape distortion can be a combination of both distortions.



(a) (b)

(c)

Figure 6. Face shape distortions due to face orientation (foreshortening) (b) and perspective projection (c): (a) look front ; (b) look side way; (c) look front but far from camera

Hence, face pose estimation amounts to inferring face orientation and position from face shape distortions. This process, however, requires two images since inferring 3D face pose from a single image is ill-posed. In the following paragraphs, we briefly discuss our technique for estimating 3D face ellipse from two image face ellipses observed from two different viewpoints.

Let C_1 and C_2 represent two image face ellipses obtained from two different view points, Q is the 3D face ellipse, and R and T are the rotation matrix and translation vector relating the two view points as shown in Figure 7. Our goal is to compute Q given C_i, i=1,2, R and T.

Based on the theory of perspective projection [18], we have

$$
\begin{aligned}
G_i^t C_i G_i &= k_i Q \\
R_2 &= R R_1 \\
T_2 &= R T_1 + T
\end{aligned}
\qquad (1)
$$

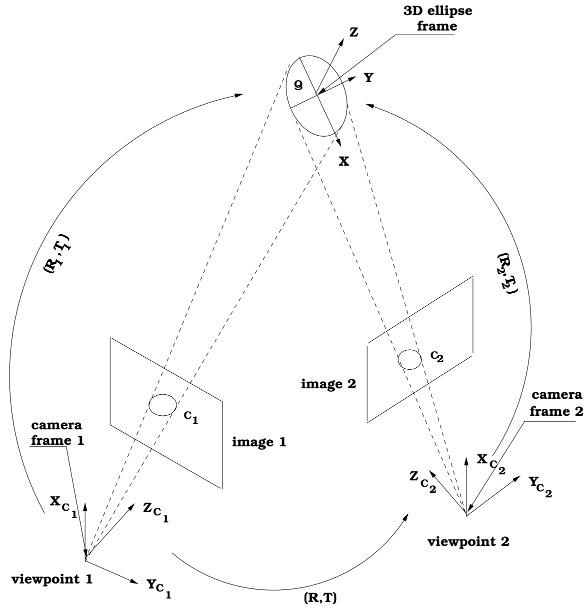


Figure 7. 3D face pose estimation from two image face ellipses

where R_i and T_i, $i = 1, 2$ relate the 3D ellipse frame to the two camera frames, and $G_i = (r_{i1} \ \ r_{i2} \ \ T_i)$.

Since the matrices in equation 1 are all 3×3 symmetric matrices, these equations provide 12 constraints for 9 independent unknowns: six for R_1 and T_1 plus four unknowns k_1, k_2, and two ellipse parameters (major and minor axes). The problem is therefore over-determined. We can therefore solve for the 10 unknowns, which yields the position and orientation and the 3D face ellipse that has generated the two observed image ellipses.

According to the second approach [8], the following four lengths are used to model the face: L_f (eye-to mouth), L_n (nose-base to nose-tip), L_e (eye-to-eye) and L_m (mouth to nose-tip). Figure 8 illustrates these lengths. Although these lengths usually vary from person to person, ratios between them do not change very much. Specifically, each of the above lengths is divided by L_f, thus, yielding three normalized ratios ($R_n{=}L_n/L_f$, $R_e{=}L_e/L_f$ and $R_m{=}L_m/L_f$) which are used to determine face orientation (i.e., the orientation of the vector from the base to the tip of the nose). The method proceeds as follows: first, the symmetry axis of the face plane is determined by finding the midpoints of the eye and mouth points. Then, the nose base is determined using the nose ratio R_n. Finally, the projec-

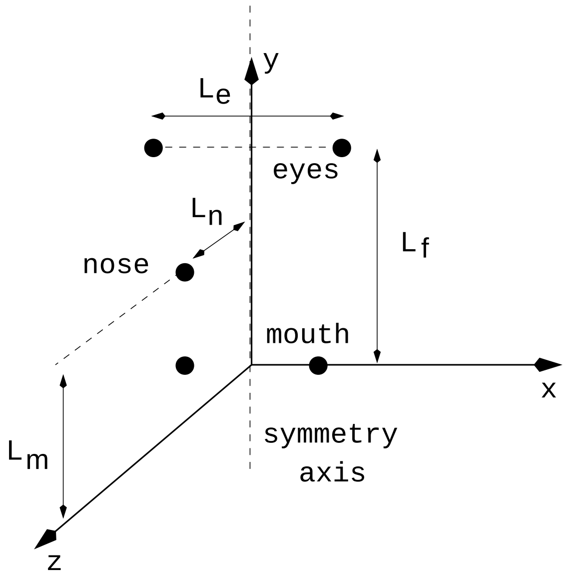


Figure 8. The face model used by Gee and Cipolla [8]

tion of the face normal is found by connecting the base to the tip of the nose. The tilt angle is simply the angle between the x-axis to the projection of the face normal. The slant angle is computed using R_n and R_m. It should be mentioned that weak perspective is assumed and that the above angles are with respect to the camera coordinate system.

We have performed a number of experiments to evaluate this approach. Figure 9 shows the images used while Table 1 shows the angles computed (the camera is lower that the face). The approach is effective and has very low computational requirements, however, the main challenge is in locating the five facial points reliably. The bright-dark eye pupil effect will be very useful in detecting the eye corners while the approximate location of the other features will be determined by using model ratios. A local search will then be applied to locate these features accurately.

7 Eye Movement Monitoring and Gaze Tracking

Eye movement monitoring includes measurement of percentage of eye closure, the PERCLOS measure, eye blinking frequency, and eye closure duration. To obtain these measurements, we propose to continuously track the driver's pupil and monitor its size (as characterized by its area). This may be achieved using the pupil detection algorithm discussed in section 5.1, coupled with Kalman filtering. An eye closure starts when the pupil size shrinks to a fraction of its nominal size.

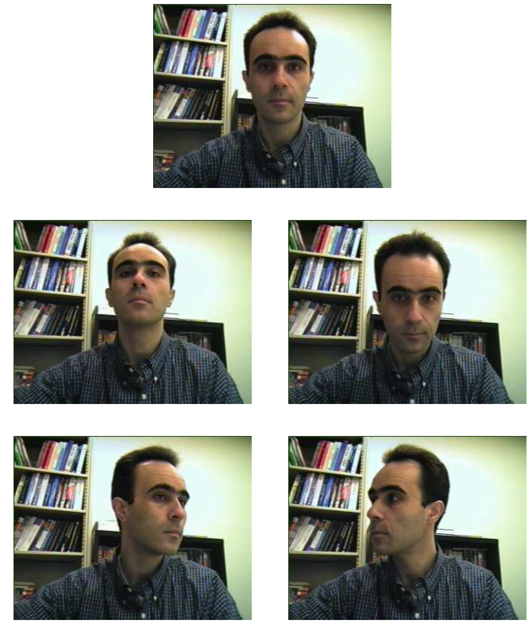


Figure 9. Some results of face orientation estimation using the method of Gee and Cipolla [8]

Table 1. Face orientation results

Images	Tilt (degree)	Slant (degree)
g1	96.581951	29.79006
g2	98.494375	47.597893
g3	280.175516	26.133798
g4	30.481387	57.755945
g5	158.519443	66.078906

The next step is to estimate the driver's eye gaze, that is, to determine where the driver is looking from the appearance of his/her eyes. Current eye gaze tracking methods rely on intrusive techniques such as measuring the reflection of some light that is shone onto the eye, measuring the electric potential of the skin around the eyes or applying special contact lenses.

We propose to use a non-intrusive approach for gaze monitoring. Of the numerous techniques proposed for gaze estimation [1] [25], [14], the one proposed by Ebisawa [5] appears very promising and is directly applicable to this project. Their technique estimates the gaze direction based on the relative position between pupil and the glint [1] as shown in Figure 10.

This technique seems to be sufficient for our purpose since an accurate gaze estimation is not necessary. The technique quantizes the gaze direction into five directions: center, left, right, up, and down as

[1] the brightest light spot in the eye due to light reflection

shown in Figure 10. Based on the relative position between gaze and glint as well as the light direction, it is possible to determine which of the five directions the driver is looking at. For normal driving, the gaze should be around the front position.

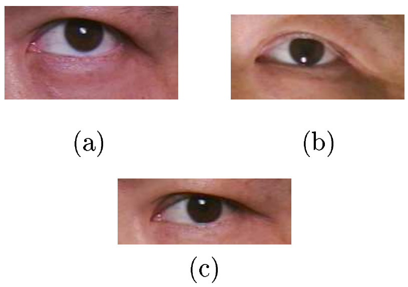


Figure 10. Examples of gaze directions. The bright white spots represent glint. The location of glint in eyes is determined by the light direction. (a) line of sight is *in* the direction of light; b) line of sight looks *up* with respect to the light direction; c) line of sight looks *down* with respect to the light direction

The above technique produces only an approximate gaze direction. To improve this, we propose to analytically estimate the local gaze direction based on pupil location. This can be accomplished by approximating the eyeball with a sphere and the local gaze by the normal of the point on the sphere, where the pupil is located as shown in Figure 11.

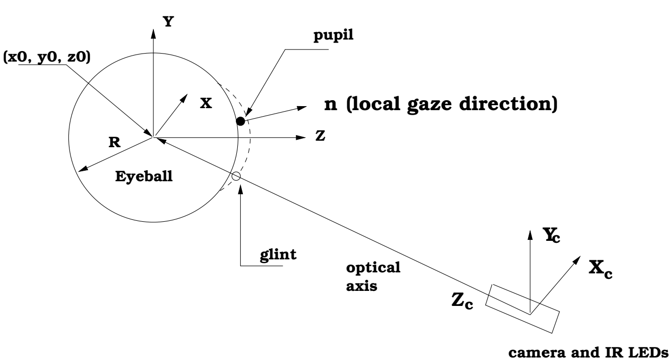


Figure 11. Local gaze estimation

Despite the above approach, we have also been investigating another idea which involves the recovery of gaze direction using the shape of the iris or the pupil in the image. Since circles are projected on the image plane as ellipses, the idea is to estimate the pose of the iris from the distortion of its projection (i.e., ellipse) in the image. The idea is similar to what we have proposed for face pose estimation (see section 5.3), however, it is simpler because the iris is circular, not elliptical. This simplifies things considerably since only one image is now required. We have performed a number of experiments using the pose-from-ellipse algorithm given in [19]. Figure 12 illustrates some of the initial results we have obtained. The main challenge with this approach is that it requires a close view of the eye.

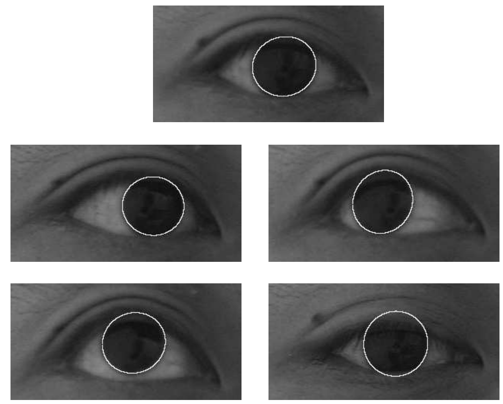


Figure 12. Gaze estimation using pose from ellipse algorithm

It should be mentioned that there are two main components to gaze estimation: the orientation of the driver's head and the orientation of the driver's eyes within their sockets. To compensate for gaze change due to head orientation change, we will incorporate the head orientation (estimated from face pose) into the final gaze direction. Figure 13 shows the complete algorithm for gaze estimation.

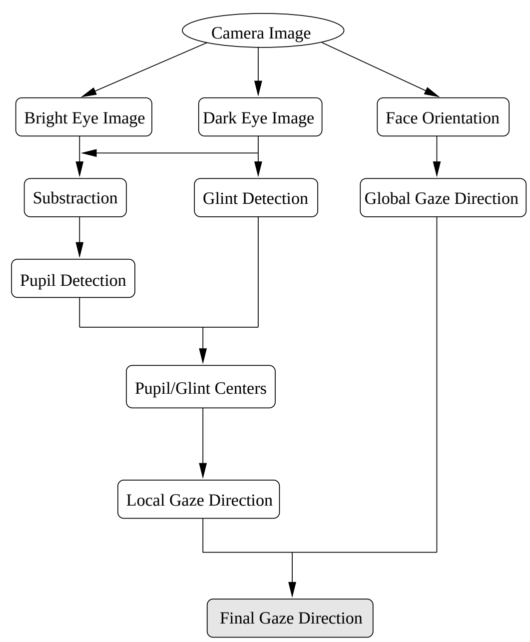


Figure 13. An accurate gaze estimation technique

8 Potential Applications

The computer vision technologies to be developed through this research may find applications in other areas of automobile industry. For example, the technology for determining the line of sight of a person may find potential applications in the peripheral control of automobile equipment such as switch on/off an automotive equipment using one's fixation (e.g. Eye Switch).

The facial recognition technology can also be used to identify and recognize the car driver. This information may be used for automatic adjustments of cockpit configurations like seat position, mirror adjustments, and radio selections, and may also be used for driver authentication to provide anti-theft functions.

9 Conclusions and Future Work

In this paper, we summarize our latest work on extracting various visual cues to characterize a driver's vigilance level. The work presented in this paper is preliminary and on-going. The results, however, are very encouraging and meet with our expectations.

In the subsequent months, we will further validate these techniques with more image data and implement our proposed techniques for computing eye movement measurements. Another important part of our future work is to develop the proposed Bayesian network so that the extracted visual cues and the available contextual information can be integrated to achieve a robust and accurate characterization of driver's vigilance level.

10 Acknowledgment

This work is supported by a grant from HONDA R & D Americas, Inc.

References

[1] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. *Technical Report CMU-CS-94-102, Carnegie Mellon University*, 1994.

[2] S. Birchfield. An elliptical head tracker. *31st Asilomar Conference*, pages 1710–1714, 1997.

[3] A. Blake, R. Curwen, and A. Zisserman. A framework for spatiotemporal control in the tracking of visual contours. *Int. Journal of Computer Vision*, 11(2):127–145, 1993.

[4] S. Boverie, J. M. Leqellec, and A. Hirl. Intelligent systems for video monitoring of vehicle cockpit. *1998 International congress and exposition ITS: Advanced controls and vehicle navigation systems*, pages 1–5, 1998.

[5] Y. Ebisawa. Unconstrained pupil detection technique using two light sources and the image difference method. *Visualization and Intelligent Design in Engineering*, pages 79–89, 1989.

[6] M. K. et al. Development of a drowsiness warning system. *11th international conference on enhanced safety of vehicle, Munuch, 1994*, 1994.

[7] J. Feraric, M. Kopf, and R. Onken. Statistical versus neural bet approach for driver behavior description and adaptive warning. *11th European annual manual*, pages 429–436, 1992.

[8] A. Gee and R. Cipoll. Determining the gaze of faces in images. *Image and Vision Computing*, 30:639–647, 1994.

[9] T. Horprasert, Y. Yacoob, and L. Davis. Computing 3d head orientation from a monocular image. *Proc. of Int. Conf. on automatic face and gesture recognition*, pages 242–247, 1996.

[10] T. Ishii, M. Hirose, and H. Iwata. Automatic recognition of driver's facial expression by image analysis. *Journal of JSAE*, 41(12):1398–1403, 1987.

[11] C. Morimoto, D. Koons, A. Amir, and M. Flickner. Real-time detection of eyes and faces. *1998 Workshop on perceptual user interfaces*, pages 117–120, 1998.

[12] R. Onken. Daisy, an adaptive knowledge-based driver monitoring and warning system. *Proceedings of 1994 Vehicle navigation and information systems conference, Yokohama, Japan, Aug. 1994*, pages 3–10, 1994.

[13] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *CVPR-94*, pages 84–91, 1994.

[14] R. Rae and H. Ritter. Recognition of human head orientation based on artificial neural networks. *IEEE Transactions on Neural Networks*, 9(2):257–265, 1998.

[15] E. O. S. Gong and S. McKenna. Learning to associate faces across views in vector space of similarities of prototypes. *Proc of British Machine Vision*, 1998.

[16] H. Saito, T. Ishiwaka, M. Sakata, and S. Okabayashi. Applications of driver's line of sight to automobiles-what can driver's eye tell. *Proceedings of 1994 Vehicle navigation and information systems conference, Yokohama, Japan, Aug. 1994*, pages 21–26, 1994.

[17] S. Saito. Does fatigue exist in a quantitative of eye movement ? *Ergonomics*, 35, 1992.

[18] D. M. Song. Conics-based stereo, motion estimation, and pose determination. *International journal of computer vision*, 10(1):7–25, 1993.

[19] E. Trucco and A. Verri. Introductory techniques for 3d computer vision. *Prentice Hall*, 1998.

[20] A. Tsukamoto, C. Lee, and S. Tsuji. Detection and pose estimation of human face with synthesized image models. *ICPR 94*, pages 754–757, 1994.

[21] H. Ueno, M. Kaneda, and M. Tsukino. Development of drowsiness detection system. *Proceedings of 1994 Vehicle navigation and information systems conference, Yokohama, Japan, Aug. 1994*, pages 15–20, 1994.

[22] Wierville. Overview of research on driver drowsiness definition and driver drowsiness detection. *ESV, Munich*, 1994.

[23] W. W. Wierville, L. A. Ellsworth, and et al. Research on vehcile based driver status/performance monitoring: devlopment, validation, and refinement of algorithms for detection of driver drowsiness. *National Highway Traffic Safety Admistration Final Report: DOT HS 808 247*, 1994.

[24] K. Yammamoto and S. Higuchi. Development of a drowsiness warning system. *Journal of SAE of Japan*, 46(9), 1992.

[25] G. Yang and A. Waibel. A real-time face tracker. *Workshop on Applications of Computer Vision*, pages 142–147, 1996.