

Real Time and Non-intrusive Driver Fatigue Monitoring

Zhiwei Zhu*

Department of Electrical, Computer, and
Systems Engineering
Rensselaer Polytechnic Institute
Troy, New York, USA
zhuz@rpi.edu

Qiang Ji**

Department of Electrical, Computer, and
Systems Engineering
Rensselaer Polytechnic Institute
Troy, New York, USA
qji@ecse.rpi.edu

Abstract—This paper describes a real-time non-intrusive prototype driver fatigue monitor. It uses remotely located CCD cameras equipped with active IR illuminators to acquire video images of the driver. Various visual cues typically characterizing the alertness of the driver are extracted in real time and systematically combined to infer the fatigue level of the driver. The visual cues employed characterize eyelid movement, gaze movement, head movement, and facial expression. A probabilistic model is developed to model human fatigue and to predict fatigue based on the observed visual cues and the available contextual information. The simultaneous use of multiple visual cues and their systematic combination yields a much more robust and accurate fatigue characterization than using a single visual cue. The feasibility of our system is demonstrated using synthetic data. Further validation of our system under real life fatigue conditions with human subjects shows that it was reasonably robust, reliable and accurate in fatigue characterization.

I. INTRODUCTION

The ever-increasing number of traffic accidents in the U.S. due to a diminished driver's vigilance level has become a problem of serious concern to society. Drivers with a diminished vigilance level suffer from a marked decline in their abilities of perception, recognition, and vehicle control and therefore pose serious danger to their own life and the lives of other people. Statistics show that a leading cause for fatal or injury-causing traffic accidents is due to drivers with a diminished vigilance level. In the trucking industry, 57% fatal truck accidents are due to driver fatigue. It is the number one cause for heavy truck crashes. 70% of American drivers report driving fatigued. With the ever-growing traffic conditions, this problem will further deteriorate. For this reason, developing systems actively monitoring a driver's level of vigilance and alerting the driver of any insecure driving conditions is essential to accident prevention.

Many efforts have been reported in the literature on developing active real-time image-based fatigue monitoring systems [1], [2], [3], [4], [5], [6], [7]. But most of them focus on only a single visual cue such as facial expression, eyelid movement or line of gaze or head orientation to characterize driver's state of alertness. The system relying on a single visual cue may encounter difficulty when the required visual features cannot be acquired accurately or reliably.

All those visual cues, however imperfect they are individually, if combined systematically, can provide an accurate

characterization of a driver's level of vigilance. It is our belief that simultaneous extraction and use of multiple visual cues can reduce the uncertainty and resolve the ambiguity present in the information from a single source. The systematic integration of these visual parameters, however, requires a fatigue model that models the fatigue generation process and is able to systematically predict fatigue based on the available visual as well as the relevant contextual information. The system we propose can simultaneously, non-intrusively, and in real time monitor several visual behaviors that typically characterize a person's level of alertness while driving. These visual cues include eyelid movement, gaze movement, head movement and facial expression. The fatigue parameters computed from these visual cues are subsequently combined probabilistically to form a composite fatigue index that could robustly, accurately, and consistently characterize one's vigilance level. Figure 1 gives an overview of our driver vigilance monitoring system.

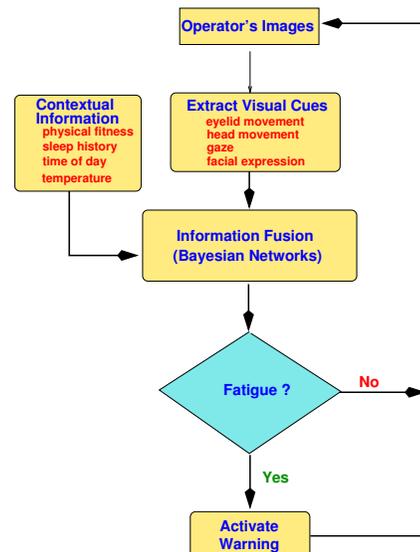


Fig. 1. A flowchart of the proposed driver vigilance monitoring system

The paper consists of three parts. First, the paper focuses on discussion of the computer vision algorithms and the necessary hardware components to extract the needed visual cues. Second, after extracting these visual cues, the issue of

sensory data fusion and fatigue modeling and inference is discussed. Finally, experiments under real life conditions are conducted to validate our driver vigilance monitoring system.

II. VISION-BASED VISUAL CUES EXTRACTION

Fatigue monitoring starts with extracting visual parameters typically characterizing a person's level of vigilance. This is accomplished via a computer vision system. In this section, we discuss the computer vision system we developed to achieve this goal. Figure 2 provides an overview of our visual cues extraction system for driver fatigue monitoring. The system consists of two cameras: one wide angle camera focusing on the face and another narrow angle camera focusing on the eyes. The wide angles camera monitors head movement and facial expression while the narrow angle camera monitors eyelid and gaze movements.

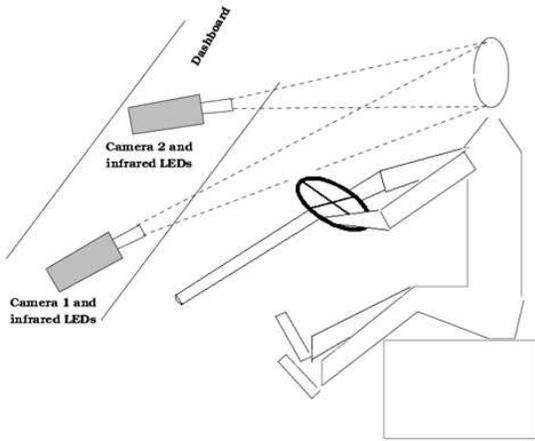


Fig. 2. Overview of the driver vigilance monitoring system

A. Eye Detection and Tracking

The system starts with eye detection and tracking. The goal of eye detection and tracking is for subsequent eyelid movement monitoring, gaze determination, face orientation estimation and facial expression analysis. A robust, accurate, and real-time eye tracker is therefore crucial. In this research, we proposed a real-time robust method for eye tracking under variable lighting conditions and face orientations, based on combining the appearance-based methods and the active IR illumination approach. When the eyes are completely closed or the eyes are partially occluded due to the large face orientations, our eye tracker can still track them successfully as shown in Figure 3. Details about the algorithm can be found in [8]. Video demos of the eye tracker may be found at <http://www.ecse.rpi.edu/~cvrl/Demo/demo.html>.

B. Eyelid Movement Parameters

Eyelid movement is one of the visual behaviors that reflect a person's level of fatigue. The primary purpose of



Fig. 3. Image results with successfully tracked eyes.

eye tracking is to monitor eyelid movements and compute the relevant eyelid movement parameters. Here, we focus on two ocular measures to characterize the eyelid movement. The first one is Percentage of Eye Closure Over Time (*PERCLOS*) and the second is Average Eye Closure Speed (*AECS*). *PERCLOS* has been validated and found to be the most valid ocular parameter for monitoring fatigue [1].

The eye closure/opening speed is a good indicator of fatigue. It's defined as the amount of time needed to fully close the eyes or to fully open the eyes. Our previous study indicates that the eye closure speed of a drowsy person is distinctively different from that of an alert person [9].

The degree of eye opening is characterized by the shape of pupil. It is observed that as eyes close, the pupils start getting occluded by the eyelids and their shapes get more elliptical. So, we can use the ratio of pupil ellipse axes to characterize degree of eye opening. The cumulative eye closure duration over time excluding the time spent on normal eye blinks is used to compute *PERCLOS*. To obtain a more robust measurement for these two parameters, we compute their running average (time tracking). To obtain running average of *PERCLOS* measurement, for example, the program continuously tracks the person's pupil shape and monitors eye closure at each time instance. We compute these two parameters in 30 seconds window and output them onto the computer screen in real time, so we can easily analyze the alert state of the driver. The plots of the two parameters over time are shown in Figure 4.

C. Face (Head) Orientation Estimation

Face (head) pose contains information about one's attention, gaze, and level of fatigue. Face pose determination is concerned with computation of the 3D face orientation and position to detect such head movements as head tilts. Frequent head tilts indicate the onset of fatigue. Furthermore, the nominal face orientation while driving is frontal. If the driver faces in the other directions (e.g., down or sideways) for an extended period of time, this is due to either fatigue or inattention. Face pose estimation, therefore, can indicate both fatigued and inattentive drivers. For this study, we focus on the former, i.e., detection of frequent head tilts.

We present a new technique to perform the 2D face tracking and 3D face pose estimation synchronously. In our method, 3D face pose is tracked by Kalman Filtering. The initial estimated 3D pose is used to guide face tracking in the image, which is subsequently used to refine the 3D face

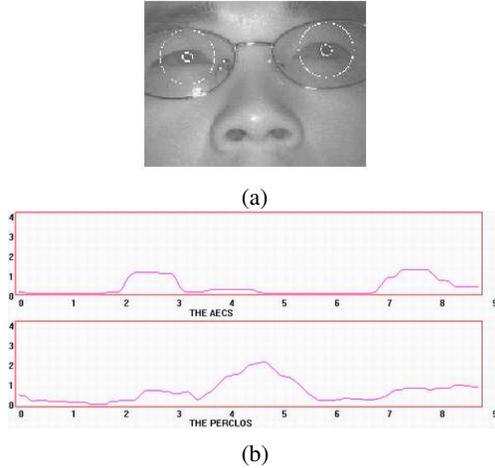


Fig. 4. (a) Detected eyes and pupils (b) Plots for eyelid movement parameters: the top one displays *AECS* parameter and the bottom one displays *PERCLOS* parameter.

pose estimation. Face detection and pose estimation work together and benefit from each other. Details about our 3D face pose tracker can be found in [10].

The proposed algorithm is tested with numerous image sequences of different people. The image sequences include a person rotating his head before an un-calibrated camera, which is approximately 1.5 meter from the person. Figure 5 shows some tracking results under different face rotations. The accuracy of the estimated face pose can be achieved around 2 degrees [10]. Video demos of the 3D face pose tracker may be found at <http://www.ecse.rpi.edu/~cvrl/Demo/demo.html>.



Fig. 5. Face pose tracking results for images randomly selected from one video sequence. The white rectangle indicates the tracked face region and the white line represents the normal of the face plane, which is drawn according to the estimated face pose.

To quantitatively characterize one's level of fatigue by face pose, we introduce a new fatigue parameter called *NodFreq*, which measures the frequency of head tilts over time. Figure 6 shows the running average of the estimated head tilts for a period of 140 seconds. As can be seen, our system can accurately detect head tilts, which are represented in the curve by the up-and-down bumps.

D. Eye Gaze Determination and Tracking

Gaze has the potential to indicate a person's level of vigilance. A fatigued individual tends to have a narrow

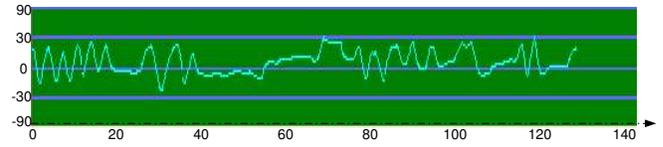


Fig. 6. Head tilts monitoring over time (seconds)

gaze. Gaze may also reveal one's needs and attention. The direction of a person's gaze is determined by two factors: the orientation of the face (face pose), and the orientation of eye (eye gaze). Face pose determines the global direction of the gaze, while eye gaze determines the local direction of the gaze. Global gaze and local gaze together determine the final gaze of the person. So far, the most common approaches for ocular-based gaze estimation [11], [12], [13] are based on local gaze by using the relative position between pupil and the glint (cornea reflection) via a remote IR camera. Therefore, these methods work well only for a static head, and need to perform a rather cumbersome calibration process whenever his/her head moved. This is because only local gaze is accounted for while global gaze due to face pose is ignored. This poses a significant hurdle for practical application of the system.

In view of these limitations, we present a gaze estimation approach [14] that accounts for both the local gaze computed from the ocular parameters and the global gaze computed from the head pose. The global gaze (face pose) and local gaze (eye gaze) are combined together to obtain the precise gaze information of the user. Our approach, therefore, allows natural head movement while still estimating gaze accurately. Another effort is to make the gaze estimation calibration free. New users or the existing users who have moved, do not need undergo a personal gaze calibration before using the gaze tracker. Therefore, the proposed gaze tracker can perform robustly and accurately without calibration and under natural head movements.

Experiments were conducted to study the performance of our gaze estimation technique. An average of gaze classification accuracy of (96% accuracy) was achieved. Details on our gaze estimation algorithm may be found in [14].

Given the gaze, we can compute a new fatigue parameter named *GAZEDIS*, which represents the gaze distribution over time to indicate the driver's fatigue or attention level. *GAZEDIS* measures the driver's situational awareness. Another fatigue parameter we compute is *PERSAC*, which is the percentage of saccade eye movement over time. Saccade eye movements represent the deliberate and conscious driver action to move eye from one to another place. It therefore can measure the degree of alertness. The value of *PERSAC* is very small for a person in fatigue. Figure 7 plots the *PERSAC* parameter over 30 seconds.

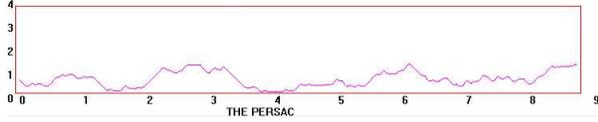


Fig. 7. Plot of PERSAC parameter over 30 seconds.

E. Facial Expression Analysis

Besides eye and head movements, another visual cue that can potentially capture one's level of fatigue is his/her facial expression. In general, people tend to exhibit different facial expressions under different levels of vigilance. The facial expression of a person in fatigue or in the onset of fatigue can usually be characterized by lagging facial muscles, expressionless, and frequent yawnings.

Our recent research has led to the development of a feature-based facial expression analysis algorithm [15]. The facial features around eyes and mouth represent the most important spatial patterns composing the facial expression. Generally, these patterns with their changes in spatio-temporal spaces can be used to characterize facial expressions. For the fatigue detection application, in which there are only limited facial expressions, the facial features around eyes and mouth include enough information to capture these limited expressions. So in our research, we focus on the facial features around eyes and mouth.

In our method, the multi-scale and multi-orientation Gabor wavelets are used to represent and detect each facial feature. The detected features and their spatial connections are used to characterize facial expressions. Details can be found in [15].

A series of experiments are conducted in [15], and results show that our proposed algorithm can track the facial features robustly under large head movements, self-occlusion and different facial expressions. Figure 8 shows some results of a typical sequence of a person in fatigue.



Fig. 8. Tracked facial features and local graphs.

For now, we focus on monitoring mouth movement to detect yawning. A yawning is detected if the features around mouth significantly deviate from its closed configuration, especially in vertical direction. The degree of mouth opening is characterized by the shape of mouth. Therefore, the openness of the mouth can be represented by the ratio of mouth height and width.

We develop a new measure of facial expression, *Yawn-*

Freq, which computes the occurrence frequency of yawning over time. Figure 9 shows the plot of *YawnFreq* over time, and a yawning is represented by an up-and-down bump.

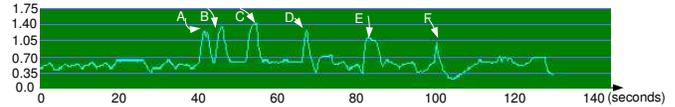


Fig. 9. The plot of the openness of the mouth over time. The bumps *A, B, C, D, E, F* are the detected yawns.

III. FATIGUE MODELING USING BAYESIAN NETWORKS

As we discussed above, human fatigue generation is a very complicated process. Several uncertainties may be present in this process. First, fatigue is not observable and it can only be inferred from the available information. In fact, fatigue can be regarded as the result of many contextual variables such as working environments, health and sleep history. Also, it is the cause of many symptoms, e.g. the visual cues, such as irregular eyelid movements, yawning and frequent head tilts. Second, human's visual characteristics vary significantly with age, height, health and shape of face. To effectively monitor fatigue, a system that integrates evidences from multiple sources into one representative format is needed. Naturally, a Bayesian Networks (BN) model is the best option to deal with such an issue. After integrating these evidences, the BN model will output a value (or composite fatigue index) to represent the fatigue level of the driver.

A BN provides a mechanism for graphical representation of uncertain knowledge and for inferring high level activities from the observed data. The main purpose of a BN model is to infer the unobserved events from the observed or contextual data. The vision system discussed in previous sections can compute several visual fatigue parameters. They include *PERCLOS* & *ACSE* for eyelid movement, *NodFreq* for head movement, *GAZEDIS* and *PERSAC* for gaze movement, and *YawnFreq* for facial expression. Putting all these factors together, the BN model for fatigue is constructed as shown in Fig.10. The target node is fatigue. The nodes above the target node represent various major factors that could lead to one's fatigue. They are collectively referred to as contextual information. The nodes below the target node represent visual observations from the output of our computer vision system. These nodes are collectively referred to as observation nodes.

IV. SYSTEM VALIDATION

A. Validation with Synthetic Data

Before using the BN for fatigue inference, the network is parameterized as discussed in [16]. Given the parameterized model, fatigue inference can then commence upon the arrival of visual evidences via belief propagation. Here we use some typical combination of evidences and their results are summarized in Table I.

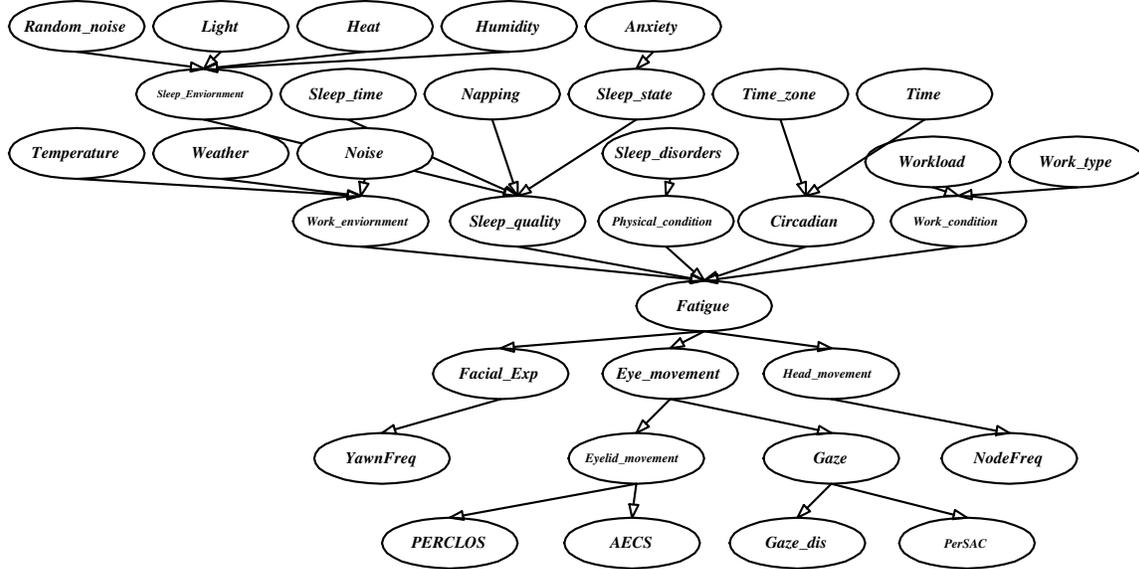


Fig. 10. Bayesian Network model for monitoring human fatigue

TABLE I
THE INFERENCE RESULTS OF FATIGUE BAYESIAN NETWORK
MODEL

Ref. No.	Evidences Instantiated	Fatigue Prob.
1	No any evidence	0.5755
2	YawnFreq (high)	0.8204
3	PERCLOS (high)	0.8639
4	AECS (slow), Sleep time (insufficient), Time (drowsy time)	0.9545
5	YawnFreq (high), AECS(slow)	0.9552
6	Sleep time (insufficient), Time (drowsy time), Temperature (high)	0.8363

From Table I, we can see that the prior probability of fatigue (e.g. when there is not any evidence) is about 0.5755 (ref#1). The observation of single visual evidence does not usually provide conclusive finding since the estimated fatigue probability is less than the critical value 0.95¹ (ref#2 and ref#3). Even when PERCLOS is instantiated, the fatigue probability reaches 0.8639, which is still below the threshold 0.95. This indicates that one visual cue is not sufficient to conclude if the person is fatigued. On the other hand, when combined with some contextual evidences, any visual parameter can lead to a high fatigue probability (ref#4). This demonstrates the importance of contextual information. The simultaneous observation of abnormal values for two visual parameters (ref#5) such as NodeFreq and PerSAC can lead to a fatigue probability higher than 0.95. This makes sense since they quantify fatigue from two different perspectives: one is gaze and the other is

¹a hypothesized critical fatigue level. It may vary from application to application

head movement. Any simultaneous observation of abnormal values of three or more visual parameters guarantees that the estimated fatigue probability exceeds the critical value. The simultaneous presence of several contextual evidences only leads to a high probability of fatigue, even in the absence of any visual evidence. These inference results, though preliminary and synthetic, demonstrate the utility of the proposed framework for predicting and modelling fatigue.

B. System Validation with Real Data

The last part of this research is to experimentally and scientifically demonstrate the validity of the computed fatigue parameters as well as the composite fatigue index. The validation consists of two parts. The first part involves the validation of the measurement accuracies of our computer vision techniques, and the second part studies the validity of the fatigue parameters and the composite fatigue index that our system computes in characterizing fatigue.

Experiments are conducted to quantitatively characterize the measurement accuracies of our computer vision techniques in measuring eyelid movement, gaze, face pose and facial expressions. The measurements from our system are compared with those obtained either manually or using conventional instruments. The study shows that our eye tracker is quite accurate, with a false alarm rate of 0.05% and a misdetection rate of 4.2%. For the 3D face pose tracker, quantitatively, the RMS errors for the estimated pan and tilt angles are 1.92 degrees and 1.97 degrees respectively. It demonstrates that our face pose estimation technique is sufficiently accurate.

To study the validity of the proposed fatigue parameters and that of the composite fatigue index, we performed a human subject study. The study included a total of 8

subjects. Two test bouts were performed for each subject. The first test was done when they first arrived in the lab at 9 pm and when they were fully alert. The second test was performed about 12 hours later early in morning about 7 am the following day, after the subjects have been deprived of sleep for a total of 25 hours.

During the study, the subjects are asked to perform a TOVA (Test of Variables of Attention) test. The TOVA test consists of a 20-minute psychomotor test, which requires the subject to sustain attention and respond to a randomly appearing light on a computer screen by pressing a button. TOVA test was selected as the validation criterion because driving is primarily a vigilance task requiring psychomotor reactions, and psychomotor vigilance. The response time is used as a metric to quantify the subject's performance.

Figure 11 plots the TOVA performance versus the composite fatigue score. It clearly shows that the composite fatigue score (based on combining different fatigue parameters) highly correlates with the subject's response time.

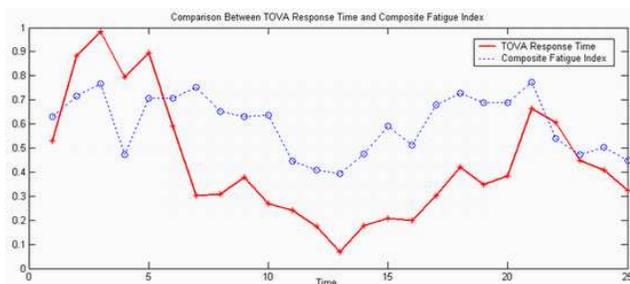


Fig. 11. The estimated composite fatigue index (dashed line) versus the normalized TOVA response time. The two curves track each other well.

It is clear that the two curves' fluctuations match well, proving their correlation and co-variation, therefore proving the validity of the composite fatigue score in quantifying performance.

V. CONCLUSION

Through research presented in this paper, we developed a non-intrusive prototype computer vision system for real-time monitoring a driver's vigilance. First, the necessary computer vision algorithms are developed to simultaneously extract multiple visual cues that typically characterize a person's level of fatigue. Then, a probabilistic framework is built to model fatigue, which systematically combines different visual cues and the relevant contextual information to produce a robust and consistent fatigue index.

Experiment studies in a real life environment with subjects of different ethnic backgrounds, different genders and ages were scientifically conducted to validate the fatigue monitoring system. Experiment results show that our fatigue monitor system is reasonably robust, reliable and accurate in characterizing human fatigue. It represents state of the art in real time on-line and non-intrusive fatigue monitoring.

REFERENCES

- [1] David F. Dinges, M. Mallis, G. Maislin, and J. W. Powell, "Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management," *Department of Transportation Highway Safety publication 808 762*, vol. April, 1998.
- [2] J. H. Richardson, "The development of a driver alertness monitoring system," *Fatigue and Driving: Driver impairment, driver fatigue and driver simulation*, Taylor and Francis; London, L. Hartley (Ed), 1995.
- [3] W. Mullally and M. Betke, "Preliminary investigation of real-time monitoring of a driver in city traffic," *IEEE International Conference on Intelligent Vehicles, Dearborn, MI*, 2000.
- [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*, pp. 1-5, 1998.
- [5] M. Kaneda et al, "Development of a drowsiness warning system," *11th international conference on enhanced safety of vehicle, Munich, 1994*, 1994.
- [6] Richard Grace, "A drowsy driver detection system for heavy vehicles," *Conference on ocular measures of driver alertness*, vol. April, 1999.
- [7] Dixon Cleveland, "Unobtrusive eyelid closure and visual of regard measurement system," *Conference on ocular measures of driver alertness*, vol. April, 1999.
- [8] Zhiwei Zhu, Qiang Ji, Kikuo Fujimura, and Kuang chih Lee, "Combining kalman filtering and mean shift for real time eye tracking under active ir illumination," in *International Conference on Pattern Recognition*, Quebec, Canada, 2002.
- [9] Q. Ji and X. Yang, "Real time visual cues extraction for monitoring driver vigilance," in *Proc. of International Workshop on Computer Vision Systems*, Vancouver, Canada, 2001.
- [10] Zhiwei Zhu and Qiang Ji, "3d face pose tracking from an uncalibrated monocular camera," *International Conference on Pattern Recognition, August 23-26, 2004*, 2004.
- [11] Takehiko Ohno, Naoki Mukawa, and Atsushi Yoshikawa, "Freegaze: A gaze tracking system for everyday gaze interaction," *Eye Tracking Research and Applications Symposium, 25-27 March, New Orleans, LA, USA*, 2002.
- [12] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Frame-rate pupil detector and gaze tracker," *IEEE ICCV'99 FRAME-RATE WORKSHOP*, 1999.
- [13] Y. Ebisawa, "Improved video-based eye-gaze detection method," *IEEE Transactions on Instrumentation and Measurement*, vol. 47, no. 2, pp. 948-955, 1998.
- [14] Qiang Ji and Zhiwei Zhu, "Eye and gaze tracking for interactive graphic display," in *2nd International Symposium on Smart Graphics*, Hawthorne, NY, USA, 2002.
- [15] Haisong Gu, Qiang Ji, and Zhiwei Zhu, "Active facial tracking for fatigue detection," *IEEE Workshop on Applications of Computer Vision, Orlando, Florida*, 2002.
- [16] Peilin Lan, Qiang Ji, and Carl Looney, "Information fusion of visual cues using dynamic bayesian belief networks for inferring human cognitive states," *the Invited Session on Information Fusion using Bayesian Networks, Fusion 2002, Annapolis, Maryland*, 2002.