Evaluating Driving Fatigue Detection Algorithms Using Eye Tracking Glasses

Xiang-Yu Gao, Yu-Fei Zhang, Wei-Long Zheng and Bao-Liang Lu* *Senior Member*, *IEEE*

***Abstract*— Fatigue is a status of human brain activities, and driving fatigue detection is a topic of great interest all over the world. In this paper, we propose a measure of fatigue produced by eye tracking glasses, and use it as the ground truth to evaluate driving fatigue detection algorithms. Particularly, PERCLOS, which is the percentage of eye closure over the pupil over time, was calculated from eyelid movement data provided by eye tracking glasses. Experiments of a vigilance task were carried out in which both EOG signals and eyelid movement were recorded. The evaluation results of an effective EOG-based fatigue detection algorithm convinced us that our proposed measure is an appropriate candidate for evaluating driving fatigue detection algorithms.**

I. INTRODUCTION

Driving fatigue is believed to be one of the major causes of traffic accidents. Detection and prediction of driving fatigue are important for the safety of drivers. Related topics have been studied for decades [1],[2].

There are various driving fatigue detection algorithms under development. In order to evaluate these algorithms, we need a reliable measure of the subject's current fatigue level served as the ground truth. In this paper, we present several measures of fatigue used in laboratory and analyze their characteristics.

We propose to use eye tracking glasses (SensoMotoric Instruments) to define a measure of fatigue. Compared to other measures, there are several advantages of using eye tracking glasses (ETG). They are wearable, accurate, robust and objective, and a lot of eye movement parameters related to fatigue are recorded. We can not only use them in driving simulation experiments but also in real driving conditions.

PERCLOS, which is one of the most widely accepted measures in the scientific literature for driving fatigue detection [3], is calculated from the eyelid movement parameters of eye tracking glasses (called '$PERCLOS_{ETG}$' in this paper). It was validated by comparison with that of a pupil detection procedure using the videos of eyes. Unlike other fatigue-related research, we use $PERCLOS_{ETG}$ as the ground truth for evaluating fatigue detection algorithms rather than a single indicator of driving fatigue.

This work was partially supported by the National Basic Research Program of China (Grant No.2013CB329401), the National Natural Science Foundation of China (Grant No.61272248), and the Science and Technology Commission of Shanghai Municipality (Grant No.13511500200).

Xiang-Yu Gao, Yu-Fei Zhang, Wei-Long Zheng and Bao-Liang Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University and the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.

*Corresponding author: Bao-Liang Lu (bllu@sjtu.edu.cn)

Experiments on a vigilance task were carried out together with recording electro-oculogram (EOG) signals and wearing eye tracking glasses. In our previous work, several features related to fatigue were extracted from EOG and used to construct effective driving fatigue detection models [4],[5]. In this paper, we evaluate these algorithms in terms of $PERCLOS_{ETG}$. We confirm from the evaluation results that $PERCLOS_{ETG}$ can be applied to evaluate fatigue detection algorithms.

II. MEASURES OF FATIGUE

In general, measures of fatigue for evaluating driving fatigue detection algorithms used in current research can be categorized into the following three types: subjective assessment, observer ratings, and performance measures. The characteristics and drawbacks of these three types of measures are described as follows:

- **subjective assessment**: The Karolinska Sleepiness Scale (KSS), which is a 9-step subjective scale, is the most commonly used measure for the subjective self-assessment of sleepiness [6]. Subjective measures like KSS are recorded over relatively long intervals. This is because intrusive feedback would be recorded if subjects are asked to report their fatigue status too frequently. Hence, subjective measures cannot reflect fatigue level continuously and report sudden alertness variation caused by different driving situations.
- **observer ratings**: Trained observer-raters can evaluate the levels of fatigue of drivers, given the videotapes recording the drivers' faces. Experiment results show that such ratings are reliable and consistent, and covary with other indicators of fatigue [7]. However, one of the shortcomings of this method is that there would be a huge amount of workload for observers to rate fatigue levels given hours of videotapes from different individuals.
- **performance measures**: Performance measures such as reaction times are regarded as indicators of fatigue and have been used frequently. Since performance measures are based on specific tasks such as psychomotor vigilance task (PVT) [8], even if they are reliable to evaluate fatigue level, they are unfit to be conducted in real driving situations because of the distraction for driving.

All these measures are not practical in real driving conditions. In this paper, we present a novel measure of driving fatigue by using eye tracking glasses, and apply this new measure to evaluate driving fatigue detection algorithms.

III. PROPOSED MEASURE

A. Eye Tracking Glasses

The advent of wearable devices such as eye tracking glasses (ETG) makes it possible to monitor human cognition reliably and conveniently. We use the SensoMotoric Instruments (SMI) eye tracking glasses [9] in this paper, which are designed for accurate recording of gaze behavior. The glasses use infrared (IR) lighting sources to track the eye gaze and eye movements, and function well in different light conditions. Combining with a high definition scene camera, the glasses track two eyes with automatic parallax compensation at a sample rate of 30Hz. Fig. 1 illustrates the SMI eye tracking glasses used in our experiments.

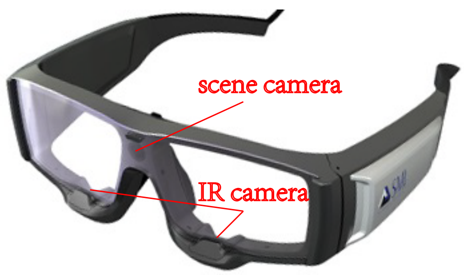


Fig. 1. SMI Eye Tracking Glasses

There are already applications of eye tracking glasses in various fields such as market research, usability, and psychology. However, using eye tracking glasses for driving fatigue detection is new in current studies.

B. PERCLOS by Eye Tracking Glasses

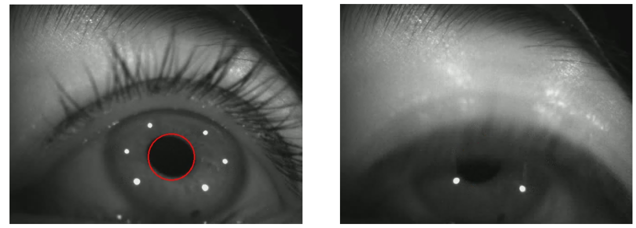
PERCLOS is the percentage of eyelid closure over the pupil over time and reflects slow eyelid closures ("droops") rather than blinks, and it was considered to be the most reliable and valid determination of a driver's alertness level.

The eye tracking glasses divide eye status into three types, i.e. fixations, blinks, and saccades, and can record all the durations of these status based on videos of two eyes. If slow eyelid closures or long-time closures occur, the eyes cannot be tracked, hence are recognized as none of these three status by eye tracking glasses (called 'CLOS'). Fixations and saccades mean that eyes are open, while blinks and 'CLOS' status can be regarded as eye closures over the pupils. Therefore, PERCLOS can be approximated by the percentage of durations of blinks and 'CLOS' over a specified time interval.

$$PERCLOS_{ETG} = \frac{blink + CLOS}{interval} \quad (1)$$

$$interval = blink + fixation + saccade + CLOS \quad (2)$$

However, the eye movement data from eye tracking glasses is provided as 'black boxes', i.e., we don't have the opportunity to examine the underlying image-processing procedures. To be rigorous, we validate the reliability of $PERCLOS_{ETG}$ using an effective pupil detection procedure with the down sampled eye videos provided by SMI.



(a) Open eye (b) Eye closure over the pupil

Fig. 2. Pupil Detection

If the pupil is detected via the pupil detection procedure, eyelids are not covering the pupils. Otherwise, pupils are covered by eyelids. PERCLOS can then be calculated through the results of pupil detection. Fig. 2 illustrates the pupil detection of both open and close status of eyes.

The comparison is performed between the two PERCLOS measures through the experiments described below. The average correlation coefficients come to 0.893 among all subjects, indicating that it is reasonable to use $PERCLOS_{ETG}$ to approximate PERCLOS.

Previous studies usually place cameras in front of drivers, and whether drivers' eyes are open or close can be determined with mature face recognition and face aligned techniques. However, the IR cameras of eye tracking glasses are placed closely to eyes so that the images of two eyes remain stable and reliable under different situations and have few individual differences. Besides, we could do more than just determine the open or close status of eyes with the precise eye movement parameters provided by eye tracking glasses, and we are going to involve these parameters in our future studies. The accuracy and robustness of $PERCLOS_{ETG}$ compared to traditional video-based methods makes it suitable for representing driving fatigue level and evaluating driving fatigue detection algorithms.

IV. EXPERIMENTS AND METHODS

It is reasonable that if subjects have a level of fatigue in which they don't respond to a visual stimulus in time, that level of fatigue is likely to make them have a high risk in real driving environment. Thus, we performed experiments of a vigilance task with visual stimulus, and evaluated an effective fatigue detection algorithm in terms of $PERCLOS_{ETG}$.

A. Experiment Description

Eight sessions from 6 men aged around 20 years old participated in the experiments. Subjects wore the eye tracking glasses. Horizontal and vertical EOG signals were recorded by the NeuroScan system, sampled at 500Hz.

All the subjects were asked to get up early in the morning, and took part in the experiment after lunch in order to induce fatigue during the experiment.

The subjects sit in a comfortable chair, and faced the computer screen in a room with no noise and gentle light. The screen displayed traffic signs of four different colors, i.e. red, yellow, blue, and green, randomly. The traffic signs were presented every 5-7 seconds, and remained on the screen for only 0.5 seconds. Each subject held a pad with four buttons. When a traffic sign appeared, the subject was supposed to push the button of the same color with the sign [2]. The experiment lasted for about an hour, and was easy to induce fatigue because of the monotonousness.

Fig. 3 illustrates the experiment scene.

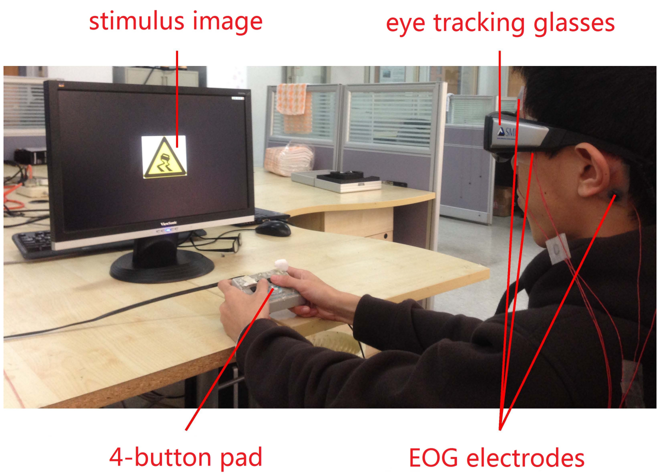


Fig. 3. Experiment Scene

B. *Feature Extraction*

Our previous studies have found the relations between fatigue and features extracted from EOG signals, and these features can be used to construct driving fatigue detection algorithms with good performance. In this paper, we used the blink features, slow eye movement (SEM) features, and rapid eye movement (REM) features extracted from EOG signals using the methods developed in [4],[5].

1) Blinks features: We tried to detect blinks from raw EOG signals. First, the vertical EOG signals were filtered by a low-pass filter with 10Hz. Then we calculated the differences of the signals, and each point means the current change rate of the EOG signals. Several thresholds with carefully consideration were used to mark blinks in EOG signals, and slow blinks were also recognized [4].

After all the blinks were identified, eye blink features were extracted. The features include blink duration, closing time, closure time, reopening time, blink intervals, the velocity of eye closure, the velocity of eye opening, the energy of blinks, and several other statistical features.

2) SEM features: Slow eye movements are related to fatigue. To extract SEM features, EOG signals are first decomposed into components in different frequency bands using discrete wavelet transformation (DWT). Daubechies order 4 (db4) wavelet was used to process EOG signals into 10 levels. Because SEMs are in the range of 0.2-0.6Hz typically, scale 7-10 are necessary to catch the band of SEMs.

Then the energy computation is performed in scale 7-10 using wavelet coefficients, and the energy ratio of the SEM band was obtained. If the result is higher than the specified threshold, the corresponding time period is regarded as an SEM duration.

It is necessary to unify the coefficient resolution of each scale when we compute the energy during a fixed time interval. Scale 6 was used as the standard, and the coefficients were grouped for scale lower than 6 because they are closer in the time axis. For scale higher than six, the coefficients need to be scattered since they are sparse.

A discriminant function judging whether the energy ratio is larger than the threshold was used after we get the energy from each scale [5]. The SEM proportion within each time window was used as the feature.

Additionally, Fourier transformation was also used to get SEMs accurately, the details can be found in [4].

3) REM features: Similar to SEM features, rapid eye movement (REM) features are also closely related to fatigue. The number of REMs, duration of REMs, and energy of REMs were extracted from EOG. We use Fourier transformation and wavelet transformation methods adaptively to extract these features. The details can be found in [4],[5].

All the features extracted were combined to form a feature vector over a fixed interval.

C. *Evaluation*

$PERCLOS_{ETG}$ was calculated with 60 seconds time window, and moving step was 10 seconds. Then sliding average was performed to smooth the results, because fatigue level was not supposed to vary abruptly in the monotonous experiment.

The data from one session was divided into six adjacent segments. The first, third, and fifth segments were used to train the driving fatigue detection model, and the second, fourth, and sixth segments were used to evaluate the model. Different levels of fatigue were expected to distribute evenly between train and test data with the segmentation way. The proportion between training set and test set was $7:3$.

We used support vector machine (SVM) with $PERCLOS_{ETG}$ and features extracted from EOG to construct regression models. The mean squared error was minimized over the training data from each session, respectively. The constructed model then was applied to predict the fatigue level of test segments from the same session. The comparison was performed between the fatigue level represented by $PERCLOS_{ETG}$ and predicted values of fatigue.

Fig. 4 gives a typical example of evaluation results from one session. Three red segments are the predicted fatigue level by the trained model, and blue ones are $PERCLOS_{ETG}$ through the session. It could be observed that the predicted fatigue level is fairly consistent with $PERCLOS_{ETG}$ in this session. Therefore, the fatigue model performed well.

The mean squared errors and correlation coefficients were calculated between $PERCLOS_{ETG}$ and predicted fatigue level by the trained model over the three test segments within the eight sessions. The results are given in Table 1.

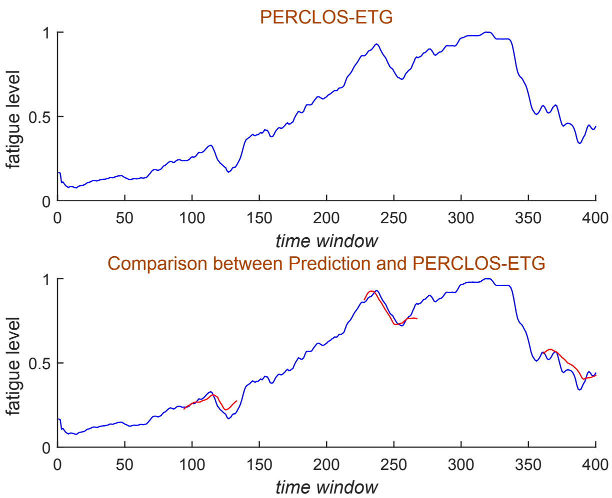


Fig. 4. Evaluation result from one session

TABLE I
EVALUATION RESULTS

Session	mean squared error			correlation coefficients		
	1st	2nd	3rd	1st	2nd	3rd
1	0.0040	0.0086	0.0039	-0.0356	-0.2342	0.5444
2	0.0012	0.0010	0.0024	0.7718	0.9089	0.8446
3	0.0044	0.0029	0.0084	0.6685	0.0065	-0.1539
4	0.0059	0.0043	0.0020	0.5932	-0.3210	-0.0269
5	0.0540	0.0276	0.0034	0.0549	-0.5492	-0.7394
6	0.0189	0.0242	0.0031	0.7243	0.1544	0.2559
7	0.0697	0.0036	0.0043	-0.5419	0.9056	0.7565
8	0.0093	0.0058	0.0042	0.6146	0.7308	0.8920

The mean squared errors are fairly small in almost all the sessions, and are below 0.01 in all the test segments of the 1,2,3,4,8 sessions, which is an acceptable performance for the fatigue level scale ranged from 0 to 1. In regard to correlation coefficients, half of the test segments produce the values higher than 0.5. Generally, the importance of correlation coefficients is not as important as mean squared errors in the evaluation of fatigue detection methods. Therefore, the fact that the correlation results are not as good as the mean squared errors is also acceptable.

We checked the data of sessions with high mean squared errors, and we found that it might be caused by the division strategy for training and testing segments. The sessions with poor evaluation results were divided into six segments in a way that the subjects were alert most of the time in training segments, while they were sleepy in terms of $PERCLOS_{ETG}$ most of the time in testing segments, leading to distribution differences between training segments and testing segments.

V. DISCUSSION AND FUTURE

Actually, our real purpose is not to prove the performance of the EOG-based fatigue detection algorithm, but to make an attempt to evaluate fatigue detection algorithms based on the proposed $PERCLOS_{ETG}$. Additionally, the acceptable evaluation results could justify this role of $PERCLOS_{ETG}$ to some extent.

The eye tracking glasses also provide many other parameters possibly related to human fatigue, such as blink frequency, fixation duration, and saccade speed. The gaze behaviors of drivers recorded by eye tracking glasses also provide important information about driving fatigue. If the driver is dull-looking or doesn't look at the points that they are supposed to, he or she is very likely in a state of fatigue. Combining $PERCLOS_{ETG}$ with other eye movement parameters and information of gaze points is supposed to provide more reliable measures of driving fatigue level, and this is under consideration in our future work.

Although eye tracking glasses are fairly convenient and there are also driving fatigue detection products including sensors in glasses [10], remote cameras are presumably favored as it is less obtrusive. Besides, because of the high cost and inconvenience for drivers who already wear glasses, eye tracking glasses are not aimed to detect driving fatigue as the final product. We think eye tracking glasses are more suitable for the evaluation of driving fatigue detection methods.

All kinds of sensors can be assembled on real cars and wearable devices such as smart bracelet can be worn by drivers. We could then collect all these data in real driving environment, label the driving fatigue level with eye tracking glasses, and train driving fatigue detection models using machine learning approaches. This is the vision of our future research.

REFERENCES

[1] Y.-C. Dong, Z.-C. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596-614, Jun. 2011.

[2] L.-C. Shi, B.-L. Lu, "EEG-based vigilance estimation using extreme learning machines," *Neurocomputing*, vol. 102, pp. 135-143, Feb. 2013.

[3] D. F. Dinges, and R. Grace, "PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance," *Fed. Highway Admin., Office Motor Carriers, Washington, DC, Tech. Rep.*, 1998.

[4] Z.-P. Wei, B.-L. Lu, "Online vigilance analysis based on electrooculography," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 17, 2012.

[5] J.-X. Ma, L.-C. Shi, and B.-L. Lu, "Vigilance Estimation by Using Electrooculographic Features," in *32nd Annual International Conference of the IEEE EMBS*, pp. 6591-6594, Jan. 2010.

[6] K. Kaida, M. Takahashi, T. kerstedt, A. Nakata, Y. Otsuka, T. Haratani, and K. Fukasawa, "Validation of the Karolinska sleepiness scale against performance and EEG variables," *Clinical Neurophysiology*, vol. 117, no. 7, pp. 1574-1581, 2006.

[7] W. W. Wierwille, and L. A. Ellsworth. Evaluation of driver fatigue by trained raters, *Accident Analysis & Prevention*, vol. 26, no. 5, pp. 571-581, 1994.

[8] D. F. Dinges, and J. W. Powell, "Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations," *Behavior Research Methods, Instruments, & Computers*, vol. 17, no. 6, pp. 652-655, 1985.

[9] [Online]. Available: www.smivision.com

[10] M. W. Johns, A. Tucker, R. Chapman, K. Crowley, and N. Michael, "Monitoring eye and eyelid movements by infrared reflectance oculography to measure fatigue in drivers," *Somnologie-Schlafforschung und Schlafmedizin*, vol. 11, no. 4, pp. 234-242, Jul. 2007.