

On Using Drivers' Eyes to Predict Accident-Causing Drowsiness Levels

Alyssa Byrnes and Cynthia Sturton

Abstract—We examine the use of video data to determine a driver's drowsiness level. We conduct a user study to collect video of a user reading, watching a driving simulation, and playing a video game that simulates driving. Alongside each video is the user's sleepiness as measured by the Stanford Sleepiness Scale, the Epworth Sleepiness Scale, and eight questions that have been shown to coincide with unsafe driving. Using this data, we replicate the results of prior art, showing that on average, changes in eye movement do correlate with drowsiness. We find however, that the measurements appear to have no predictive value for the drowsiness metrics that are known to coincide with unsafe driving. We determine that additional research in detecting driver drowsiness is needed. Our user study data is publicly available.

I. INTRODUCTION

This paper presents an analysis of the use of eye measurements of a driver in order to determine accident-causing drowsiness levels. In level 3 semi-autonomous cars, control is switched between the driver and the vehicle [1]. One challenge for these vehicles is deciding when the car should be in control and when the user should be in control. Various aspects of a driver's mental state, such as driver attentiveness, drowsiness, stress, and sobriety are used to determine if the driver is ready to be given control. In this work, we focus on monitoring driver drowsiness, as it plays a key role in driver safety [2], [3].

Many controllers have been developed to monitor drowsiness. Some systems involve attaching pulse sensors and electrodes to a driver to monitor heart rate, blink rate, brain signals, and other physical indicators of drowsiness [4], [5]. However, attaching devices to the driver is not appealing to the consumer. Therefore, the most marketable method for monitoring a driver's drowsiness is via video camera. The majority of video based techniques use the driver's eyes – blink duration, blink frequency, and percentage eye closure – to measure drowsiness. The results overall have been positive [6], [7], [5], [8], [9], [10].

This material is based upon work supported by the National Science Foundation under Grant No. 1544924.

The authors are affiliated with the University of North Carolina at Chapel Hill. {abyrnes1, csturton}@cs.unc.edu.

In this paper, we examine the prior work done in monitoring a driver's eye information to determine drowsiness levels. While some work shows correlations between eye information and drowsiness, they do not show that eye information has any predictive value [7], [9]. Other works have shown success in detecting drowsiness using different drowsiness metrics, but the metrics used have not been shown to coincide with vehicle accidents [6], [8].

We designed and conducted a user study to create a dataset for testing. We filmed a user reading, watching a driving simulation, and playing a driving simulation video game using steering wheel and foot pedal controllers. Along with each video is the user's sleepiness as measured by the Stanford Sleepiness Scale [11], the Epworth Sleepiness Scale [12], and their responses to eight questions regarding their sleep habits that have been shown to coincide with safe driving [2], [3].

We later collected the eye landmark information of each user's video performance in the driving simulation. To minimize possibility of error related to incorrect eye tracking, these landmarks were collected by hand rather than automatically. In our analysis, we first replicate the results of prior art, showing that on average blink duration is shorter, percentage eye closure (PERCLOS) is higher, and blink rate is lower for drivers that are more drowsy. However, we then find that although these measures correlate on average with drowsiness, there is sufficient variability that these measures do not provide strong predictive value.

Our dataset is publicly available. To our knowledge, it is the first publicly available dataset of its kind. Additionally, the script and the survey used in the user study are available to encourage comparative work.

II. BACKGROUND

A. Drowsiness Scales

The *Epworth Sleepiness Scale* (ESS) is designed to measure a person's average sleepiness levels while conducting day to day activities [12]. A user answers eight questions about their daily sleepiness levels and provides a score from 0–3 for each question, 0 indicating they would never fall asleep during a particular

activity and 3 indicating there is a high chance they would fall asleep during the activity. The scores are totaled for an overall ESS Score of 0–24. A score of 10 or higher indicates a person has mild to severe excessive daytime sleepiness.

The *Stanford Sleepiness Scale* (SSS) is an introspective measure of a person’s current sleepiness level. The user provides a score between 1–7 to indicate their current drowsiness level. A score of 1 means they are “feeling active, vital, alert, or wide awake” and a score of 7 means they are “no longer fighting sleep, sleep onset soon; having dream-like thoughts” [11].

B. Accident-causing Drowsiness Metrics

TABLE I
SUMMARY OF PREVIOUS WORK DETERMINING WHAT SLEEPINESS METRICS CORRELATE WITH ACCIDENTS

	Epworth	Stanford	Hours Sleep
Connor et al. [2]	✗	✓	✓
Stutts et al. [3]	✓	NA	✓

Previous work has shown that driver drowsiness significantly increases the likelihood of an accident. A summary of this work is shown in Table I. A check means the study found the metric to be a good indicator of likelihood of an accident; an X means it did not.

In 1999, Stutts et al. interviewed 1,403 drivers [3]. 467 of those drivers had been involved in a police-reported accident in North Carolina who were identified as “asleep” or “fatigued” by the officer, 529 of those drivers were in a police-reported accident in North Carolina but were not reported to be asleep or fatigued, and 407 of the drivers were not involved in a recent accident. They used multiple logistic regression models to produce estimates of the odds ratio for the occurrence of a sleep-related crash given a particular risk factor, adjusted for driver age and gender. They found that the drivers who got in sleep-related crashes were more than twice as likely to have Epworth Sleepiness Scores greater than 10. They also found a high percentage of sleep-related crash drivers had less than six hours of sleep the night before when compared to the drivers who were in crashes that were not sleep related.

In 2002, Connor et al. surveyed 571 drivers who had been involved in a crash in which someone was admitted to the hospital or killed [2]. They recruited 588 additional drivers to act as their control group. They found an increased likelihood of an injury-causing accident with subjects who had a Stanford Sleepiness

Score of 4 or above, who had five or fewer hours of sleep in the past 24 hours, or were driving between 2 and 5 A.M. They found no increase in likelihood of an accident with drivers who had an Epworth Sleepiness Scale measure greater than or equal to 10 compared to drivers with an Epworth Sleepiness Scale less than 10.

Both studies showed “hours of sleep” is an indicator of accident-causing drowsiness. Connor et al. found the Stanford Sleepiness Scale to be an indicator, but the studies conflicted on whether or not the Epworth Sleepiness Scale is. We include all three metrics in our user study to ensure completeness of the resulting data.

C. Using Eye Data to Measure Drowsiness

Table II summarizes previous work using eye data to measure driver drowsiness. In 2003, Caffier et al. found that blink duration, eyelid closing time, reopening time, proportion of long closure duration blinks, and blink frequency are on average higher when a person is drowsy [9]. The study used 60 volunteers who filled out a survey about their sleepiness levels. They then wore glasses that captured eye data while performing a driving simulation.

In 2005, Ingre et al. found that on average, blink duration increases with drowsiness, but blink duration differs widely between individuals [7]. In the study ten subjects performed a two-hour driving simulation while reporting their drowsiness levels on the Karolinska Sleepiness Scale (KSS) every five minutes. Eye data was collected using an Electrooculogram (EOG).

In 2009, Shuyan et al. detected drowsiness using eye data [8]. Study participants performed a driving simulation while drowsiness was measured using an Electroencephalogram (EEG), the Karolinska Drowsiness Score (KDS), and KSS. An EOG was used to collect 11 eyelid related features including upper threshold, lower threshold, open time, close time, blink amplitude, and blink duration, which were used to train and test a support vector machine. They showed a correct detection rate of 100% when the driver was very sleepy, 87% when the driver was sleepy, and 83% when they were awake.

In 2012, Lee et al. found that the percentage of a person’s eye closure (PERCLOS) and average blink duration can be used to detect drowsiness with a 55% false detection rate and an 82% correct detection rate using a Bayesian neural network [6]. The study used ten volunteers who filled out a survey about basic drowsiness indicators and performed a driving simulation. Eye data was collected from video of the driver.

TABLE II
SUMMARY OF PRIOR WORK ON HOW EYE DATA INDICATES DROWSINESS

	Eye Data Collected	Sleepiness Metric	Prediction of Sleepiness Demonstrated	Metric Correlates with Accident Likelihood
Caffier et al. [9]	Blink Duration, Closing Time, Reopening Time, Proportion of Long Closure Duration Blinks, and Blink Frequency	Survey	✗	✗
Lee et al. [6]	Percentage Eye Closure and Blink Duration	Survey	✓	✗
Ingre et al. [7]	Standard Deviation of the Lateral Position and Blink Duration	KSS	✗	✗
Shuyan et al. [8]	Upper Threshold, Lower Threshold, Open Time, Close Time, Blink Amplitude, and Blink Duration	EEG, KDS, KSS	✓	✗

III. METHODOLOGY

A. Data Collection

We recruited 23 subjects through our university’s user study recruitment tool. The subjects ranged in age from 26–55. Of the 23 participants 52% identified as female and 48% identified as male; 13 participants identified as White, 4 identified as Asian, 2 identified as Indian, 1 identified as African American, 1 identified as Hispanic, and 2 identified as other.

Each participant filled out a survey to indicate their current and average drowsiness levels. The questions on the survey include the participant’s Epworth Sleepiness Score, Stanford Sleepiness Score, and how much sleep they had gotten in the past 24 hours. Then the subjects’ faces were filmed while they performed the following activities: reading a passage [13], watching a video of a driving simulation [14], and performing a driving simulation. The survey and filming took less than 30 minutes per participant.

During driving simulation users drove in “freeplay mode” in Driving Simulator 2013 using the Logitech Driving Force G29 Racing Wheel for PlayStation 4 and PlayStation 3. The subjects were given three minutes to acclimate to the simulator before we began filming. The video was recorded at 300 frames per second.

The script used to administer the study, the passages and videos used by the participants, and the surveys are all included in the dataset, which is freely available to other researchers.

B. Data Analysis

We analyzed the first 100 frames of the driving simulation video for each of our 24 subjects. We ex-



Fig. 1. Image of experimental setup

plored algorithmic options including supervised descent method (SDM) [15] and online eye tracking software. However, SDM did not provide sufficient accuracy when drivers’ eyes were partially closed, and online eye tracking software did not provide the necessary API access. Therefore, for each 720×1280 frame, we manually recorded the coordinate location of the top of the left eyelid and the bottom of the left eyelid (see Figure 2).

We then calculated *blink duration*, *percentage eye closure*, *blink rate*, and *percentage eye movement*.

- *Blink duration* is the number of frames it takes for a user to blink.
- *Percentage eye closure (perclos)* is the percentage the eye is closed given a baseline eye size.



Fig. 2. Left: Example of landmark detection using the Supervised Descent Method; Right: Example of landmarks collected manually

It is calculated by $\frac{1 - \text{measured_distance}}{\text{eye_size}}$, where *measured_distance* is the distance between the upper and lower eyelid, and *eye_size* is the baseline eye size as determined by the largest distance each subject had their eyes open in the 100 frames.

- *Blink rate* is the number of blinks that occurred per 100 frames.
- *Percentage eye movement* is the percentage of frames where the upper and lower eyelid distance changes by more than 1 pixel.

A blink starts when a person's perclos is greater than 70% and stops when a person's perclos is less than or equal to 70%.

We analyzed the video data against seven definitions of *Awake*. (If a participant is not *Awake*, they are *Drowsy*.) The first three come from the Epworth Sleepiness Scale score (ESS), the Stanford Sleepiness Scale score (SSS), and the hours of sleep the participants had in the past 24 hours (HS):

- ESS. The driver has an ESS score ≤ 10 .
- SSS. The driver has an SSS score < 4 .
- HS. The driver received more than 6 hours of sleep in the past 24 hours.

The last four definitions are combinations of the above three. For example, under the ESS\SSS definition, a participant had to have an Epworth Sleepiness Scale score of ≤ 10 AND a Stanford Sleepiness Scale score < 4 to be considered *Awake*, otherwise they would be considered *Drowsy*.

We measured both correlation between eye data and drowsiness, and the predictive value of the eye data for determining drowsiness.

To measure predictive value of the eye measurements, we trained a support vector machine (SVM) on a training set of randomly chosen participants and tested its accuracy of predicting awakeness or drowsiness in the remaining participants. For blink duration, perclos, and percentage movement, the SVM trained on 15

Correlation of Drowsiness Measures

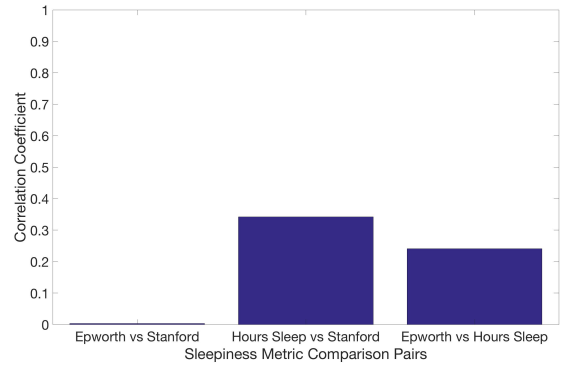


Fig. 3. Each bar represents represents the absolute value of the correlation between drowsiness metrics.

Correlation of Eye Information

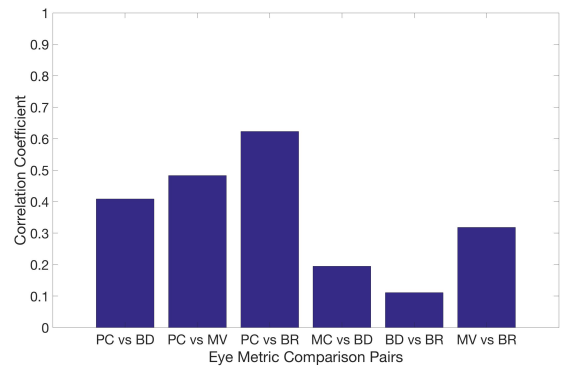


Fig. 4. Each bar represents the absolute value of the correlation between the two eye measurements.

images and tested on 8. Since only 12 participants blinked in the first 100 frames, the training set for blink rate was 8 and the testing set was 4. To account for the unbalanced data set we added a penalty to the SVM for missed detections.

IV. RESULTS

Table III shows the average value of each eye measurement per *Awake* person and *Drowsy* person, using the seven definitions of *Awake* and *Drowsy*.

Table IV shows the number of *Awake* participants according to each of the seven definitions of *awake*. Less than half of participants qualified as *Awake* under the $ESS \wedge SSS \wedge HS$ definition.

Figure 3 shows the absolute value of the 2-dimensional correlation between each subject's drowsiness metrics. There was not a high correlation between any of the drowsiness metrics.

TABLE III

AVERAGE VALUE OF EACH EYE MEASUREMENT FOR DROWSY AND AWAKE DRIVERS. THIS REPLICATES THE WORK OF CAFFIER ET AL. [9] AND INGRE ET AL. [7]

Sleepiness Metric	Blink Duration		Percentage Eye Closure		Blink Rate		Percentage Movement	
	Awake	Drowsy	Awake	Drowsy	Awake	Drowsy	Awake	Drowsy
ESS	2.8889	2.3333	11.35%	11.73%	.5882	.6667	7.62%	9.42%
SSS	2.75	NA	11.53%	10.53%	.6667	0	8.7%	1.68%
HS	2.8889	2.3333	10.96%	13.19%	.5556	.8	7.47%	1.03%
ESS \wedge HS	3.1429	2.2	11.16%	11.82%	.6154	.6	6.7%	9.9%
SSS \wedge HS	2.8889	2.3333	11.02%	12.43%	.625	.5714	8.2%	7.84%
ESS \wedge SSS	2.8889	2.3333	11.46%	11.43%	.6667	.5	8.41%	7.49%
ESS \wedge SSS \wedge HS	3.1429	2.2	11.27%	11.61%	.7273	.5	7.61%	8.53%

TABLE IV

NUMBER OF AWAKE DRIVERS PER DEFINITION

Definition of Awake	Number of Users
ESS	17
SSS	21
HS	18
ESS \wedge HS	16
SSS \wedge HS	16
ESS \wedge SSS	15
ESS \wedge SSS \wedge HS	11

Figure 4 shows the correlation between two different eye measurements for each user. Perclos and Blink Rate have a strong correlation, and Perclos and Percentage Eye Movement have a strong correlation. Therefore we expect that the accuracy of drowsiness predicted using Perclos will be similar to the accuracy predicted using Blink Rate and Percentage Eye Movement.

Figure 5 shows the histogram values of users' percentage eye movement. They were classified as awake or drowsy based on their ESS score. We include this as an example of the high overlap of eye metric values in sleepy and awake users.

Figure 6 shows the percentages of detection accuracy, false positives, and missed detections for determining ESS-based drowsiness using every combination of eye metrics. Similarly, Figure 7 shows the percentages of detection accuracy, false positives, and missed detections for determining Hours-of-Sleep-based drowsiness using every combination of eye metrics. Accuracy is the number of correct drowsy and awake classifications divided by the total number of users. False positive is the number of awake users marked as drowsy divided by the number of awake users. Missed detection is the number of drowsy users marked as awake divided by the number of drowsy

Histogram of Percentage Eye Movement Results

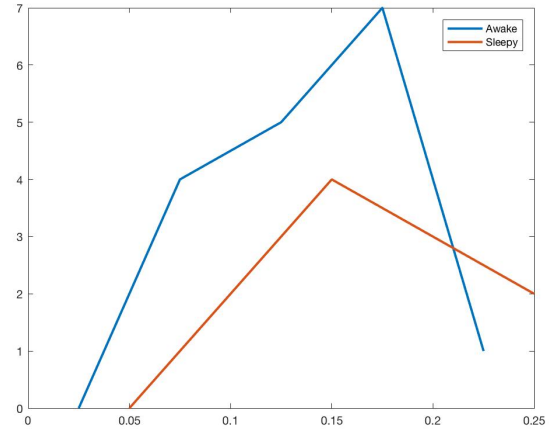


Fig. 5. Each line represents the histogram data of different users' percentage eye movement values for users that were classified as either awake or drowsy using ESS.

users.

V. DISCUSSION

As is shown in Table III, blink duration was always decreased when the driver was drowsy and percentage eye closure was higher when the driver is drowsy except for when considering just SSS or ESS \wedge SSS. It can also be seen that the blink rate decreases when the driver is drowsy, except for when considering ESS or HS. We found no correlation between drowsiness and percentage eye movement.

As seen in Figure 3, the drowsiness metrics are not highly correlated, so it is important to consider all three measurements.

As seen in Figure 4, Perclos and Blink Rate are strongly correlated, and Perclos and Movement are

Balanced SVM Drowsiness Detection Accuracies - ESS



Fig. 6. Each bar represents the percentage of each success measure of the SVM using a combination of eye metrics for training. The color of the bar denotes what success measure was used.

Balanced SVM Drowsiness Detection Accuracies - Hours of Sleep

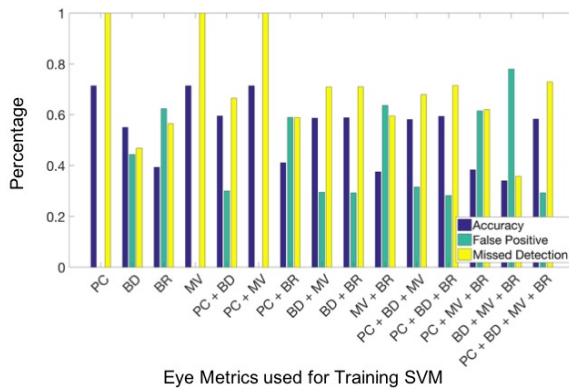


Fig. 7. Each bar represents the percentage of each success measure of the SVM using a combination of eye metrics for training. The color of the bar denotes what success measure was used.

moderately correlated. This leads us to believe that just looking at one of these measurements instead of all three could be sufficient in predicting a driver’s drowsiness.

Figure 5 demonstrates an argument against the predictive value of the eye metrics. The histogram for percentage eye movement in drowsy users highly overlaps the histogram for percentage eye movement in awake users. This suggests that classification of a user as awake or drowsy would be extremely difficult if not impossible using these metrics.

As seen in figures 6 and 7, overall accuracy ranged from around 40–60%. The cases where accuracy was greater than 60% only occurred when there was a missed detection rate of 100%. This result demonstrates that these eye metrics don’t have any predictive value

for determining accident-causing drowsiness levels.

VI. CONCLUSION

We extended prior work in showing that eye measurements correlate with driver drowsiness when using drowsiness measures that have been shown to indicate increased risk of accidents. We then showed that these measurements appear to have little to no predictive value for the studied drowsiness measures. We conclude that more research is needed before drowsiness detection from eye data can be reliably used. We make our data set freely available to other researchers to encourage this research.

REFERENCES

- [1] “Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles,” lvm.org/docs/LangRef.html, SAE International, Nov. 2016.
- [2] J. Connor, R. Norton, S. Ameratunga, E. Robinson, I. Civil, R. Dunn, J. Bailey, and R. Jackson, “Driver sleepiness and risk of serious injury to car occupants: population based case control study,” *BMJ*, vol. 324, no. 7346, 2002.
- [3] J. C. Stutts, J. W. Wilkins, and B. V. Vaughn, “Why do people have drowsy driving crashes,” 1999.
- [4] A. Reyes-Muñoz, M. C. Domingo, M. A. López-Trinidad, and J. L. Delgado, “Integration of body sensor networks and vehicular ad-hoc networks for traffic safety,” *IEEE Sensors Journal*, vol. 16, no. 1, p. 107, 2016.
- [5] Y. Sun and X. B. Yu, “An innovative nonintrusive driver assistance system for vital signal monitoring,” *IEEE journal of biomedical and health informatics*, vol. 18, no. 6, 2014.
- [6] B.-G. Lee and W.-Y. Chung, “Driver alertness monitoring using fusion of facial features and bio-signals,” *IEEE Sensors Journal*, vol. 12, no. 7, pp. 2416–2422, 2012.
- [7] M. Ingre, T. Åkerstedt, B. Peters, A. Anund, and G. Kecklund, “Subjective sleepiness, simulated driving performance and blink duration: examining individual differences,” *Journal of sleep research*, vol. 15, no. 1, pp. 47–53, 2006.
- [8] S. Hu and G. Zheng, “Driver drowsiness detection with eyelid related parameters by support vector machine,” *Expert Systems with Applications*, vol. 36, no. 4, 2009.
- [9] P. P. Caffier, U. Erdmann, and P. Ullsperger, “Experimental evaluation of eye-blink parameters as a drowsiness measure,” *European journal of applied physiology*, vol. 89, 2003.
- [10] I. Park, J.-H. Ahn, and H. Byun, “Efficient measurement of eye blinking under various illumination conditions for drowsiness detection systems,” in *ICPR*. IEEE, 2006.
- [11] E. Hoddes, V. Zarcone, and W. Dement, “Development and use of stanford sleepiness scale (SSS),” in *Psychophysiology*, vol. 9, no. 1, 1972, p. 150.
- [12] R. D. Chervin, “Epworth sleepiness scale?” 2003.
- [13] E. A. Taub, “Sleepy behind the wheel? some cars can tell,” *The New York Times*, p. B4, 3 2017.
- [14] T6Games, “Driving simulator 2013 gameplay,” Published online, 4 2014, youtube video.
- [15] X. Xiong and F. De la Torre, “Supervised descent method and its applications to face alignment,” in *Computer Vision and Pattern Recognition*. IEEE, 2013.