Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines

Yulan Liang, Michelle L. Reyes, and John D. Lee

***Abstract*—As use of in-vehicle information systems (IVISs) such as cell phones, navigation systems, and satellite radios has increased, driver distraction has become an important and growing safety concern. A promising way to overcome this problem is to detect driver distraction and adapt in-vehicle systems accordingly to mitigate such distractions. To realize this strategy, this paper applied support vector machines (SVMs), which is a data mining method, to develop a real-time approach for detecting cognitive distraction using drivers' eye movements and driving performance data. Data were collected in a simulator experiment in which ten participants interacted with an IVIS while driving. The data were used to train and test both SVM and logistic regression models, and three different model characteristics were investigated: how distraction was defined, which data were input to the model, and how the input data were summarized. The results show that the SVM models were able to detect driver distraction with an average accuracy of 81.1%, outperforming more traditional logistic regression models. The best performing model (96.1% accuracy) resulted when distraction was defined using experimental conditions (i.e., IVIS drive or baseline drive), the input data were comprised of eye movement and driving measures, and these data were summarized over a 40-s window with 95% overlap of windows. These results demonstrate that eye movements and simple measures of driving performance can be used to detect driver distraction in real time. Potential applications of this paper include the design of adaptive in-vehicle systems and the evaluation of driver distraction.**

***Index Terms*—Classification, driving performance, eye movement, logistic regression, secondary task, support vector machine (SVM).**

I. Introduction

DRIVER DISTRACTION is an important safety problem. Between 13% and 50% of crashes are attributed to driver distraction, resulting in as many as 5000 fatalities and $40 billion in damages each year [1]–[3]. Increasing use of in-vehicle information systems (IVISs) such as cell phones, GPS navigation systems, and satellite radios has exacerbated the problem by introducing additional sources of distraction [4]–[8]. Enabling drivers to benefit from IVISs without diminishing safety is an important challenge.

Manuscript received March 28, 2006; revised December 17, 2006 and January 19, 2007. This work was sponsored by the National Highway Traffic Safety Administration (NHTSA) (Project Manager: M. Perel) and administered by the John A. Volpe National Transportation Systems Center (Project Manager: M. D. Stearns). The Associate Editor for this paper was Y. Liu.

Y. Liang and J. D. Lee are with the Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA 52242 USA (e-mail: Yulan-liang@uiowa.edu; jdlee@engineering.uiowa.edu).

M. L. Reyes is with the Center for Computer-Aided Design, The University of Iowa, Iowa City, IA 52242 USA (e-mail: mlries@engineering.uiowa.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2007.895298

One promising strategy involves classifying the driver state in real time and then using this classification to adapt the in-vehicle technologies to mitigate the effects of distraction [9]. This strategy is not new. For example, "attentive autos," which monitor driver attention and emit an alert when the driver looks away from the road or when driving demands require a high level of attention, have been studied [10]. Smith *et al.* [11] developed a robust system using head orientation and eye blinks to monitor decrements in visual attention due to fatigue while driving. The degree of driver stress [12] and vigilance [13] has been predicted from physiological measures and used to help manage IVIS functions. Also, studies have used advanced data mining techniques to predict lane-change intent [14]–[16] so that driver-assistance systems can take driver intention into consideration. Obviously, measuring driver state in real time is a core function in such systems. Often, however, the means of ascertaining this state are often intrusive (e.g., measuring galvanic skin response using electrodes). This paper presents an unobtrusive approach that uses eye movements and driving behavior to detect driver cognitive distraction.

A. Driver Distraction and Eye Movements

Visual and cognitive distractions are two major types of driver distraction. Both can degrade driving performance (e.g., [4]–[7]). Visual distraction is straightforward, occurring when drivers look away from the roadway (e.g., to adjust a radio); it can be reasonably measured by the length and frequency of glances away from the road [6]. Unlike visual distraction, cognitive distraction occurs when drivers think about something not directly related to the current vehicle control task (e.g., conversing on a hands-free cell phone or route planning). Some studies have shown that cognitive distraction undermines driving performance by disrupting the allocation of visual attention to the driving scene and the processing of attended information. Cognitive distraction impaired the ability of drivers to detect targets across the entire visual scene and caused gaze to be concentrated in the center of the driving scene [17]. Cognitive distraction associated with cell phone conversations negatively affected both implicit perceptual memory and explicit recognition memory for items that drivers looked at while driving [7]. Because the mental state of drivers is not observable, no simple measure can index cognitive distraction precisely [18]. Currently, researchers use primary and/or secondary task performance measures, subjective measures, and physiological measures—among which eye movements are one of the most promising—to assess driver distraction.

Three fundamental types of eye movements help define eye movement parameters that might reflect allocation of visual attention and consolidation of fixated information. These movements are fixation, saccade, and smooth pursuit. Fixations occur when an observer's eyes are nearly stationary. The fixation position and duration may relate to attention orientation and the amount of information perceived from the fixated location, respectively [19]. Saccades are very fast movements that occur when visual attention shifts from one location to another. Smooth pursuits occur when the observer tracks a moving object, such as a passing vehicle. They serve to stabilize the object on the retina so that visual information can be perceived while the object is moving relative to the observer. In the context of driving, smooth pursuits have a particularly important function; they capture information in the dynamic driving scene as fixations do in a static scene. Thus, in order to account for both of the ways in which drivers acquire visual information, both fixations and smooth pursuit movements need to be considered.

Some studies have shown links between eye movement, cognitive workload, and distraction [19]. Saccade distance decreases as tasks become increasingly complex, which indicates that saccades may be a valuable index of mental workload [20]. Rantanen and Goldberg [21] found that visual field shrank 7.8% during a moderate-workload counting task and 13.6% during a cognitively demanding counting task. Similarly, increased cognitive demand during driving decreased spatial gaze variability, defined by the area covered by one standard deviation (s.d.) of gaze location in both the horizontal and vertical directions [17], [22]. These links between eye movement and cognitive load show that eye movement measures are good candidates to predict cognitive distraction.

Although some studies have related cognitive distraction to eye movements and disruptions in visual attention, little research has considered how eye movement data may be used to detect such distraction in real time [18]. Furthermore, most studies [7], [17], [20]–[22] consider the relationship between cognitive distraction and eye movement using linear, univariate approaches (e.g., linear regression and ANOVA). This paper develops a method of real-time detection of cognitive distraction and degraded driving performance using support vector machines (SVMs), which can capture nonlinear relationships and the interaction of multiple measures.

B. Support Vector Machines

First proposed by Vapnik [23], SVMs are based on the statistical learning technique and can be used for pattern classification and inference of nonlinear relationships between variables [23], [24]. This method has been successfully applied to the detection, verification, and recognition of faces, objects, handwritten characters and digits, text, speech, and speakers, and the retrieval of information and images [25].

Fig. 1 shows the basic idea of classification using SVMs in 2-D space. Labeled binary-class training data $D = \{(x_i, y_i)\}_{i=1}^{l}$, where x_i is a vector containing multiple features, and y_i is a class indicator with value either -1 or 1, are illustrated as circles and dots in Fig. 1, respectively.

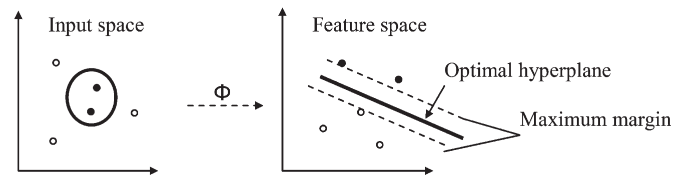


Fig. 1. SVM algorithm [25].

They are mapped onto a high-dimensional feature space via a function Φ. When the mapped data are linearly separable in the feature space, a hyperplane maximizing the margin from it to the closest data points of each class exists. The hyperplane yields a nonlinear boundary in the input space. The maximum margin represents the minimized upper bound of generalization error. The function Φ is written in the form of a kernel function $K(x_i, x_j) = \Phi(x_i)^{\mathrm{T}}\Phi(x_j)$ used in the SVM calculation. When data are not linearly separable in the feature space, the positive penalty parameter C [26] allows for training error ε by specifying the cost of misclassified training instances.

The learning technique of the SVM method makes it very suitable for measuring the cognitive state of humans. First, human cognition can seldom be represented by a linear model. SVMs can generate both linear and nonlinear models and are able to compute the nonlinear models as efficiently as the linear ones. Second, SVMs can extract information from noisy data [25] and do not require prior knowledge before training. Although nobody knows exactly how cognitive distraction affects eye movement patterns and impairs driving performance, we believe that systematic relationships do exist, and that the SVM method may be able to extract these relationships. Third, the SVM method avoids overfitting by minimizing the upper bound of the generalization error [27] to produce more robust models than traditional learning methods (e.g., logistic regression), which only minimize training error. The SVM method may have computational difficulty if the volume of training data becomes great. However, in general, the SVM method is a robust technique for detecting changes in human behavior and represents a particularly promising approach to detecting cognitive distraction given the complex dynamics of eye movement.

To evaluate the proposed method of using SVMs to detect driver distraction and degraded driving performance in real time and to work toward the implementation of such a system, this paper evaluates the detection performance of SVMs, compares SVM performance with that of logistic regression, and discusses the effects of three SVM model characteristics on performance. Testing accuracy and signal detection theory measures of sensitivity and response bias are used to assess the models. Based on the characteristics of SVMs and eye movement, we expect that the SVM method will be able to detect cognitive distraction based on eye movement and driving measures and that this method will outperform logistic regression.

II. Model Construction

Eye and driving data were collected in a simulator experiment in which ten participants performed three driving tasks while interacting with an IVIS. Raw eye data were classified into fixation, saccade, and smooth pursuit eye movements, and measures of the eye movements were calculated. Three driving performance measures—steering wheel angle, lane position, and steering error transformed from steering wheel angle [28]—were also obtained. SVM models were trained for each participant with different values for three characteristics, including different definitions of the model outputs (distraction definition), different combinations of input variables (feature combination), and different summarizing parameters for inputs (window size with overlap between windows). The resulting models generated binary predications of the state of distraction (i.e., distracted or not distracted). Testing accuracy, model sensitivity, and response bias were used to measure and compare model performance.

A. Data Source

1) Participants: Ten drivers (six male and four female) with normal or corrected to normal vision participated in this experiment. They were between the ages of 35 and 55 $(\text{mean} = 45, \text{sd} = 6.6)$, possessed a valid U.S. driver's license, had at least 19 years of driving experience $(\text{mean} = 30)$, and drove at least five times a week. Participants were compensated \$15/h and could earn a bonus (up to \$10) based on their performance of the secondary task.

2) Driving Task: The experiment took place in a fixed-based medium-fidelity driving simulator. The driving scene was displayed at 768×1024 resolution on a rear-projection screen located approximately 2 m in front of the drivers, producing approximately 50° of visual field. The simulator collected data at 60 Hz.

Participants drove along a straight suburban street with two lanes in each direction. The subject vehicle (SV; vehicle driven by the participants) was equipped with a simulated cruise control system that engaged automatically at 45 mi/h and disengaged when drivers pressed the brake pedal. The participants were instructed to follow the vehicle in front of them [lead vehicle (LV)] and to use the cruise control as much as possible. The LV was coupled to the SV by a 1.8-s tailway.

The participants performed three driving tasks during each of the six 15-min drives. The first task was to follow the LV and respond to six LV braking events during each drive. The timing of each braking event was determined by the status of the IVIS task and was initiated by the researcher. During the events, the LV braked at a rate of 0.2 g until it reached a minimum speed of no more than 20 mi/h and the participant had braked at least once. Following a brief, random delay (0 to 5 s), the LV accelerated at a rate of 0.25 g until it reached a speed of 25 mi/h. Then, the LV was coupled to the SV, again with a 1.8-s tailway. The second task was to keep the SV from drifting toward the lane boundaries and to drive in the center of the lane as much as possible. The final task was to detect the appearance of a bicyclist on the right side of the road in the driving scene by pressing a button on the steering wheel. The bicyclist appeared about three times per minute and was visible, on average, for approximately 2.8 s.

3) IVIS Task: During four of the six drives, participants interacted with the IVIS: an auditory stock ticker. The stock ticker was composed of three-letter stock symbols (e.g., BYS, NUD, VBZ, etc.) each followed by its value (a whole number from 1 to 25). The participants were instructed to keep track of the values of the first two stocks (i.e., the target stocks) presented during each interaction. Each time the drivers heard one of the target stock symbols, they determined whether the value of that stock had increased or decreased since the last time they had heard it and pressed the corresponding button on the steering wheel. At the end of the interaction with the IVIS, drivers identified the overall trend of the target stocks from four choices: hill, valley, upward, and downward. Each IVIS drive included four interactions with the stock ticker, whose lengths were 1, 2, 3, and 4 min. The order of interactions was counterbalanced for the four IVIS drives using a Latin square, and a 1-min interval separated consecutive interactions.

TABLE I
CHARACTERISTICS OF FIXATIONS, SACCADES, AND SMOOTH PURSUITS

Types	Dispersion	Velocity
Fixation	Small (≤ 1°)	Low, random direction
Saccade	Large (> 1°)	400-600 °/sec, straight
Smooth pursuit	Target decided (> 1°)	1-30 °/sec, target trajectory

4) Eye Movement Measures: Eye movement data were collected at 60 Hz using a Seeing Machines' faceLab eye tracking system (version 4.1). The eye tracking system uses two small video cameras, located on the dashboard on either side of the steering wheel, to track head and eye movements. The images captured by the cameras are analyzed by the faceLab software to calculate, among other measures, horizontal and vertical coordinates for a gaze vector that intersects the simulator screen. Initial calibration of the eye tracker for each participant took 5 to 15 min, and the calibration of the gaze vector with the simulator screen was verified immediately before each experimental drive. After calibration, the tracking error was within approximately 5° of visual angle for most participants. The participants did not wear glasses or eye make-up because these conditions can negatively affect tracking accuracy; some participants wore contact lenses, which caused no adverse effects. This eye tracking system does not require any head-mounted or chin-rested hardware. Although the system may not track eyes as accurately as those using fixing hardware, it can do so without disturbing the driver and, thus, is more suitable for detecting drivers' cognitive distraction.

The gaze vector-screen intersection coordinates were transformed into a sequence of fixations, saccades, and smooth pursuits. To identify these three eye movements, segments of raw eye data were categorized based on two characteristics: dispersion and velocity (see Table I). Dispersion describes the span (in radians of visual angle) that the gaze vector covered during a movement, and velocity describes the speed (in radians of visual angle per second) and direction of the gaze vector (in radians) during a movement. The identification process began with a segment of six frames, and based on these characteristics, the posterior probabilities of the eye movements were calculated for the segment (see Fig. 2). If the highest probability was greater than a threshold, the segment was identified as an eye movement. Then, the segment was increased by one frame, and the process was repeated to check if the

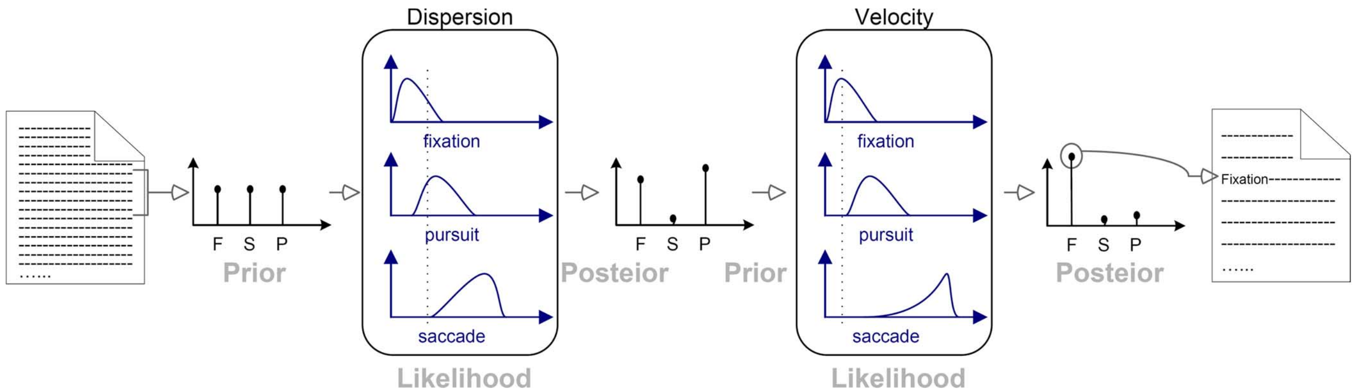


Fig. 2. Illustration of the algorithm used to identify eye movements.

TABLE II
FEATURE COMBINATIONS USED AS MODEL INPUT

Eye-movement and driving measures	Feature combinations: eye minus spatial data	eye data	eye plus driving
fixation duration	√	√	√
mean of horizontal position of fixation		√	√
mean of vertical position of fixation		√	√
s.d. of fixation position	√	√	√
pursuit duration	√	√	√
pursuit distance	√	√	√
pursuit direction	√	√	√
pursuit speed	√	√	√
percentage of pursuit in time	√	√	√
mean of blinking frequency	√	√	√
s.d. of steering wheel position			√
mean of steering error			√
s.d. of lane position			√

eye movement continued in the new frame. If no movement could be identified, the segment was decreased by one frame, and the posterior probabilities were calculated again. If only three frames remained in the segment, the eye movement was identified using only the speed characteristic. When speed was high, the movement was labeled as a saccade; when the speed was low, it was labeled as a smooth pursuit. After each eye movement was completely identified, the identification process began again with a new six-frame segment. The likelihood of eye movements and the threshold of posterior probability were chosen according to the literature [29] and adjusted according to the particular characteristics of our data.

After identification, measures of eye movements were summarized over various windows to create instances that became the SVM model inputs. As shown in Table II, the ten categories of eye movement measures were fixation duration, horizontal and vertical fixation location coordinates, sd of fixation location, pursuit duration, pursuit distance, pursuit direction, pursuit speed, and percentage of time spent performing pursuit movements. Fixation and pursuit duration represent the temporal characteristics of eye movements; horizontal and vertical position of fixation relates to spatial distribution of gaze, and sd more explicitly represents the variability of gaze. The pursuit distance, direction and speed, and the percentage of time spent in pursuit movements capture the characteristics of smooth pursuits. Saccade measures were not included because the study was interested in how cognitive distraction may interfere with the driver's acquisition of visual information, and this only occurs during fixation and pursuit eye movements.

5) Driving Performance Measures: The driving measures consist of sd of steering wheel position, mean of steering error, and sd of lane position. The driving simulation directly outputted steering wheel position and lane position at 60 Hz. Steering error was calculated at 5 Hz, which describes the difference between the actual steering wheel position and the steering wheel position predicted by a second-order Taylor expansion [28]. The driving performance measures were summarized across the same windows as eye movement measures.

B. Model Characteristics and Training

1) Distraction Definitions: Four different distraction definitions classified the binary states of distraction (i.e., "distracted" and "not distracted") for the SVMs (see Table III). The first two defined the assumed driver cognitive state based on experimental conditions. In the experiment, participants drove four drives with the IVIS task that included both IVIS and non-IVIS stages and two baseline drives without the IVIS task. The DRIVE definition classified the IVIS drives as "distracted" and the baseline drives as "not distracted." The STAGE definition classified the stages with IVIS interaction as "distracted" and the non-IVIS stages and baseline drives as "not distracted." Thus, the difference between DRIVE and STAGE lies in the 1-min non-IVIS intervals between IVIS interactions: DRIVE

TABLE III
MODEL CHARACTERISTICS AND THEIR VALUES

Model characteristics	Values				
Distraction definition	DRIVE	STAGE	STEER	RT	
Feature combination	Eye - spatial info	Eye	Eye + driving		
Window size	5, 10, 20, 40 seconds				
Overlap	1%	25%	50%	75%	95%

defined these as "distracted," while STAGE defined these as "not distracted." Since cognitive states can change gradually and continuously, the participants may remain distracted during the non-IVIS intervals, even though they were not performing the secondary task. Thus, we speculated that DRIVE and STAGE would capture distraction differently.

The other two definitions (STEER and RT) were based on driving performance. We categorized the upper quartile of all steering error values and accelerator release reaction times for each participant as "distracted" and the remainder as "not distracted." Accelerator release reaction time was defined by the interval from when the LV began to brake until the participant released the gas pedal. STEER explores the possible influence of eye movements on driving control, while RT predicts driver response to the braking events using the antecedent eye movement and driving measures.

2) Feature Combinations: Three different combinations of input variables consisting of eye movement and driving measures were investigated. As may be seen in Table II, the inputs are the ten categories of eye movement measures and the three driving measures. Three feature combinations were designed to compare the importance of some specific variables for distraction detection. First, we compared "eye minus spatial data" and "eye data." "Eye minus spatial data" excluded horizontal and vertical fixation position from "eye data" because the spatial distribution of fixations may be an important indicator of the eye-scanning pattern and may be a particularly helpful means of detecting driver distraction. Second, the comparison of "eye data" with "eye plus driving" evaluated how valuable driving measures might be for detection. Since the driving measures included mean steering error, the "eye plus driving" combination was not used to identify the STEER-based definition of distraction.

3) Summarizing Parameters of Inputs: The summarizing parameters for the inputs were window size and overlap between windows. Window size denotes the period over which eye movement and driving data were averaged (see Table III). The comparisons of window size serve to identify the appropriate length of data that can be summarized to reduce the noise of the input data without losing useful information. We chose four window sizes: 5, 10, 20, and 40 s.

Overlap is the percentage of data that were shared between windows. The summaries of eye movement and driving measures across the specific combinations of window size and overlap formed the "instances." Overlap reflects the redundancy between instances and influences the detection frequency of models. In a real-time application, window size and overlap interact to affect the computational load on the detection system. For the DRIVE, STAGE, and STEER definitions, five overlaps (1%, 25%, 50%, 75%, and 95%) were used for window sizes 5–20 s. Only three overlaps (50%, 75%, and 95%) were used for the 40-s window because there were not enough instances to train and test the models if 1% and 5% overlaps were used. For the RT definition, the overlaps were not applied because the braking events were discrete. In this case, the models used only the instance occurring immediately before the braking events to predict the RT performance.

4) Model Training: First, the data were summarized across the windows, which partially overlapped, to form instances. Then, the instances using the same window size and overlap were normalized using z-score and labeled as either "distracted" or "not distracted" according to the distraction definitions. For the DRIVE, STAGE, and STEER definitions, we randomly selected 200 training instances (100 for each class) and used the remaining instances, which accounted for at least two thirds of total instances, for testing. That is, all models were trained with 200 instances and tested with various numbers of instances. For the RT definition that considered response to discrete braking events, about 25% of the total instances (about 8–10, evenly divided between the two classes) were used for training. We also used the same training and testing datasets to build logistic regression models.

The radial basis function (RBF) was chosen as the kernel function for the SVM models

$$K(x_i, x_j) = e^{-\gamma|x_i - x_j|^2} \tag{1}$$

where x_i and x_j represent two data points, and γ is a predefined positive parameter. The RBF is a very robust kernel function. Using the RBF, it is possible to implement both nonlinear and linear mapping by manipulating the values of γ and the penalty parameter C. The RBF can reduce numerical difficulties and tends to obtain more robust results than other kernels, such as polynomial [26]. Moreover, using the RBF to train SVM models requires only two parameters (C and γ). In training, we searched for C and γ in the exponentially growing sequences ranging of 2^{-5} to 2^5 to obtain good parameter settings for the models [30]. We used the "LIBSVM" Matlab toolbox [30] to train and test the SVM models. For the logistic regression models, we used the Matlab Statistics Toolbox (i.e., the glmfit and glmval m-functions).

Data from ten participants were used for the DRIVE, STAGE, and STEER definitions, and data from nine participants were used for the RT definition. Some data were excluded because the participants failed to follow instructions, or the methods of data collection malfunctioned. For example, one participant drove too slowly to follow the LV, which produced too few braking events to train models for RT. For most participants, we built 54 models (three feature combinations $\times$ 18 window size-overlap combinations) for each of the DRIVE and STAGE definitions, 36 models (two feature combinations $\times$ 18 window size-overlap combinations) for the STEER definition, and 12 models (three feature combinations $\times$ four window sizes) for the RT definition. In all, 1548 SVM models and 1548 logistic models were constructed from the training data. These models were tested with the remaining data from the corresponding participants and with the same combination of model characteristics; preliminary analyses had indicated

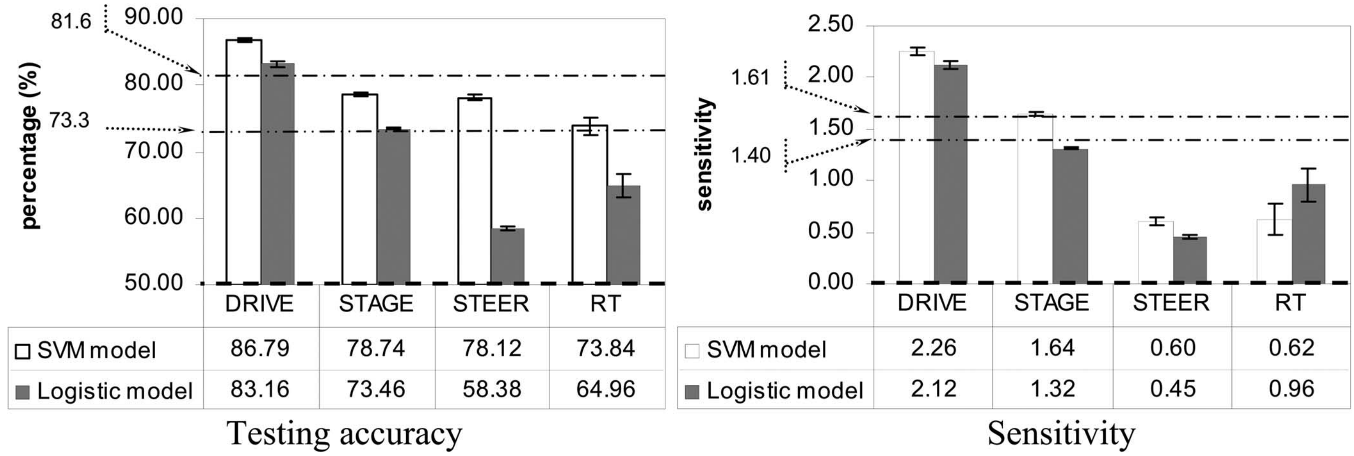


Fig. 3. Results of the SVM and logistic models. The dashed, dash-dot, and dash-double dot lines indicate chance selection, SVM average, and logistic average, respectively. The error bars indicate standard error from comparison tests.

that the prediction across participants varies greatly and, in general, is worse than that within participants.

5) Model Performance Measures: Model performance was evaluated with three different measures. The first was testing accuracy, which is the ratio of the number of instances correctly identified by the model to the total number of instances in the testing set. The other two measures were associated with signal detection theory: sensitivity (d') and response bias (β), which were calculated according to

$$d' = \Phi^{-1}(\mathrm{HIT}) - \Phi^{-1}(\mathrm{FA})$$
$$\beta = e^{\left\{\frac{\left|\Phi^{-1}(\mathrm{FA})\right|^2 - \left|\Phi^{-1}(\mathrm{HIT})\right|^2}{2}\right\}} \tag{2}$$

where HIT is hit rate equal to [true positive/(true positive + false negative)] and also called sensitivity (different than d') in some literature; FA is false alarm rate defined as [false positive/(false positive + true negative)] and equal to (1-specification) used in some literature; and $\Phi - 1$ presents the function of calculating z-score. d' represents the ability of the model to detect driver distraction. The larger the value of d', the more sensitive the model. β signifies the strategy used by the model. When β equals 1, models favor neither "distracted" nor "not distracted," and false alarms and misses (not detecting distraction when it is present) tend to occur at similar rates. When $\beta < 1$, the models are classified as liberal and are more likely to overestimate driver distraction with higher false alarm rates than miss rates. When $\beta > 1$, the models are classified as conservative and are more likely to underestimate driver distraction with more misses than false alarms. Both d' and β can affect detection accuracy. By separating the sensitivity to distraction from the bias of the model in identifying distraction [31], these signal detection theory measures allow a more refined evaluation of the detection system.

III. Results

A. Performance of the SVM Models

The performance of the SVM models was better than chance, which is defined as 50% accuracy and zero sensitivity. Averaging across all four distraction definitions, the SVM models detected driver distraction with 81.1% accuracy (sd $= 9.05\%$). The mean sensitivity for all SVM models was 1.54 (sd $= 1.04$), and the mean response bias was 3.91 (sd $= 35.22$). Both accuracy and sensitivity significantly exceeded chance performance (accuracy: $t_9 = 21.66$, $p < 0.0001$; sensitivity: $t_9 = 11.71$, $p < 0.0001$). Accuracy and sensitivity had a moderately positive relationship with a correlation coefficient of 0.68 ($p < 0.0001$). The dotted lines for the SVM models in Fig. 3 illustrate these results.

B. Comparison With the Logistic Regression

The SVM method outperformed logistic regression. The paired t-tests show that the SVM models were more accurate (SVM: 81.1%, logistic: 72.7%, $t_9 = 11.74$, $p < 0.0001$) and more sensitive (SVM: 1.54, logistic: 1.37, $t_9 = 5.59$, $p = 0.0003$) than the logistic models (see Fig. 3). In the refined comparisons shown in Fig. 3, the SVM models were more accurate than the logistic models for all definitions (DRIVE: $t_9 = 4.27$, $p = 0.0021$; STAGE: $t_9 = 5.71$, $p = 0.0003$; STEER: $t_9 = 11.47$, $p < 0.0001$; RT: $t_8 = 2.43$, $p = 0.041$). The SVM models also had greater sensitivity than the logistic models for DRIVE and STAGE (DRIVE: $t_9 = 3.02$, $p = 0.0144$; STAGE: $t_9 = 8.84$, $p < 0.0001$) and marginally greater sensitivity for STEER ($t_9 = 2.22, p = 0.054$). No difference in sensitivity was seen, however, for RT ($t_8 = -1.30, p = 0.23$).

When comparing response bias, we found that the SVM models took an overall conservative strategy ($\beta = 3.91$), while the logistic models were neutral ($\beta = 1.05$). However, no significant difference in response bias was found either for average value or for DRIVE, STAGE, and RT (Average: $t_9 = 1.63$, $p = 0.137$; DRIVE: $t_9 = 0.81$, $p = 0.440$; STAGE: $t_9 = 1.00$, $p = 0.345$; RT: $t_8 = 1.71$, $p = 0.12$). The SVM models were marginally significantly more conservative than the logistic models ($t_9 = 2.24, p = 0.052$) for STEER.

Because the difference between the mean response bias for the two types of models was relatively large but not statistically significant, the variance of response bias was examined. The variance of the SVM models' response bias was significantly greater than that of the logistic models for the first

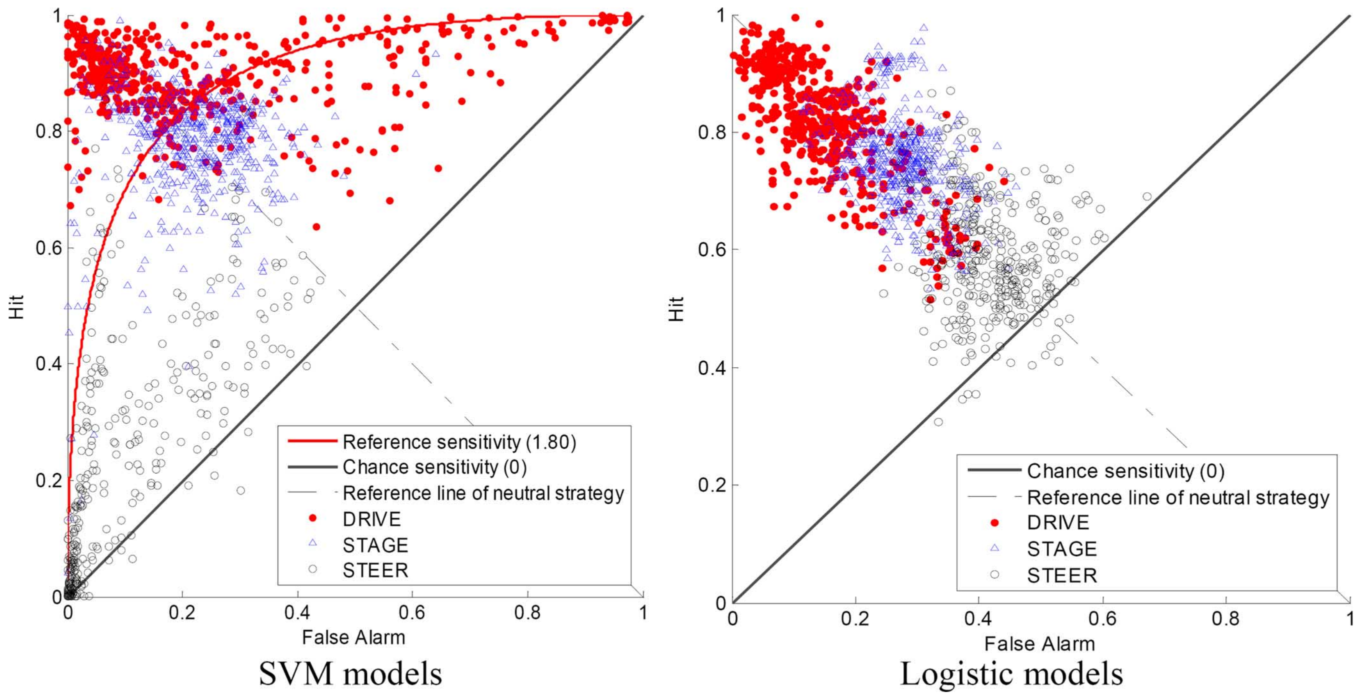


Fig. 4. ROC curves for the SVM and logistic models for DRIVE, STAGE, and STEER.

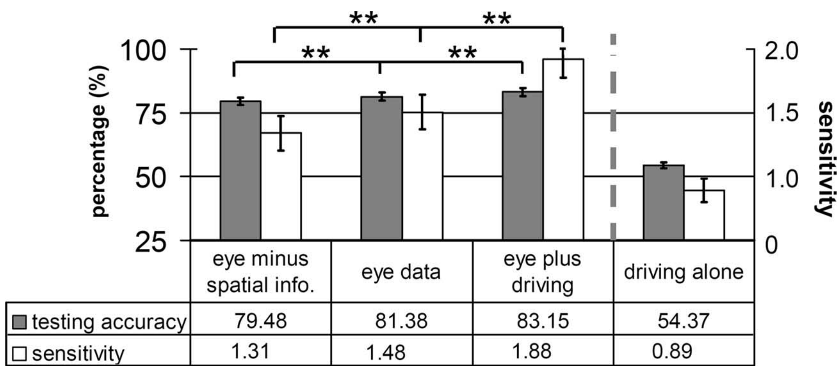
three definitions (DRIVE: $F_{9,9} = 102.20, p < 0.0001$; STAGE: $F_{9,9} = 2862.28,\ p < 0.0001$; STEER: $F_{9,9} = 93201.8,\ p < 0.0001$) but not for RT ($F_{8,8} = 0.34, p = 0.92$). The receiver operating characteristic (ROC) plots in Fig. 4 clearly show these differences between model types for each definition. The dashed diagonal represents a neutral strategy ($\beta = 1$). Models using liberal strategies ($\beta < 1$) are located on the right side of the neutral line, while models using conservative strategies ($\beta > 1$) are located on the left side. The logistic models (the right graph) are located along the diagonal, while the SVM models (the left graph) are span a much wider area and show different characteristics for different definitions. That is, misses and false alarms equally contributed to detection errors for the logistic models. This happens because during training, the logistic method minimizes training error. When the number of training instances from the two classes ("distracted" and "not distracted") is the same, the trained logistic models tend to make both types of errors (false alarms and misses) at the same rate. Meanwhile, the SVM models used different values of C and γ in training to vary miss and false alarm rates. The variability of C and γ could be one of the reasons why the SVM models achieved higher testing accuracy and sensitivity than the logistic models. The SVM models using DRIVE and STAGE tended to use neutral or slightly liberal strategies, and STEER used substantially more conservative strategies.

C. *Effect of Model Characteristics*

The results thus far have demonstrated that the SVM method is a more powerful technique for identifying driver distraction compared to logistic regression. To guide the implementation of real-time distraction detection systems using SVMs, the effects of the distraction definitions, feature combinations, window sizes, and overlaps on model performance were studied. The analyses used the mixed linear model with subject as a repeated measure and performed *posthoc* comparisons using the Tukey–Kramer method with SAS 9.0.

1) Distraction Definition: Accuracy and sensitivity (d') were significantly affected by the different distraction definitions (accuracy: $F_{3,26} = 254.32,\ p < 0.0001$; sensitivity: $F_{3,26} = 570.17, p < 0.0001$). The models for DRIVE had the highest accuracy and sensitivity, and the models for STEER and RT had the lowest accuracy and sensitivity (see the left graphs in Fig. 3). The differences in model sensitivity (d') are depicted in Fig. 4. Most points for STAGE and STEER are below the reference curve for sensitivity equal to 1.80, whereas most points for DRIVE are above the reference curve. Despite rather large differences in mean bias (DRIVE: 1.9; STAGE: 3.6; STEER: 8.33), the definitions did not differ significantly due to the large variance in bias, as discussed earlier.

Comparing the two definitions based on the experimental conditions, the models for DRIVE were more accurate and more sensitive than the models for STAGE (accuracy: $t_{26} = 21.84, p < 0.0001$; sensitivity: $t_{26} = 17.05, p < 0.0001$). This suggests that the differences in eye movement patterns and driving performance are more discernible when comparing IVIS and baseline drives than when comparing the conditions distinguished by whether the driver was interacting with the IVIS system or not. One possible interpretation of this result is that drivers' eye movements continued to be affected after the IVIS task had ended, and so, the eye movement patterns during the one-minute non-IVIS intervals between IVIS interactions maintained the same or similar patterns as during the interactions.



	eye minus spatial info.	eye data	eye plus driving	driving alone
testing accuracy	79.48	81.38	83.15	54.37
sensitivity	1.31	1.48	1.88	0.89

Fig. 5. SVM testing accuracy and sensitivity for the feature combinations. The braces represent the *posthoc* comparisons between the successive combinations using the Tukey–Kramer method. ** indicates $p < 0.05$. The error bars indicate standard error from comparison tests.

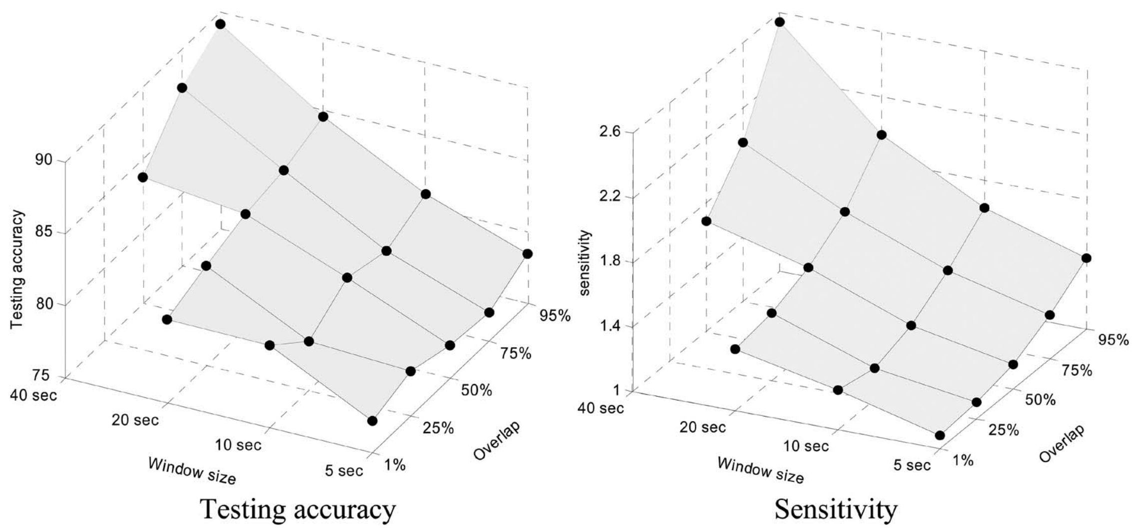


Fig. 6. Testing accuracy and sensitivity for different summarizing parameters of input data.

Comparing the distraction definitions based on driving performance, STEER was significantly more accurate than RT ($t_{26} = 5.99, p = 0.0002$) but was not different in sensitivity. The results show eye movements could reflect drivers' steering performance and predict response time to the lead-vehicle brake but not with a high degree of sensitivity.

2) Feature Combinations: Feature combinations had a significant effect on testing accuracy ($F_{2,18} = 25.84, p < 0.0001$) and sensitivity ($F_{2,18} = 44.68, p < 0.0001$) but not on response bias ($F_{2,18} = 2.48, p = 0.1117$). Testing accuracy and sensitivity increased with the number of input variables (see Fig. 5). Specifically, the multiple comparisons show that the spatial information of eye movements and driving measures both contributed to the detection of driver distraction ($t_{18} = 2.97$, $p = 0.0212$, and $t_{18} = 6.67$, $p < 0.0001$, respectively) by increasing model sensitivity. Adding the driving measures to the eye data increased sensitivity more (0.4102, $t_{18} = 6.67$, $p < 0.0001$) than adding the spatial information to the other eye movements (0.1650, $t_{18} = 2.97$, $p = 0.0212$).

Since the driving measures improved model performance so dramatically, the importance of driving measures compared to eye movement features for the detection of distraction was of interest. We built additional models using only driving measures as inputs. "Driving alone" achieved accuracy of only 54.4%, sensitivity of 0.89, and bias of 1.87 (see Fig. 5). This finding that gaze-related features led to much better prediction accuracy, compared to driving performance measures alone, agrees with previous research [18].

3) Summarizing Parameters of the Input Data: Window size affected testing accuracy ($F_{3,27} = 33.35, p < 0.0001$) and sensitivity ($F_{3,27} = 44.68, p < 0.0001$) but not response bias ($F_{3,27} = 2.48, p = 0.1117$). The models' accuracy and sensitivity increased with window size, suggesting that using longer periods to summarize the data made the distraction signal easier for the models to detect. Similarly, overlap had significant effects on testing accuracy ($F_{4,36} = 19.01, p < 0.0001$) and sensitivity ($F_{4,36} = 72.47, p < 0.0001$) but not on response bias ($F_{4,36} = 0.79, p = 0.5421$). The results suggest that increasing the redundancy of input data between adjacent windows improves the model performance. More importantly, window size and overlap significantly interacted to affect testing accuracy ($F_{17,153} = 35.36, p < 0.0002$) and sensitivity ($F_{17,153} = 51.01, p < 0.0001$). This is evident in Fig. 6. This interaction was not significant for response bias ($F_{17.153} = 1.30$, $p = 0.1966$).

Finally, we studied the effects of summarizing parameters for the distraction definition with the best performance: DRIVE. Window size and overlap showed the same trend as seen in

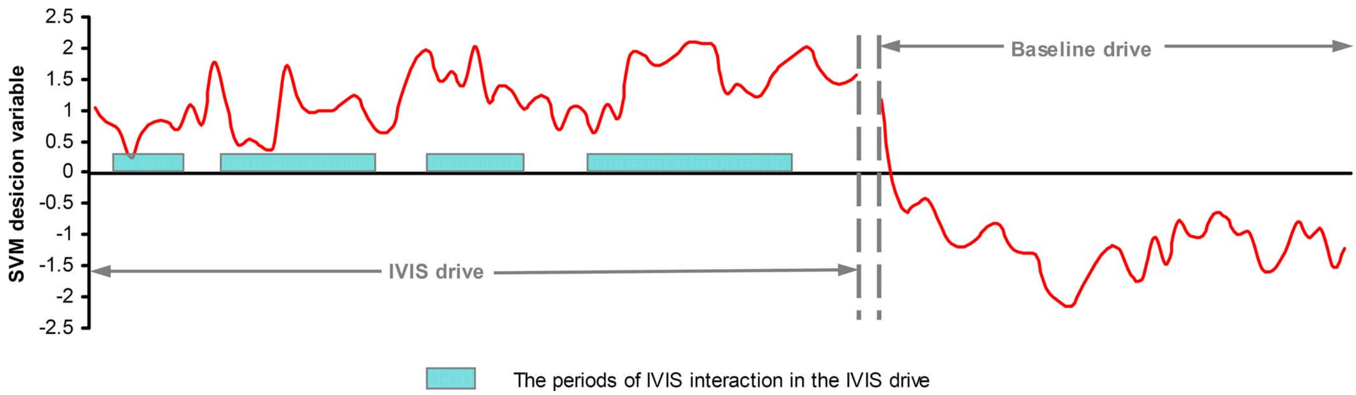


Fig. 7. SVM decision variable along the timeline of an IVIS drive and a baseline drive for participant SF7.

Fig. 6. The best model, using 40-s window size with 95% overlap, obtained 96.08% accuracy, sensitivity of 3.84, and response bias of 4.25.

IV. Discussion

This paper develops a methodology using SVMs, which is a data mining method, to detect driver distraction from eye movements and driving performance in real time. The connections between eye movements, driving performance, and driver distraction have already been demonstrated with the measures aggregated over experimental conditions [17], [20]–[22]. This paper explored the change of eye movements and driving performance with driver distraction in a smaller window, the results of which can be used in the real-time to detect driver's cognitive distraction. The results show that the SVM models were not only able to detect driver distraction substantially better than chance performance, but they also outperformed more traditional logistic regression models. The comparisons for the model characteristics show that, on average, the DRIVE definition, "eye plus driving" feature combination, and 40-s window size with 95% overlap led to the best model performance.

To make a prediction, the SVM models generated a decision variable for each testing instance. When the decision variable was positive, SVM models outputted a binary state 1; when negative, the models outputted 0. The value of the decision variable can indicate the distance from the instance to the classification boundary and can be interpreted as the model's confidence in its prediction. Fig. 7 shows how the decision variable changed for one model with excellent performance using the DRIVE distraction definition, "eye plus driving" feature combination, and 40-s window with 75% overlap. The decision variable is positive in the IVIS drive and is below zero for most of the baseline drive. This figure shows that the binary states produced by the SVM models follow the expected trend corresponding to the IVIS task. Also, Fig. 7 suggests a delay in the decision variable with IVIS interactions in the IVIS drive. This delay may be attributed to either the inertia of driver attention or the aggregation of the data over a large window or a combination of both. Clearly, delay is an issue that must be addressed before SVMs are implemented in a real-time system and will be discussed later in this section.

The SVM methods not only provided reasonable results but also outperformed the logistic regression and have several important advantages. First, the SVM models used the RBF kernel function that can fit both linear and nonlinear relationships, whereas the logistic models can only fit linear models. Second, from a theoretical point of view, during training, the SVM method minimizes the upper bound of the generalization error [27], whereas the logistic method only minimizes training error. This makes the SVM more robust by making overfitting less likely than with logistic regression. The biggest benefit of these two advantages is increased sensitivity for DRIVE, STAGE, and STEER compared with the logistic method. Third, with the same training datasets, response bias of the SVM method can be adjusted to increase performance by choosing parameter values (i.e., C and γ) before training, regardless of the number of training instances in each class ("distracted" and "not distracted"). The response bias of logistic models largely depends on the ratio of training instances from the two classes in the training dataset because the logistic method minimizes training errors. Therefore, choosing appropriate parameter values is critical for an SVM model to perform well.

The feature selection shows that the more related features are added to the detection algorithm, the better the models perform. In addition, eye movement features play a more important role than driving performance features though adding driving performance features to the eye movement features boosted the accuracy and sensitivity of the models. Future studies will add more distraction-related measures, such as blink duration, in detection models to see if they can improve model performance.

Choosing summarizing parameters across time for input data is a very important issue for a real-time detection system. These parameters influence the delay from when driver distraction begins to when the distraction is detected, the salience of the change in driver's performance due to the distraction, and model computational load. Longer window sizes generated more accurate models, which is consistent with a previous study [18].

One limitation of this paper is that the bicyclist detection task likely caused drivers to scan the driving environment in a manner differently than they would in normal driving. The discrepancy between the experimental data and reality suggests a need for caution in generalizing the results. However, the bicyclist detection events simulated an important driving task—recognizing potential hazards. Moreover, in right-driving countries (such as the U.S.), drivers may more likely check the right-side of road compared with the left side because it takes less time for right-side hazards to enter the driving lane. Because the detection task reflected representative demands of driving, the scenarios used in the experiment reasonably simulated normal driving. For future studies, a more realistic scenario or on-road driving is needed to verify the SVM as a method for real-time detection cognitive distraction.

There are some additional, practical limitations for implementing detection systems. The first is how to obtain the consistent and reliable sensor data. Eye trackers may lose tracking accuracy when vehicles are traveling on rough roads or when the lighting conditions are variable. More robust eye tracking techniques are needed to make these detection systems a reality. Steering data can be obtained directly from the angle of the steering wheel and some have developed robust measures of lane position in real driving environments [14]. Second, delay of detection needs to be accurately measured to evaluate whether it is appropriate for the application. The delay in real-world systems can come from three sources. One source is sensor delay. For example, the eye tracker used in this paper took approximately 2.6 s to transfer camera image to numerical data. The second source is the data-reduction and computational time of SVM models. It took about one second to reduce data and compute 15-s-long data in this paper. These two kinds of delays can be reduced with the advance of sensor and computer technology and the improvement of the data reduction algorithms. Nonetheless, these delays are not reflected in Fig. 7 because model testing was conducted off-line. The third source is from summarizing data across windows. Larger windows may cause longer delay. Unfortunately, this paper was unable to quantify the lags precisely. The future study will create the procedure that allows us to calculate such lags. The consequence of these lags will depend on the particular distraction mitigation strategy they support. Developing a systematic approach to balance the cost of time lags with the precision of distraction estimates for particular mitigation strategies represents an important research issue.

V. Conclusion

SVMs provide a viable means of detecting cognitive distraction in real-time and outperformed the traditional approach of logistic regression. Comparing the performance of SVMs using eye movements to SVMs using only driving performance clearly demonstrates the importance of eye movements as an input to a distraction detection algorithm. Including both eye and driving measures as inputs to a distraction-detection algorithm is recommended. These data were collected in a driving simulator with relatively homogenous traffic and roadside. On-road data in a more diverse set of conditions is needed to assess the generality of the results.

Acknowledgment

The work presented herein is part of the SAfety VEhicle(s) using adaptive Interface Technology (SAVE-IT) program. The authors would like to thank the graduate students in the Cognitive Systems Laboratory for offering valuable comments, as well as the assistance of T. Lopes at The University of Iowa Public Policy Center in preparing this manuscript.

References

[1] E. D. Sussman, H. Bishop, B. Madnick, and R. Walter, "Driver inattention and highway safety," *Transp. Res. Rec.*, no. 1047, pp. 40–48, 1985.

[2] J. Wang, R. R. Knipling, and M. J. Goodman, "The role of driver inattention in crashes; New statistics from the 1995 crashworthiness data system (CDS)," in *Proc. 40th Annu.: Assoc. Advancement Automotive Med.*, 1996, pp. 377–392.

[3] J. C. Stutts, D. W. Reinfurt, L. Staplin, and E. A. Rodgman, "The role of driver distraction in traffic crashes," *AAA Foundation of Traffic Safety*, 2001.

[4] J. D. Lee, B. Caven, S. Haake, and T. L. Brown, "Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway," *Hum. Factors*, vol. 43, no. 4, pp. 631–640, 2001.

[5] A. J. McKnight and A. S. McKnight, "The effect of cellular phone use upon driver attention," *Accident Anal. Prev.*, vol. 25, no. 3, pp. 259–265, Jun. 1993.

[6] D. Haigney and S. J. Westerman, "Mobile (cellular) phone use and driving: A critical review of research methodology," *Ergonom.*, vol. 44, no. 2, pp. 132–143, Feb. 2001.

[7] D. L. Strayer, F. A. Drews, and W. A. Johnston, "Cell phone-induced failures of visual attention during simulated driving," *J. Exp. Psychol., Appl.*, vol. 9, no. 1, pp. 23–32, Mar. 2003.

[8] D. L. Strayer and W. A. Johnston, "Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone," *Psychol. Sci.*, vol. 12, no. 6, pp. 462–466, Nov. 2001.

[9] B. Donmez, L. Boyle, and J. D. Lee, "Taxonomy of mitigation strategies for driver distraction," in *Proc. Hum. Factors Ergonom. Soc. 47th Annu. Meeting*, Denver, CO, 2003, pp. 1865–1869.

[10] W. W. Gibbs, "Considerate computing," *Amer. Mag.*, vol. 292, no. 1, pp. 54–61, Jan. 2005.

[11] P. Smith, M. Shah, and N. V. Lobo, "Determining driver visual attention with one camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 4, pp. 205–218, Dec. 2003.

[12] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[13] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 63–77, Mar. 2006.

[14] J. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 20–37, Mar. 2006.

[15] J. McCall, D. Mipf, M. M. Trivedi, and B. Rao, "Lane change intent analysis using robust operators and sparse Bayesian learning," *IEEE Trans. Intell. Transp. Syst.,* to be published.

[16] H. M. Mandalia and D. D. Salvucci, "Using support vector machines for lane-change detection," in *Proc. Hum. Factors and Ergonomics Soc. 49th Annu. Meeting*, Orlando, FL, 2005, pp. 1965–1969.

[17] M. A. Recarte and L. M. Nunes, "Mental workload while driving: Effects on visual search, discrimination, and decision making," *J. Exp. Psychol., Appl.*, vol. 9, no. 2, pp. 119–137, Jun. 2003.

[18] Y. Zhang, Y. Owechko, and J. Zhang, "Driver cognitive workload estimation: A data-driven perspective," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Washington, DC, 2004, pp. 642–647.

[19] M. M. Hayhoe, "Advances in relating eye movements and cognition," *Infancy*, vol. 6, no. 2, pp. 267–274, 2004.

[20] J. G. May, R. S. Kennedy, M. C. Williams, W. P. Dunlap, and J. R. Brannan, "Eye movement indices of mental workload," *Acta Psychol.*, vol. 75, no. 1, pp. 75–89, Oct. 1990.

[21] E. M. Rantanen and J. H. Goldberg, "The effect of mental workload on the visual field size and shape," *Ergonom.*, vol. 42, no. 6, pp. 816–834, Jun. 1999.

[22] M. A. Recarte and L. M. Nunes, "Effects of verbal and spatial-imagery tasks on eye fixations while driving," *J. Exp. Psychol., Appl.*, vol. 6, no. 1, pp. 31–43, Mar. 2000.

[23] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[24] N. Cristianini and J. S. Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[25] H. Byun and S. W. Lee, "Applications of Support Vector Machines for pattern recognition: A survey," in *Proc. 1st Int. Workshop, SVM Pattern Recog. With Support Vector Mach.*, Niagara Falls, ON, Canada, 2002, pp. 213–236.

[26] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, *A Practical Guide to Support Vector Classification*. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

[27] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Netw.*, vol. 12, no. 6, pp. 783–789, Jul. 1999.

[28] O. Nakayama, T. Futami, T. Nakamura, and E. R. Boer, "Development of a steering entropy method for evaluating driver workload," presented at the Int. Congr. Expo., Detroit, MI, Mar. 1–4, 1999, Paper SAE Tech. Paper Series: 1999-01-0892.

[29] R. J. K. Jacob, "Eye tracking in advanced interface design," in *Advanced Interface Design and Virtual Environments*, W. Barfield and T. Furness, Eds. Oxford, U.K.: Oxford Univ. Press, 1995, pp. 258–288.

[30] C.-C. Chang and C.-J. Lin, in *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf

[31] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behav. Res. Meth., Instrum., Comput.*, vol. 31, no. 1, pp. 137–149, Feb. 1999.

Yulan Liang received the B.E. and M.S. degrees in precision instruments and mechanics from Tsinghua University, Beijing, China, in 2000 and 2002, respectively. She is currently working toward the Ph.D. degree in industrial engineering at The University of Iowa, Iowa City.

Her research interests include multitask performance, user modeling and performance prediction, driver distraction, and data mining in human behavior.

Michelle L. Reyes received the B.S. degree in industrial engineering from The University of Iowa, Iowa City, in 2002, where she is currently working toward the M.S. degree.

Since then, she has been a Research Assistant with the Cognitive Systems Laboratory, The University of Iowa, and is also a part-time student. Her research interests include driver distraction, workload, and performance, as well as in-vehicle information systems, eye movements, and physiological measures.

John D. Lee received the B.A. degree in psychology and the B.S. degree in mechanical engineering from Lehigh University, Bethlehem, PA, in 1987 and 1988, respectively, and the M.S. degree in industrial engineering and the Ph.D. degree in mechanical engineering from the University of Illinois, Urbana–Champaign, in 1989 and 1992, respectively.

He is currently a Professor with the Department of Mechanical and Industrial Engineering and the Director of the Cognitive Systems Laboratory, The University of Iowa, Iowa City. His research enhances the safety and acceptance of complex human-machine systems by considering how technology mediates attention. Specific research interests include technology-mediated attention, trust in technology, supervisory control, and collision warning systems and driver distraction.