Proceedings, Digital Signal Processing for in-Vehicle and mobile systems, Istanbul, Turkey, June 2007.

MACHINE LEARNING SYSTEMS FOR DETECTING DRIVER DROWSINESS

Esra Vural, Mujdat Cetin, Aytul Ercil □

Sabanci University
Faculty of
Engineering and Natural Sciences
Orhanli, Istanbul

Gwen Littlewort, Marian Bartlett, Javier Movellan

University Of California San Diego
Institute of
Neural Computation
La Jolla, San Diego

ABSTRACT

The advance of computing technology has provided the means for building intelligent vehicle systems. Drowsy driver detection system is one of the potential applications of intelligent vehicle systems. Previous approaches to drowsiness detection primarily make pre-assumptions about the relevant behavior, focusing on blink rate, eye closure, and yawning. Here we employ machine learning to datamine actual human behavior during drowsiness episodes. Automatic classifiers for 30 facial actions from the Facial Action Coding system were developed using machine learning on a separate database of spontaneous expressions. These facial actions include blinking and yawn motions, as well as a number of other facial movements. In addition, head motion was collected through automatic eye tracking and an accelerometer. These measures were passed to learning-based classifiers such as Adaboost and multinomial ridge regression. The system was able to predict sleep and crash episodes during a driving computer game with 96% accuracy within subjects and above 90% accuracy across subjects. This is the highest prediction rate reported to date for detecting real drowsiness. Moreover, the analysis revealed new information about human behavior during drowsy driving.

1. INTRODUCTION

In recent years, there has been growing interest in intelligent vehicles. A notable initiative on intelligent vehicles was created by the U.S. Department of Transportation with the mission of prevention of highway crashes [1]. The ongoing intelligent vehicle research will revolutionize the way vehicles and drivers interact in the future.

The US National Highway Traffic Safety Administration estimates that in the US alone approximately 100,000 crashes each year are caused primarily by driver drowsiness or fatigue [2]. Thus incorporating automatic driver fatigue detection mechanism into vehicles may help prevent many accidents.

One can use a number of different techniques for analyzing driver exhaustion. One set of techniques places sensors on standard vehicle components, e.g., steering wheel, gas pedal, and analyzes the signals sent by these sensors to detect drowsiness [3]. It is important for such techniques to be adapted to the driver, since Abut and his colleagues note that there are noticeable differences among drivers in the way they use the gas pedal [4].

A second set of techniques focuses on measurement of physiological signals such as heart rate, pulse rate, and Electroencephalography (EEG) [5]. It has been reported by researchers that as the alertness level decreases EEG power of the alpha and theta bands increases [6]. Hence providing indicators of drowsiness. However this method has drawbacks in terms of practicality since it requires a person to wear an EEG cap while driving.

A third set of solutions focuses on computer vision systems that can detect and recognize the facial motion and appearance changes occurring during drowsiness [7] [8]. The advantage of computer vision techniques is that they are non-invasive, and thus are more amenable to use by the general public. There are some significant previous studies about drowsiness detection using computer vision techniques . Most of the published research on computer vision approaches to detection of fatigue has focused on the analysis of blinks and head movements. However the effect of drowsiness on other facial expressions have not been studied thoroughly. Recently Gu & Ji presented one of the first fatigue studies that incorporates certain facial expressions other than blinks. Their study feeds action unit information as an input to a dynamic bayesian network. The network was trained on subjects posing a state of fatigue [9]. The video segments were classified into three stages: inattention, yawn, or falling asleep. For predicting falling-asleep, head nods, blinks, nose wrinkles and eyelid tighteners were used.

Previous approaches to drowsiness detection primarily make pre-assumptions about the relevant behavior, focus-

□The first author performed the work while at University Of California San Diego

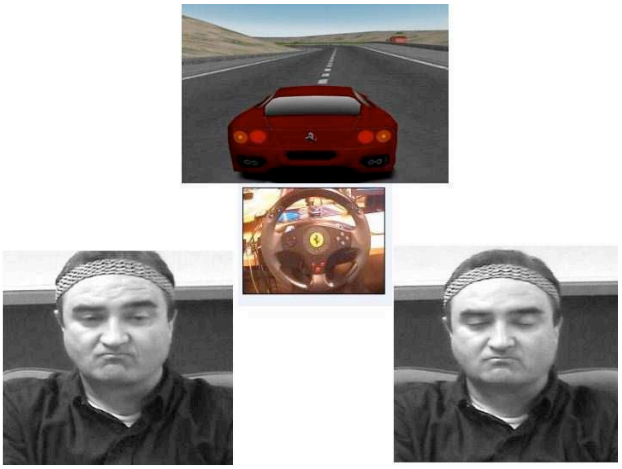

Fig. 1. Driving simulation task.

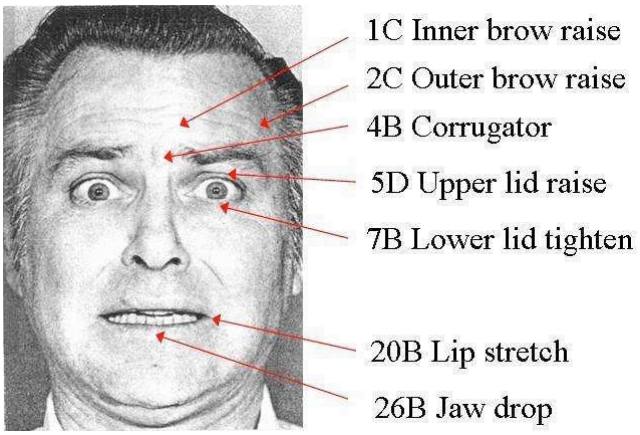

Fig. 2. Example facial action decomposition from the Facial Action Coding System.

ing on blink rate, eye closure, and yawning. Here we employ machine learning methods to datamine actual human behavior during drowsiness episodes. The objective of this study is to discover what facial configurations are predictors of fatigue. In this study, facial motion was analyzed automatically from video using a fully automated facial expression analysis system based on the Facial Action Coding System (FACS) [10]. In addition to the output of the automatic FACS recognition system we also collected head motion data using an accelerometer placed on the subject's head, as well as steering wheel data.

2. METHODS

2.1. Driving task

Subjects played a driving video game on a windows machine using a steering wheel[1] and an open source multi-platform video game[2] (See Figure 1). The windows version of the video game was maintained such that at random times, a wind effect was applied that dragged the car to the right or left, forcing the subject to correct the position of the car. This type of manipulation had been found in the past to increase fatigue [11]. Driving speed was held constant. Four subjects performed the driving task over a three hour period beginning at midnight. During this time subjects fell asleep multiple times thus crashing their vehicles. Episodes in which the car left the road (crash) were recorded. Video of the subjects face was recorded using a DV camera for the entire 3 hour session.

1 Thrustmaster® Ferrari Racing Wheel

2 The Open Racing Car Simulator(TORCS)

2.2. Head movement measures

Head movement was measured using an accelerometer that has 3 degrees of freedom. This three dimensional accelerometer[3] has three one dimensional accelerometers mounted at right angles measuring accelerations in the range of 5g to +5g where g represents earth gravitational force.

2.3. Facial Action Classifiers

The facial action coding system (FACS) [12] is arguably the most widely used method for coding facial expressions in the behavioral sciences. The system describes facial expressions in terms of 46 component movements, which roughly correspond to the individual facial muscle movements. An example is shown in Figure 2. FACS provides an objective and comprehensive way to analyze expressions into elementary components, analagous to decomposition of speech into phonemes. Because it is comprehensive, FACS has proven useful for discovering facial movements that are indicative of cognitive and affective states. In this paper we investigate whether there are Action units (AUs) such as chin raises (AU17), nasolabial furrow deepeners(AU11), outer(AU2) and inner brow raises (AU1) that are predictive of the levels of drowsiness observed prior to the subjects falling sleep

In previous work we presented a system, named CERT, for fully automated detection of facial actions from the facial action coding system [10]. The workflow of the system is based is summarized in Figure 3. We previously reported detection of 20 facial action units, with a mean of 93% correct detection under controlled posed conditions, and 75% correct for less controlled spontaneous expressions ~~with head movements a~~nd speech.

3 Vernier®

For this project we used an improved version of CERT which was retrained on a larger dataset of spontaneous as well as posed examples. In addition, the system was trained to detect an additional 11 facial actions for a total of 31 (See Table 1). The facial action set includes blink (action unit 45), as well as facial actions involved in yawning (action units 26 and 27). The selection of this set of 31 out of 46 total facial actions was based on the availability of labeled training data.

AU	Name
1	Inner Brow Raise
2	Outer Brow Raise
4	Brow Lowerer
5	Upper Lid Raise
6	Cheek Raise
7	Lids Tight
8	Lip Toward
9	Nose Wrinkle
10	Upper Lip Raiser
11	Nasolabial Furrow Deepener
12	Lip Corner Puller
13	Sharp Lip Puller
14	Dimpler
15	Lip Corner Depressor
16	Lower Lip Depress
17	Chin Raise
18	Lip Pucker
19	Tongue show
20	Lip Stretch
22	Lip Funneller
23	Lip Tightener
24	Lip Presser
25	Lips Part
26	Jaw Drop
27	Mouth Stretch
28	Lips Suck
30	Jaw Sideways
32	Bite
38	Nostril Dilate
39	Nostril Compress
45	Blink

Table 1. Full set of action units used for predicting drowsiness

The facial action detection system was designed as follows: First faces and eyes are detected in real time using a system that employs boosting techniques in a generative framework [13]. The automatically detected faces are aligned based on the detected eye positions, cropped and scaled to a size of 96 × 96 pixels and then passed through a bank of Gabor filters. The system employs 72 Gabor spanning 9

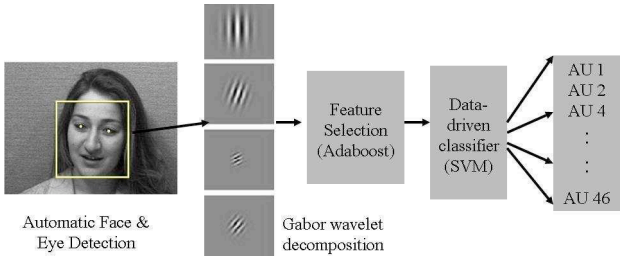

Fig. 3. Overview of fully automated facial action coding system.

spatial scales and 8 orientations. The outputs of these filters are normalized and then passed to a standard classifier. For this paper we employed support vector machines. One SVM was trained for each of the 31 facial actions, and it was trained to detect the facial action regardless of whether it occurred alone or in combination with other facial actions. The system output consists of a continuous value which is the distance to the separating hyperplane for each test frame of video. The system operates at about 6 frames per second on a Mac G5 dual processor with 2.5 ghz processing speed.

Facial expression training data The training data for the facial action classifiers came from two posed datasets and one dataset of spontaneous expressions. The facial expressions in each dataset were FACS coded by certified FACS coders. The first posed datasets was the Cohn-Kanade DFAT-504 dataset [14]. This dataset consists of 100 university students who were instructed by an experimenter to perform a series of 23 facial displays, including expressions of seven basic emotions. The second posed dataset consisted of directed facial actions from 24 subjects collected by Ekman and Hager. Subjects were instructed by a FACS expert on the display of individual facial actions and action combinations, and they practiced with a mirror. The resulting video was verified for AU content by two certified FACS coders. The spontaneous expression dataset consisted of a set of 33 subjects collected by Mark Frank at Rutgers University. These subjects underwent an interview about political opinions on which they felt strongly. Two minutes of each subject were FACS coded. The total training set consisted of 6000 examples, 2000 from posed databases and 4000 from the spontaneous set.

3. RESULTS

Subject data was partitioned into drowsy (non-alert) and alert states as follows. The one minute preceding a sleep episode or a crash was identified as a non-alert state. There was a mean of 24 non-alert episodes with a minimum of 9 and a maximum of 35. Fourteen alert segments for each subject were collected from the first 20 minutes of the driv-

ing task.[4] Our initial analysis focused on drowsiness prediction within-subjects.

3.1. Facial action signals

The output of the facial action detector consisted of a continuous value for each frame which was the distance to the separating hyperplane, i.e., the margin. Histograms for two of the action units in alert and non-alert states are shown in Figure 4. The area under the ROC (A') was computed for the outputs of each facial action detector to see to what degree the alert and non-alert output distributions were separated. The A' measure is derived from signal detection theory and characterizes the discriminative capacity of the signal, independent of decision threshold. A' can be interpreted as equivalent to the theoretical maximum percent correct achievable with the information provided by the system when using a 2-Alternative Forced Choice testing paradigm. Table 2 shows the actions with the highest A' for each subject. As expected, the blink/eye closure measure was overall the most discriminative for most subjects. However note that for Subject 2, the outer brow raise (Action Unit 2) was the most discriminative.

	AU	Name	A'
Subj1	45	Blink	.94
	17	Chin Raise	.85
	30	Jaw sideways	.84
	7	Lid tighten	.81
	39	Nostril compress	.79
Subj2	2	Outer brow raise	.91
	45	Blink	.80
	17	Chin Raise	.76
	15	Lip corner depress	.76
	11	Nasolabial furrow	.76
Subj3	45	Blink	.86
	9	Nose wrinkle	.78
	25	Lips part	.78
	1	Inner brow raise	.74
	20	Lip stretch	.73
Subj4	45	Blink	.90
	4	Brow lower	.81
	15	Lip corner depress	.81
	7	Lid tighten	.80
	39	Nostril Compress	.74

Table 2. The top 5 most discriminant action units for discriminating alert from non-alert states for each of the four subjects. A' is area under the ROC curve.

[4] Several of the drivers became drowsy very quickly which prevented extraction of more alert segments.

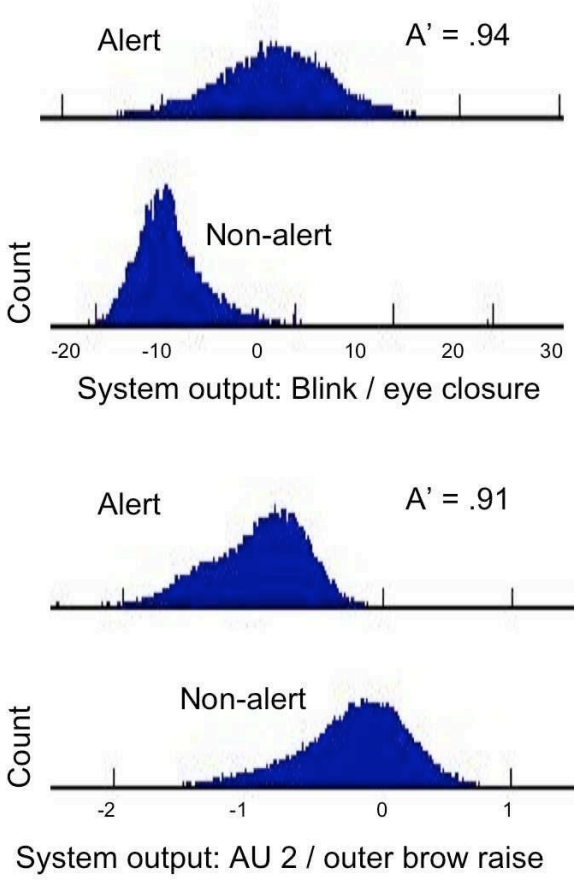

Fig. 4. Histograms for blink and Action Unit 2 in alert and non-alert states. A' is area under the ROC.

3.2. Drowsiness prediction

The facial action outputs were passed to a classifier for predicting drowsiness based on the automatically detected facial behavior. We compared two learning-based classifiers, Adaboost and multinomial ridge regression. Both within-subject prediction of drowsinss and across-subject (subject independent) prediction of drowsiness were tested.

For the within-subject prediction, 80% of the alert and non-alert episodes were used for training and the other 20% were reserved for testing. This resulted in a mean of 19 non-alert and 11 alert episodes for training, and 5 non-alert and 3 alert episodes for testing per subject.

The weak learners for the Adaboost classifier consisted of each of the 30 Facial Action detectors. The classifier was trained to predict alert or non-alert from each frame of video. There was a mean of 43,200 training samples, (24 + 11) ×60 × 30, and 1440 testing samples, (5 + 3) × 60 × 30, for each subject. On each training iteration, Adaboost se-

Classifier	Percent Correct	Hit Rate	False Alarm Rate
Adaboost	.92 ± .03	.92± .01	.06± .1
MLR	.94 ±.02	.98±.02	.13±.02

Table 3. Performance for drowsiness prediction, within subjects. Means and standard devisions are shown across subjects.

lected the facial action detector that minimized prediction error given the previously selected detectors. Adaboost obtained 92% correct accuracy for predicting driver drowsiness based on the facial behavior.

Classification with Adaboost was compared to that using multinomial ridge regression (MLR). Performance with MLR was similar, obtaining 94% correct prediction of drowsy states. The facial actions that were most highly weighted by MLR also tended to be the facial actions selected by Adaboost. 85% of the top ten facial actions as weighted by MLR were among the first 10 facial actions to be selected by Adaboost.

The ability to predict drowsiness in novel subjects was next tested by using a leave-one-out cross validation procedure. The data for each subject was first normalized to zero-mean and unit standard deviation before training the classifier. MLR was trained to predict drowsiness from the AU outputs several ways. Performance was evaluated in terms of area under the ROC. For all of the the novel subject analysis, the MLR output for each feature was summed over a temporal window of 12 seconds (360 frames) before computing A'. MLR trained on all features obtained an A' of .90 for predicting drowsiness in novel subjects.

First, MLR was trained on each facial action individually. Examination of the A' for each action unit reveals the degree to which each facial movement is associated with drowsiness in this study. The A's for the drowsy and alert states are shown in Table 4. The five facial actions that were the most predictive of drowsiness by *increasing* in drowsy states were 45, 2 (outer brow raise), 15 (frown), 17 (chin raise), and 9 (nose wrinkle). The five actions that were the most predictive of drowsiness by *decreasing* in drowsy states were 12 (smile), 7 (lid tighten), 39 (nostril compress), 4 (brow lower), and 26 (jaw drop). The high predictive ability of the blink/eye closure measure was expected. However the predictibility of the outer brow raise (AU 2) was previously unknown. We observed during this study that many subjects raised their eyebrows in an attempt to keep their eyes open, and the strong association of the AU 2 detector is consistent with that observation. Also of note is that action 26, jaw drop, which occurs during yawning, actually occurred *less* often in the critical 60 seconds prior to a crash. This is consistent with the prediction that yawning does not tend to occur in the final moments before falling asleep.

More when critically drowsy		
AU	Name	A'
45	Blink/eye closure	0.94
2	Outer Brow Raise	0.81
15	Lip Corner Depressor	0.80
17	Chin Raiser	0.79
9	Nose wrinkle	0.78
30	Jaw sideways	0.76
20	Lip stretch	0.74
11	Nasolabial furrow	0.71
14	Dimpler	0.71
1	Inner brow raise	0.68
10	Upper Lip Raise	0.67
27	Mouth Stretch	0.66
18	Lip Pucker	0.66
22	Lip funneler	0.64
24	Lip presser	0.64
19	Tongue show	0.61

Less when critically drowsy		
AU	Name	A'
12	Smile	0.87
7	Lid tighten	0.86
39	Nostril Compress	0.79
4	Brow lower	0.79
26	Jaw Drop	0.77
6	Cheek raise	0.73
38	Nostril Dilate	0.72
23	Lip tighten	0.67
8	Lips toward	0.67
5	Upper lid raise	0.65
16	Lower lip depress	0.64
32	Bite	0.63

Table 4. MLR model for predicting drowsiness across subjects. Predictive performance of each facial action individually is shown.

Second, MLR is computed using all features. MLR trained on all features obtained an A' of .90 for predicting drowsiness in novel subjects. Moreover, a new MLR classifier was trained by contingent feature selection, starting with the most discriminative feature (AU 45), and then iteratively adding the next most discriminative feature given the features already selected. These features are shown at the bottom of Table 5. Best performance of .98 was obtained with five features: 45, 2, 19 (tongue show), 26 (jaw drop), and 15. This five feature model outperformed the MLR trained on all features.

We next examined the effect of the size of the temporal window on performance. The five feature model was employed for this analysis. The performances shown to this

Feature	A'
AU45	.9468
AU45,AU2	.9614
AU45,AU2,AU19	.9693
AU45,AU2,AU19,AU26	.9776
AU45,AU2,AU19,AU26,AU15	.9792
all the features	.8954

Table 5. Drowsiness detection performance for novel subjects, using and MLR classifier with different feature combinations. The weighted features are summed over 12 seconds before computing A'.

point in the paper were for temporal windows of one frame, with the exception of the novel subject analysis (Tables **4** and **5**), which employed a temporal window of 12 seconds. The MLR output in the 5 feature model was summed over windows of N seconds, where N ranged from 0.5 to 60 seconds. Figure **5** shows the area under the ROC for drowsiness detection in novel subjects over time periods. Performance saturates at about 0.99 as the window size exceeds 30 seconds. In other words, given a 30 second video segment the system can discriminate sleepy versus non-sleepy segments with 0.99 accuracy across subjects.

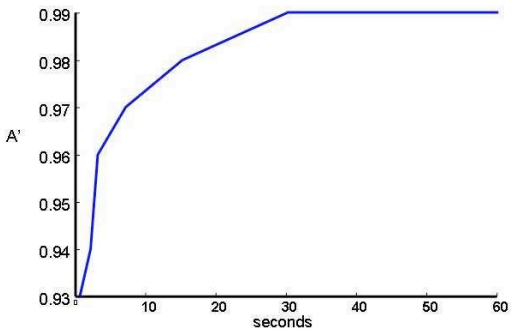

Fig. 5. Performance for drowsiness detection in novel subjects over temporal window sizes.

3.3. Coupling of Steering and Head Motion

Observation of the subjects during drowsy and nondrowsy states indicated that the subjects head motion differed substantially when alert versus when the driver was about to fall asleep. Surprisingly, head motion increased as the driver became drowsy, with large roll motion coupled with the steering motion as the driver became drowsy. Just before falling asleep, the head would become still.

We also investigated the coupling of the head and arm motions. Correlations between head motion as measured by the roll dimension of the accelerometer output and the steering wheel motion are shown in Figure **6**. For this subject (subject 2), the correlation between head motion and steering increased from 0.33 in the alert state to 0.71 in the non-alert state. For subject 1, the correlation between head motion and steering similarly increased from 0.24 in the alert state to 0.43 in the non-alert state. The other two subjects showed a smaller coupling effect. Future work includes combining the head motion measures and steering correlations with the facial movement measures in the predictive model.

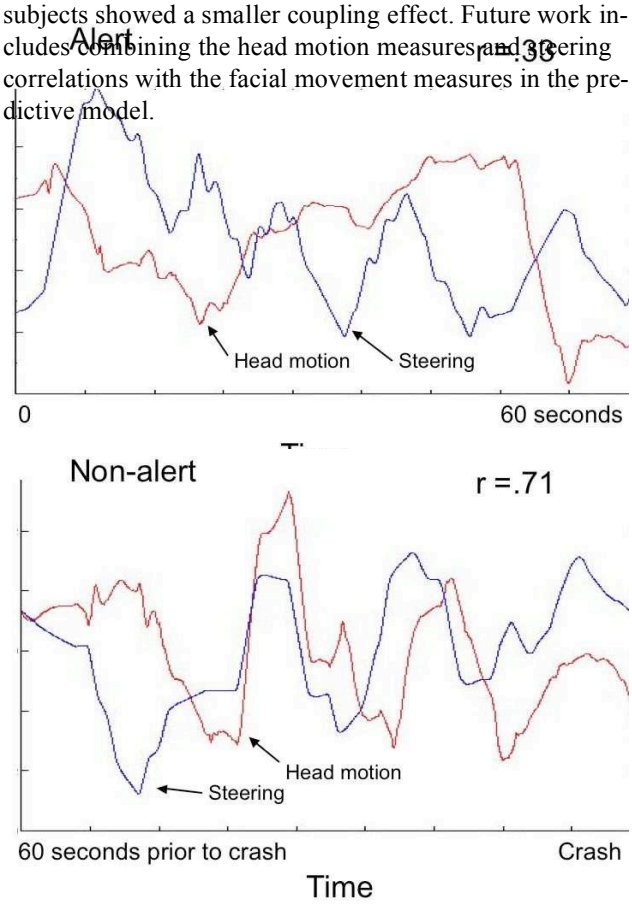

Fig. 6. Head motion and steering position for 60 seconds in an alert state (top) and 60 seconds prior to a crash. Head motion is the output of the roll dimension of the accelerometer.

4. CONCLUSION

This paper presented a system for automatic detection of driver drowsiness from video. Previous approaches focused on assumptions about behaviors that might be predictive of drowsiness. Here, a system for automatically measuring facial expressions was employed to datamine spontaneous behavior during real drowsiness episodes. This is the first work to our knowledge to reveal significant associations between facial expression and fatigue beyond eyeblinks. The project also revealed a potential association between head roll and driver drowsiness, and the coupling of head roll with steering motion during drowsiness. Of note is that a behavior that is often assumed to be predictive of drowsiness, yawn, was in fact a negative predictor of the 60-second window prior to a crash. It appears that in the moments before falling asleep, drivers yawn less, not more, often. This highlights the importance of using examples of fatigue and drowsiness conditions in which subjects actually fall sleep.

Acknowledgements This research was supported in part by NSF grants NSF-CNS 0454233,SBE-0542013 and by a grant from Turkish State Planning Organization. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

5. REFERENCES

[1] United States Department of Transportation., "Intelligent vehicle initiative.," http://www.its.dot.gov/ivi/ivi.htm./.

[2] United States Department of Transportation., "Saving lives through advanced vehicle safety technology.," http://www.its.dot.gov/ivi/docs/AR2001.pdf.

[3] Y. Takei, Y. Furukawa, "Estimate of driver's fatigue through steering motion," in *Man and Cybernetics, 2005 IEEE International Conference*, Volume: 2, On page(s): 1765- 1770 Vol. 2.

[4] Kei Igarashi, Kazuya Takeda, Fumitada Itakura, and Huseyin Abut, *DSP for In-Vehicle and Mobile Systems*, Springer US, 2005.

[5] W.A. Cobb., "Recommendations for the practice of clinical neurophysiology.," *Elsevier*, 1983.

[6] Kim Hong, Chung, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation.," *International Journal of Industrial Ergonomics.*, Volume 35, Issue 4, April 2005, Pages 307-320.

[7] Haisong Gu and Qiang Ji, "An automated face reader for fatigue detection," *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference*, vol. 00, pp. 111, 2004.

[8] Zutao Zhang and Jia shu Zhang, "Driver fatigue detection based intelligent vehicle control," in *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*,Washington, DC, USA, 2006, pp. 1262–1265, IEEE Computer Society.

[9] Haisong Gu, Yongmian Zhang, and Qiang Ji, "Task oriented facial behavior recognition with selective sensing," *Comput. Vis. Image Underst.*, vol. 100, no. 3, pp. 385–415, 2005.

[10] Bartlett M.S., Littlewort G.C., Frank M.G., Lainscsek C., Fasel I., and Movellan J.R., "Automatic recognition of facial actions in spontaneous expressions.," *Journal of Multimedia.*, 1(6) p. 22-35.

[11] Karl F. Van Orden, Tzyy-Ping Jung, and ScottMakeig, "Combined eye activity measures accurately estimate changes in sustained visual task performance," *Biological Psychology*, 2000 Apr;52(3):221-40.

[12] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto, CA, 1978.

[13] Movellan J.R. Fasel I., Fortenberry B., "A generative framework for real-time object detection and classification.," *Computer Vision and Image Understanding.*, 98, 2005.

[14] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the fourth IEEE International conference on automatic face and gesture recognition (FG'00)*, Grenoble, France, 2000, pp. 46–53.