Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring

Ralph Oyini Mbouna, Seong G. Kong, *Senior Member, IEEE*, and Myung-Geun Chun

***Abstract*—This paper presents visual analysis of eye state and head pose (HP) for continuous monitoring of alertness of a vehicle driver. Most existing approaches to visual detection of nonalert driving patterns rely either on eye closure or head nodding angles to determine the driver drowsiness or distraction level. The proposed scheme uses visual features such as eye index (EI), pupil activity (PA), and HP to extract critical information on nonalertness of a vehicle driver. EI determines if the eye is open, half closed, or closed from the ratio of pupil height and eye height. PA measures the rate of deviation of the pupil center from the eye center over a time period. HP finds the amount of the driver's head movements by counting the number of video segments that involve a large deviation of three Euler angles of HP, i.e., nodding, shaking, and tilting, from its normal driving position. HP provides useful information on the lack of attention, particularly when the driver's eyes are not visible due to occlusion caused by large head movements. A support vector machine (SVM) classifies a sequence of video segments into alert or nonalert driving events. Experimental results show that the proposed scheme offers high classification accuracy with acceptably low errors and false alarms for people of various ethnicity and gender in real road driving conditions.**

***Index Terms*—Driver alertness monitoring, driver drowsiness detection, eye state, head pose (HP), support vector machines (SVMs).**

I. Introduction

DRIVER drowsiness has been one of the major causes of fatal car accidents. According to a 2012 poll conducted by the National Sleep Foundation, one in five pilots admit that they have made a serious error, and one in six train operators and truck drivers say that they have had a "near miss" due to sleepiness [1]. In 2008, the National Highway Traffic Safety Administration estimates that 100 000 police reports on vehicle crashes were direct results of driver drowsiness resulting in 1550 deaths, 71 000 injuries, and $12.5 billion in monetary losses [2]. Driver inattention might be the result of a lack of alertness when driving due to driver drowsiness and distraction. Driver distraction occurs when an object or event draws a person's attention away from the driving task. Unlike driver distraction, driver drowsiness involves no triggering event but, instead, is characterized by a progressive withdrawal of attention from the road and traffic demands. Both driver drowsiness and distraction, however, might have the same effects, i.e., decreased driving performance, longer reaction time, and an increased risk of crash involvement.

Three main approaches have been developed to detect driver inattention, i.e., physiological, driving-behavior-based, and visual-feature-based approaches. Physiological approaches involve analysis of vital signals such as brain activity, heart rate, and pulse rate. As an example, Khushaba *et al.* [3] developed a fuzzy mutual-information-based wavelet packet transform model to estimate the drowsiness level from a set of electroencephalogram, electrooculogram, and electrocardiogram signals. However, physiological approaches often require electrodes that are attached to the driver's body, which are intrusive in nature and, therefore, may cause annoyance to the driver. Driving-behavior-information-based approaches evaluate the driver's performance over time. Based on the variations in the lateral position, speed, steering wheel angle, acceleration, and breaking, the system determines if the driver is alert or not. Liang *et al.* [4] developed a real-time approach to detecting distraction using the driver's eye movements and driving performance data collected in a simulator environment called the in-vehicle information system. Then, the data were used to train and test both support vector machine (SVM) and logistic regression models to detect driver distraction. The analysis by Liang *et al.* suggested that the SVM outperformed the traditional approach of logistic regression in detecting driver distraction. An advantage of this approach is its convenient signal acquisition. However, they highly depend on the vehicle type, driver experience, and the road condition. If a driver falls asleep on a straight road, such systems may fail because the car would not provide any significant information. The feature-based approach analyzes visual features from the driver's facial images. Drowsy people often produce unique visual features on the face such as eye blinking, yawning, and eye and head movements. Hammoud *et al.* [5] proposed a driver drowsiness detection system that estimates the status of the eyes in the near-infrared spectrum. Moriyama *et al.* [6] estimated the eye state by creating detailed templates of the shape and texture of the eyelid. As a widely accepted visual measure for drowsiness detection, the percentage of eyelid closure (PC) counts the number of eye blinks of the driver [7]. More recently, Jimenez *et al.* [8] have proposed a gaze fixation system based on a stereo camera system to detect the driver's distraction level in a driving simulator. From the viewpoint of practical applications, visual-feature-based approaches are preferred since they are natural and inherently nonintrusive to the driver.

Manuscript received January 22, 2013; revised April 17, 2013; accepted May 2, 2013. Date of publication May 22, 2013; date of current version August 28, 2013. The Associate Editor for this paper was R. I. Hammoud.

R. Oyini Mbouna and S. G. Kong are with Temple University, Philadelphia, PA 19122 USA (e-mail: oyini@temple.edu; skong@temple.edu).

M.-G. Chun is with the Department of Electronics Engineering, Chungbuk National University, Cheongju 361-763, Korea (e-mail: mgchun@chungbuk.ac.kr).

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.2013.2262098

This paper presents visual analysis of eye state and head pose (HP) using a single camera for continuous monitoring of alertness of a vehicle driver without the use of additional source of light. The proposed scheme finds in real time the eye and pupil centers and HP angles from a face object in a live video stream captured by a camera. The proposed method brings eye state and HP together to make a decision if a driver is not alert. We detect the face and the eyes using the AdaBoost algorithm [9] followed by iterative thresholding. Candidate pupil regions after thresholding are validated by a set of predefined geometric constraints to locate the pupil even in varying illumination conditions. Finally, the center of gravity of the pupil region finds the center of the pupil. A facial-feature-matching algorithm estimates three Euler angles of HP, i.e., nodding, shaking, and tilting, using a generic 3-D head model aligned with a 2-D face image. To avoid the error being accumulated from the matching of facial features on the 3-D head model and the 2-D face image, the matching process is reinitialized whenever the face comes back to its frontal position. Then, we compute visual features such as eye index (EI), pupil activity (PA), and HP from a video segment of 4-s duration. EI and PA measure eye closure and the rate of pupil movement over time. HP, which is a linear combination of the number of video segments with a large deviation of three HP angles, finds the amount of head movements that accounts for drowsiness and distraction. The t-test was used to validate the statistical significance of individual features. The significance levels of the three HP angles are used to scale the weights of the linear combination of the HP angles. Compared with other approaches that only use a discrete number of gaze fixation areas [8], the proposed approach considers all directional head and eye movements of the driver. An SVM classifier, which is trained with the three visual features of EI, PA, and HP, is used to learn the driving patterns of the driver to classify if the subject is either alert or nonalert. The nonalert state represents that the driver is either drowsy or distracted. Experiment results show that the driver's head and eye information helps achieve a better performance for driver drowsiness detection. Experiments were conducted using a total of 135 000 video frames from five test subjects of various ethnicity and gender in real road driving conditions.

II. Eye and Pupil Center Detection

A. Pupil Center Detection

The proposed pupil center detection method tracks the pupil center in real time. We first detect the face object in the scene using the face AdaBoost technique, which is known as the Viola–Jones algorithm [9], and adaptive template matching. Although accurate, the AdaBoost algorithm is sensitive to face rotation. Therefore, we implement adaptive template matching to overcome the limitation of AdaBoost in detecting the face, particularly when the head rotates. In adaptive template matching, the previously detected face region is used as template T. In the next processing cycle, we match template T against each pixel within a search window in the next image frame I. Then, the normalized sum-of-squares difference S is used as a metric of match between the face template and the search region of the face from the previous frame, i.e.,

$$S(x,y) = \frac{\sum_{x',y'} [T(x',y') - I(x+x',y+y')]^2}{\sqrt{\sum_{x',y'} T^2(x',y') \sum_{x',y'} I^2(x+x',y+y')}} \quad (1)$$

where $T(x,y)$ and $I(x,y)$ denote the brightness intensity of template T and source image I at (x,y). The matched image becomes a new template to be used for template matching in the following frame.

To detect the center of the pupil, the detected face region is divided into four quadrants to reduce the computational burden. The top two parts are considered as the region containing the eyes. To increase the resolution of the eye region, a candidate eye image is upsampled by two. The binarized image is eroded and normalized using morphological operators to reduce the effect of illumination variations. The pixels of intensity below a threshold are labeled as the pupil. An adaptive thresholding scheme was used to minimize the effect of illumination over time. An initial threshold is selected such that the threshold is large enough to preserve the pupil area. Then, the image is iteratively thresholded until eye geometric constraints are satisfied. We created geometric constraints after analyzing the physical characteristics of 1521 eye images [10]. The geometric constraints pose the conditions on the shape of the eye that an eye width is approximately twice the eye height and that a pupil width is not larger than twice the pupil height. The pupil center is then estimated by computing the center of gravity of the pupil region. When more than one candidate pupil region exists, the largest region is selected as the candidate region for the pupil. For pupil region image $I(x,y)$, pupil center (x_c, y_c) can be found using the spatial moments defined as

$$m_{pq} = \sum_x \sum_y I(x,y) x^p y^q. \quad (2)$$

We computed the pupil center as the first-order spatial moment m_{10} and m_{01} divided by the area m_{00}: $x_c = m_{10}/m_{00}$ and $y_c = m_{01}/m_{00}$. The center of gravity has shown better results in determining the center of the pupil than the center of the contour area, particularly under varying illumination conditions. In Fig. 1(a), eyelashes and illumination often cause a long shadow around the pupil blob to obscure the pupil. If we select the center of the blob contour as the center, the pupil center would be off center due to a nonconvex shape owing to a long tail of the region. Taking the center of gravity is closer to the real pupil center since it reduces the error caused by the tail. Even a closed eye shows a candidate pupil region after thresholding due to a dark region by shadow and eyelash. However, the region was not considered as a pupil since the contour does not satisfy the aspect ratio constraints.

B. Pupil Center Detection Results

We evaluate the accuracy of the proposed pupil center detection using the BioID face database [10]. The BioID database consists of face images in a practical setting with various

(a) (b) (c)

Fig. 1. Pupil center detection for three cases of eye state. The contour lines show detected pupil regions after applying adaptive thresholding. (a) Open. (b) Half closed. (c) Closed.

TABLE I
COMPARISONS OF PUPIL CENTER DETECTION ALGORITHMS

Pupil Detection Methods	Algorithm	Detection Rate
Jesorsky et al. (2001) [11]	Adaboost	91.8%
Zhou and Geng (2004) [10]	Generalized Projection Function (GPF)	94.8%
Asadifard and Shanbezadeh (2010)* [12]	Cumulative Distribution Function (CDF)	96.0%
Oyini Mbouna and Kong (2011)* [13]	Adaboost & Adaptive Thresholding	97.2%

illumination conditions, backgrounds, and face sizes. The database contains 1521 gray scale images of 23 different persons in a 384 × 286 pixel resolution. Error d_{eye} refers to the relative deviation between the estimated and the ground truth pupil centers, i.e.,

$$d_{\text{eye}} = \frac{\max\left(|C_{\text{left}} - \hat{C}_{\text{left}}|\,|C_{\text{right}} - \hat{C}_{\text{right}}|\right)}{|C_{\text{left}} - C_{\text{right}}|} \tag{3}$$

where $|C_{\text{left}} - \hat{C}_{\text{left}}|$ and $|C_{\text{right}} - \hat{C}_{\text{right}}|$ denote the Euclidean distances between the true pupil center positions C_{left} and C_{right} and the estimated pupil center positions $\hat{C}_{\text{left}}$ and $\hat{C}_{\text{right}}$. We consider that the eye was successfully detected if the relative error d_{eye} is less than a threshold of 0.25, as used in [10]. Table I shows the results obtained using the proposed algorithm in comparison with the methods published using the same BioID database. The bottom two methods marked with * did not use the images with eyeglasses in the database.

III. HEAD POSE ESTIMATION

A. HP Estimation

We compare a face image of unknown HP angle in a 2-D scene with a 3-D face model rotated by known Euler angles to estimate the HP angles. A 3-D head model is constructed using the 3-D computer graphics software *Blender* [14]. The use of a generic 3-D head model not only reduces the amount of computation but provides continuous HP estimates in all three rotation directions as well. We align and scale the 3-D head model according to the position and distance between the two eyes of the face in the 2-D image. The alignment of the 2-D face and the 3-D head model is carried out during the first few seconds when the driver maintains the face straight at the camera.

From the mapping of the 2-D facial features in the image and their corresponding points on the 3-D head model, we determine the HP (rotation and translation) of the user using the POSIT algorithm [15]. The relationship between a point on the 3-D head model and a point in the 2-D image is expressed as

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{23} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4}$$

where (X, Y, Z) denote the coordinates of a 3-D point in the world coordinate space, and (u, v) denote the coordinates of the projection point in pixels. The camera matrix contains intrinsic parameters such as image center (c_x, c_y), scale factor (s), and focal length (f_x, f_y) in pixels. The joint rotation–translation matrix $[R|t]$ contains extrinsic parameters r_{ij} and t_i. The POSIT algorithm is used to estimate the position in three dimensions of a known object. The 3-D pose of an object includes three rotation angles (nodding, shaking, and tilting) and translation. The POSIT algorithm requires image coordinates of at least four noncoplanar object's points. The 3-D model coordinates of these points must be known as well. In this case, the image coordinates of the object are the coordinates of the detected features on the face, and the 3-D model coordinates of these points are their corresponding points on the 3-D head model.

To track the changes in the head position, we apply the Lucas–Kanade optical flow method [16]. Optical flow finds an estimate of the feature points between two video frames extracted using the good features to track method [17]. The number of features selected varies between 20 and 40. If the number of features is greater than 40, the feature points are not robust enough for tracking. For the features less than 20, the tracking result becomes sensitive to the features. The HP is updated after each movement and given in the form of a matrix that can be viewed as a multiplication of three rotations: one about each principle axis. Thus

$$R = R_z(\varphi)R_y(\theta)R_x(\gamma). \tag{5}$$

Given rotation matrix R, HP angles γ, θ, and φ are then computed by equating each element in R with its corresponding element from the rotation matrix [18].

TABLE II
MAE OF THE PROPOSED HP ANGLE ESTIMATION (IN DEGREES)

Illumination	Tilting	Shaking	Nodding
Uniform	3.779	3.943	4.834
Varying	5.054	6.334	5.315
Overall	3.940	5.161	5.312

TABLE III
COMPARISON OF THE MAEs OF HP ANGLES (IN DEGREES)

Head Pose Estimation Methods	Initialization	Tilting	Shaking	Nodding
La Cascia et al. (2000) [19]	AUTO	3.3	6.1	9.8
Choi et al. (2008) [18]	MANUAL	3.92	4.04	6.71
Prasad and Aravind (2010) [21]	MANUAL	2.5	3.8	3.6
DeMenthon and Davis (1995) [15]	AUTO	5.27	6.00	6.23
Proposed Method	AUTO	3.94	5.16	5.31

B. HP Estimation Results

We used the public database of Boston University (BU), Boston, MA, USA [19], to evaluate the performance of the proposed HP estimation scheme. The BU database contains 72 video clips of free head movements of various subjects along with ground truth data of precise position and orientation of the head measured using a magnetic sensor. A set of 45 video sequences in uniform lighting conditions and a set of 27 videos in varying illumination conditions are obtained. Each video consists of 200 video frames at a rate of 30 frames per second. Table II shows the results of the head estimation method in various illumination conditions. The estimated angles and ground truth values are compared for all the 72 video data in terms of the average mean absolute error (MAE). The average MAE of tilting, shaking, and nodding angles were (3.779°, 3.943°, 4.834°) for 45 uniform illumination videos and (5.054°, 6.334°, 5.315°) for 27 varying illumination videos. The MAE decreased by an average of 1.3° due to illumination variations.

The overall performance of the proposed HP estimation scheme was similar and, in some cases, better than the other algorithms. Table III compares the HP estimation approaches. Our HP estimation method outperforms the DeMenthon and Davis algorithm [15] and other algorithms using auto reinitialization. The two manual initialization methods showed decent performances; however, they map the features to a 3-D model manually.

IV. DRIVER DROWSINESS DETECTION

A. Video Data Acquisition

We conducted experiments in a vehicle during the day with a camera mounted on the dashboard. A total of 15 video clips were collected from five subjects of different ethnicity and gender, i.e., African male (Subject 1), Asian female (Subject 2), Caucasian male (Subject 3), Indian male (Subject 4), and African male (Subject 5), with a study time of 15 min per subject. The video data acquired have a resolution of 640 × 480 pixels at a frame rate of 30 frames/s. The video data for alertness and nonalertness were created in real road driving conditions. Each subject was requested to look straight for the first 5 s and then drive alert, drowsy, or distracted for the remainder of the trip. Each driver completes three driving sessions, namely, alert, drowsy, and distracted. In the case of alert, the subjects were recorded while driving 10 mi at a speed varying from 0 to 65 mi/h on a road involving many cars and stimuli. In the case of drowsy or distracted driving sessions, the test subject was sitting on the passenger seat simulating to be driving while another person was actually driving for safety reasons. The camera was then placed in front of the passenger seat at the same angle as the camera on the driver side. Similarly to the alertness session, the subject was recorded being drowsy or distracted for 10 mi of driving with car vibrations. Then, each video was cropped to three scenes of continuous 5 min long. The drowsy sessions involved very few road stimuli on the highway, and the distracted driving sessions involved answering to phone calls, multitasking, texting while driving, or reading maps. As a result, each video session is 5 min long at a frame rate of 30 frames/s. Each video session is divided into 75 segments. Each segment is 4 s long, corresponding to 120 frames with a frame rate of 30 frames/s. The first 37 segments are used for training, and the remaining 38 segments are used for testing.

The ground truth of the alertness level was labeled using a binary sleepiness scale such that a "0" is assigned to a nonalert (drowsy or distracted) segment and a "1" to an alert segment. The test subjects were requested to assess their level of alertness. This information was mixed with the alertness scores obtained from the opinions of four human experts. Those five observers rated video segments of 4-s duration, which corresponds to 120 video frames, and assigned "0" (nonalert) or "1" (alert) to each segment. Then, using a majority decision, each video segment is assigned to a ground-truth label that a majority of observers agree on to each segment. An analysis window of 4-s duration is used to process the extracted features according to the driver's manual of the State of Pennsylvania [22], which recommends to the drivers to allow 4 s to reduce the risk of getting involved in a collision. The chosen window size is suitable for the driver drowsiness detection problem because the time window is large enough to contain sufficient information and small enough to capture the changes in driving behaviors. Most drowsiness detection systems have a delay between the moment the driver starts his fatigue behavior and the moment the system detects it. By choosing a window segment as short as 4 s, the proposed method has a minimal delay.

B. Feature Selection

Five features obtained from the eye and HP are used to determine alertness. We propose as new alertness detection measures the EI and PA that describe the state of the eye and

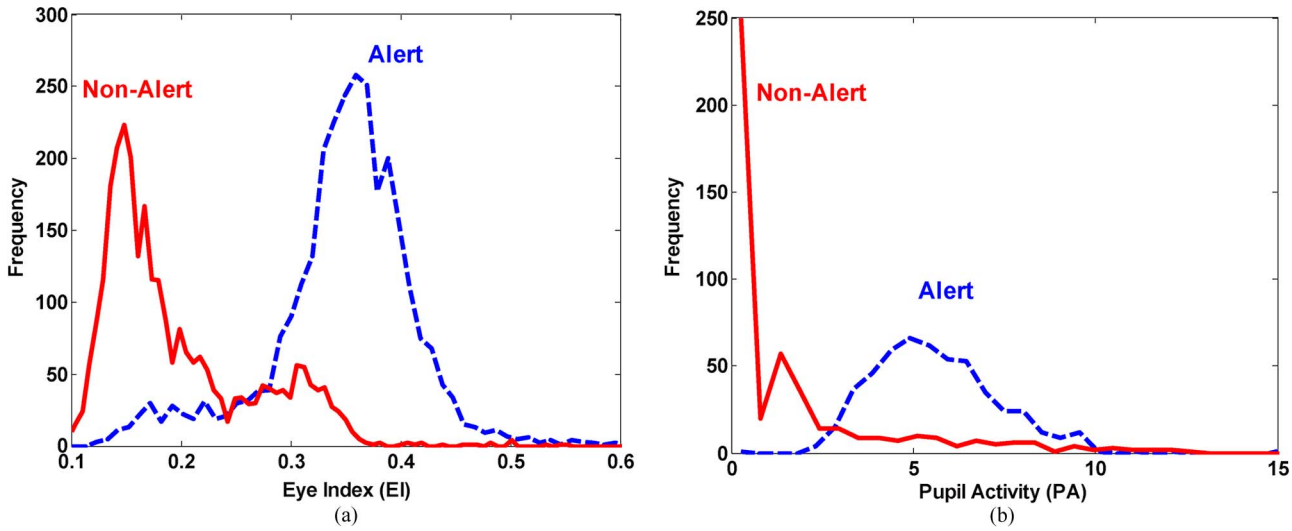

Fig. 2. Histograms of the eye features for Subject 4. (a) EI. (b) PA.

the pupil. EI is defined as the ratio of the pupil height and the height of the eye in pixels, i.e.,

$$\text{EI} = \frac{\text{Pupil Height}}{\text{Eye Height}}. \tag{6}$$

The pupil tends to show an isotropic shape when the eye is open. When the eye is half closed, the eye region becomes similar to a more rectangular shape. When the eye is closed, on the other hand, the detected eye region becomes to a flat and long shape that goes beyond the geometric constraint to be a pupil. The thresholds for the EI are chosen to determine the three states of the eye, i.e., open, half closed, and closed. We determine that an eye is closed if EI $<$ thresh1, open if EI $>$ thresh2, and half closed if thresh1 $<$ EI $<$ thresh2. For example, the thresholds are initially chosen for a subject shown in Fig. 1 as thresh1 $= 0.28$ and thresh2 $= 0.33$. In Fig. 1, the EI values were $0.35(= 38/108)$ for an open eye, 0.29 (EI $= 32/110$) for a half-closed eye, and $0.24(= 26/108)$ for a closed eye. Fig. 2(a) shows the histogram of the EI for alert and nonalert states. For the drowsiness case, the histogram shows a prominent bimodal distribution with a second peak at approximately 0.15, which corresponds to a closed eye. A major peak was observed at 0.33. This reveals the fact that the drowsiness state involves more video frames containing half-closed eyes.

PA measures the temporal activity of eye movements, which gives useful information to determine drowsiness. The coordinates (p_x, p_y) in pixels represent the pupil center with the eye center as a reference point. The relative displacement of the pupil between the two consecutive video frames is $(\Delta p_x, \Delta p_y)$, where $\Delta p_x = |p_x(t+1) - p_x(t)|$ and $\Delta p_y = |p_y(t+1) - p_y(t)|$. Then, the PA index is defined as the sum of the average displacements of the pupil movements in horizontal and vertical directions. When a subject is drowsy or distracted, the PA value tends to be greater. Thus

$$\text{PA} = \Delta\overline{p}_x + \Delta\overline{p}_y. \tag{7}$$

Fig. 2 shows the histograms of the eye features, i.e., EI and PA. The features reveal clear distinctions between alert and nonalert cases. The histograms are generated for all the video frames of Subject 4 where a pupil center is detected.

PC is a popular metric for monitoring fatigue of the driver and is even used by the U.S. Federal Highway Administration [23]. PC finds the ratio of the number of video frames containing a closed eye and the total number of frames within a period of time. Thus

$$\text{PC} = \frac{\text{Number of frames containing closed eye}}{\text{Number of total frames}}. \tag{8}$$

We apply the t-test to select statistically significant features and to reduce the dimensionality prior to the classification process. The objective of the t-test is to test if a feature is statistically significant by checking if the means of Class 1 (alert) and Class 2 (nonalert) are sufficiently separated in a two-class classification problem. Let $x_i, i = 1, 2, \ldots, N$ denote classification results of Class 1 (alert) with mean μ_1 and y_i, $i = 1, 2, \ldots, N$ of Class 2 (nonalert) with mean μ_2 obtained from the feature. N denotes the number of samples used. The classification result is either 1 for alert (A) or 0 for nonalert (NA). Let the null hypothesis be

$$H_0 : \mu_1 - \mu_2 = 0. \tag{9}$$

To test the null hypothesis, we compute the t-statistic, i.e.,

$$t = \frac{(\overline{x} - \overline{y}) - (\mu_1 - \mu_2)}{s_z\sqrt{2/N}} \tag{10}$$

$$s_z^2 = \frac{1}{2N-2}\left(\sum_{i=1}^{N}(x_i - \overline{x})^2 + \sum_{i=1}^{N}(y_i - \overline{y})^2\right). \tag{11}$$

The next step is to compute acceptance interval D based on significance level ρ. We pick $\rho = 0.05$, which means that acceptance interval D is 95% of the t distribution. On one hand, if the t-statistic falls into the D interval, it means that the test result agrees with the null hypothesis, and the feature is not selected. On the other hand, if the t-statistic lies outside interval D, it means that the test result rejects the null hypothesis, and the feature is selected. The t-test result shows that all the six

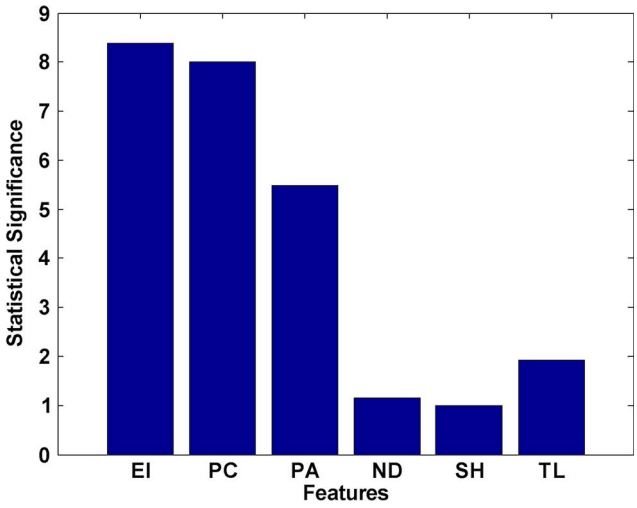

Fig. 3. Statistical significance of the features using the t-test for all subjects.

features rejected the null hypothesis with statistical significance and, therefore, are all selected for driver alertness detection. Fig. 3 shows the statistical significance level of individual features. The most significant measure was found to be EI, which replaced PC since both features account for eye closure.

Changes in HP angles provide good information to determine the drowsiness and distraction of the driver. When a person is drowsy, the nodding angle is expected to be high compared with a person that is fully alert and is keeping his head straight. Therefore, we count the number of consecutive frames when the absolute value of the nodding, shaking, and tilting angles are greater than 15°, such as

$$\text{HP} = w_1 ND + w_2 SH + w_3 TL \tag{12}$$

where ND, SH, and TL denote the number of consecutive video segments of nodding, shaking, and tilting, respectively, that exceed a threshold of 15°. When a person is distracted and looking away from the road, the HP difference is expected to be high compared with a person that is driving alert and staring at the road up front. The reason why the HP is high when distracted is because HP takes into account not only the HP being large but the duration of consecutive frames the HP stays large as well. The weights are proportional to the statistical significance and are normalized $(w_1 + w_2 + w_3 = 1)$. Each subject is then individually attributed weight values used for both training and testing.

The feature selection results confirm that both eye features and HP angles have statistical significance. However, the proposed EI measure outperforms the widely used PC measure with all test subjects. The PA measure records the dynamic motion of the driver's eye. The HP measure represents the head movements from a linear combination of the rates of nodding, tilting, and shaking angles that exceed 15°. Fig. 4 shows the clusters of the features in 3-D feature space of EI, PA, and HP for Subject 1 and Subject 2. The features show a good separation between the two classes. That clear separation between the two classes indicates that the classifier has a good chance to accurately distinguish the two classes.

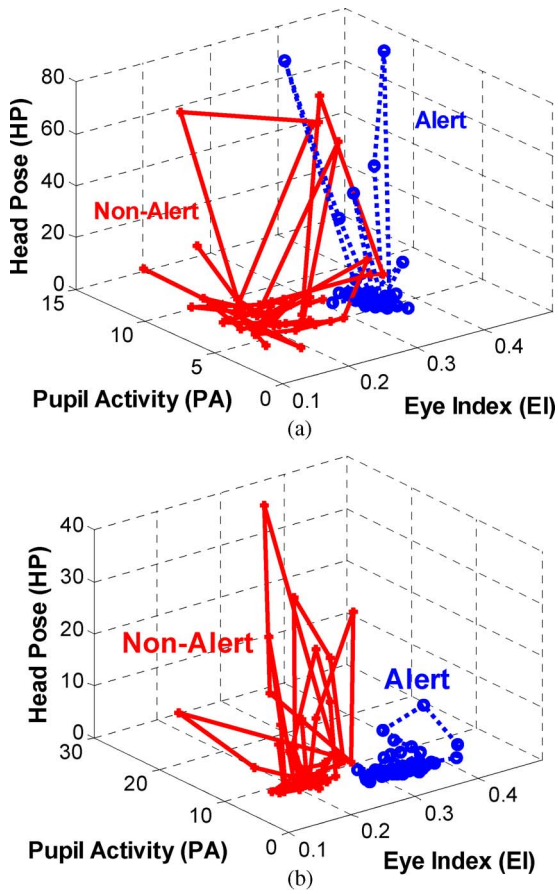

Fig. 4. Clusters of the features in 3-D feature space of EI, PA, and HP. (a) Subject 1. (b) Subject 2.

C. Classification Results

Driver alertness detection is formulated as a two-class classifier problem. Given expert scores associated with a member of each class, the goal is to find optimal function f that separates the two classes. The SVM finds the optimal decision boundary with a maximum separating margin [24]. We assume that we have a data set D of M points in n-dimensional space belonging to two classes, i.e.,

$$D = \{(\mathbf{x}_i, y_i) | i = 1, \ldots, M, \mathbf{x}_i \in \mathbb{R}^n, y_i \in \{1, 0\}\}. \tag{13}$$

A binary classifier should find function f that maps the points from their data space to their label space. The optimal hyperplane with the maximal separating margin is

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{14}$$

where N is the total number of support vectors; $\mathbf{x}_i$ denotes the support vectors; b and α_i are the solutions of the quadratic programming problem, as defined in [24]; and $K(\mathbf{x}, \mathbf{y})$ is a positive definite symmetric function that must satisfy Mercer's conditions. The training points that are closest to the optimal separating hyperplane with $\alpha_i > 0$ are called support vectors. All other training examples are irrelevant for determining the optimal hyperplane. In this paper, we use nonlinear machines to find the hyperplane that minimize the number of errors

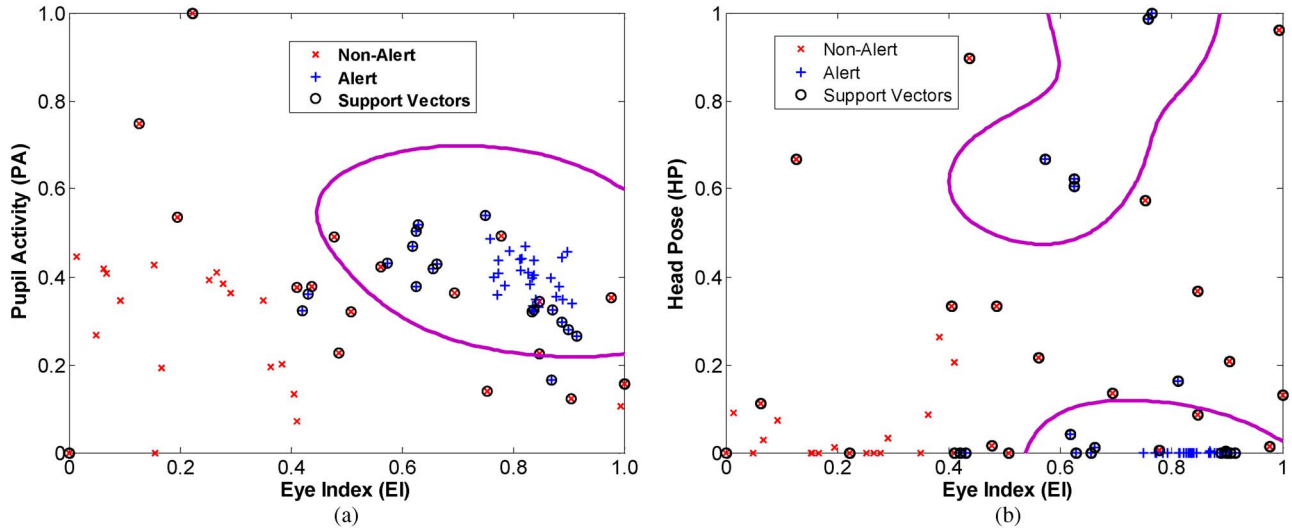

Fig. 5. Decision boundaries of the SVM classifier for two nonlinearly separable classes using two features for Subject 1. (a) EI versus PA. (b) EI versus HP.

TABLE IV
CONFUSION MATRIX

Classification	S1	S2	S3	S4	S5	Average
NA / NA	92.86	100.00	93.02	98.11	90.19	94.84
A / A	78.43	90.48	72.20	100	95.24	87.27
NA / A (Type I)	7.43	0.00	6.98	1.89	9.81	5.16
A / NA (Type II)	21.57	9.52	27.80	0.00	4.76	12.73

for the training set. Kernel function $K(\ )$ defines the nature of the decision surface that separates the data. Based on the nature of our data, we picked a nonlinear kernel function that is equivalent to a radial basis function (RBF) classifier defined as

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}\|^2}{\sigma^2}\right) \tag{15}$$

where σ^2 denotes the width of the RBF kernel. Fig. 5 shows classification boundaries using the SVM in 2-D feature space, i.e., EI versus PA and EI versus HP. Fig. 5 is generated with the alert and drowsy video segments of Subject 1. The solid line represents a nonlinear decision boundary that maximizes the separating margin from the boundary to the closest data points (support vectors) of each class.

Table IV summarizes the classification results for every subject. The first two rows indicate two cases of correct classification, where both actual and ground truth agree in terms of the classification label. Rows 3 and 4 denote two cases of misclassification, where actual and ground-truth classification labels do not coincide. There can be two types of classification errors.

Type-I Error (NA/A): The driver was classified as nonalert (NA), whereas ground truth is alert (A).

Type-II Error (A/NA): The driver was classified as alert (A), whereas ground truth is nonalert (NA).

Type-I error is considered more significant since it represents the case where the system fails to recognize that the driver is actually nonalert. Based on those two types of errors, we generate the confusion matrix for every subject individually and then for everybody together using every feature, including EI, PA, and HP. In this experiment, PC is not included because EI is closely related to PC. Moreover, EI performs better than PC, as shown in Table V. We observe from Table IV that the most crucial error is in average 5.16% for each person, and in the case of Subject 2, the crucial error is 0%.

TABLE V
TYPE-I ERROR RATES FOR DIFFERENT COMBINATIONS OF THE FEATURES

Features Used	S1	S2	S3	S4	S5	Average
PC only	42.86	0.0	37.21	3.77	29.73	22.71
EI only	28.57	0.0	37.21	11.32	27.73	20.97
EI + PA	25.0	0.0	18.60	1.89	21.62	13.42
EI + PA + HP	7.14	0.0	6.98	1.89	9.81	5.16

The combination of eye and HP information performs better than using only eye information. Table V presents Type-I error rates of the SVM classifier results when using only a single feature (PC or EI), when using only eye information (EI + PA), and when using all three major features (EI + PC + HP). Overall, EI performs better than PC. On average, EI has an error rate of 20.97%, whereas PC has a higher error rate of 22.71% for the same group of drivers. In addition, it is clear that using multiple features achieves better performance than using a single feature. The combination of all eye information (EI + PA) does not achieve the best performance. The addition of a third feature HP achieves the best performance.

V. CONCLUSION

This paper has presented visual analysis of eye state and HP using a single camera for continuous monitoring of alertness of a vehicle driver. The proposed scheme extracts visual features from the eyes and head movements of a driver in real outdoor driving conditions. The t-test ranked the features in terms of statistical significance. EI measures eye closures, PA finds

dynamic motion of the eye, and HP calculates all directional head movements. The three visual features, namely, EI, PA, and HP, are extracted in every video frame and averaged for a video segment of 120 frames or 4 s, following the "four seconds rule" according to the Pennsylvania Driver's Manual [22]. Four experts and the driver rated the video segments and attributed a label to the alertness level. Then, the final class label was obtained using majority voting. An SVM classifier was then used to identify the alertness level of each driver for every video segment of 4 s. The classification results indicate that combining eye and head information achieves the highest classification accuracy. Using the three statistically significant features, namely, EI, PA, and HP, the SVM classifier shows a low Type-I error, which is more critical than a Type-II error or a false alarm.

REFERENCES

[1] J. C. Williams, "2012 Sleep in America poll: Transportation workers' sleep," Nat. Sleep Foundation, Arlington, VA, USA, Tech. Rep., 2012.

[2] T. A. Ranney, E. Mazzae, R. Garrot, and M. J. Goodman, "NHTSA driver distraction research: Past, present, and future," Nat. Highway Traffic Safety Admin., East Liberty, OH, USA, Tech. Rep., 2000.

[3] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.

[4] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Trans. Transp. Syst.*, vol. 8, no. 2, pp. 340–350, Jun. 2007.

[5] R. I. Hammoud, G. Witt, R. Dufour, A. Wilhelm, and T. Newman, "On driver eye closure recognition for commercial vehicles," *SAE Int. J. Commercial Veh.*, vol. 1, no. 1, pp. 454–463, Apr. 2009.

[6] T. Moriyama, T. Kanade, J. Xiao, and J. F. Cohn, "Meticulously detailed eye region model and its application to analysis of facial features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 738–752, May 2006.

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

[8] P. Jimenez, L. M. Bergasa, J. Nuevo, N. Hernandez, and I. G. Daza, "Gaze distraction system for the evaluation of driver distractions induced by IVIS," *IEEE Trans. Transp. Syst.*, vol. 13, no. 3, pp. 1167–1178, Sep. 2012.

[9] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[10] O. Jesorsky, K. Kirchberg, and R. Frischholz, "Robust face detection using the Hausdorff distance," in *Proc. Audio Video Based Biometric Person Authentification*, 2001, vol. 3, pp. 90–95.

[11] Z. Zhou and X. Geng, "Projection functions for eye detection," *Pattern Recognit.*, vol. 37, no. 5, pp. 1049–1056, May 2004.

[12] M. Asadifard and J. Shanbezadeh, "Automatic adaptive center of pupil detection using face detection and CDF analysis," in *Proc. Int. MultiConf. Eng. Comput. Scientists*, 2010, vol. 1, pp. 130–133.

[13] R. Oyini Mbouna and S. G. Kong, "Pupil center detection with a single webcam for gaze tracking," *J. Meas. Sci. Instrum.*, vol. 3, no. 2, pp. 133–136, 2012.

[14] V. Gumster, *The Complete Guide to Blender Graphic: Computer Modeling and Animation*. Boca Raton, FL, USA: CRC, 2012.

[15] D. DeMenthon and L. S. Davis, "Model-based object pose in 25 lines of code," *Int. J. Comput. Vis.*, vol. 15, no. 1/2, pp. 123–141, Jun. 1995.

[16] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.

[17] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 593–600.

[18] G. G. Slabaugh, "Computing Euler angles from a rotation matrix," City Univ. London, London, U.K., Tech. Rep., 1999.

[19] M. La Cascia, S. Sclaro, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 322–336, Apr. 2000.

[20] S. Choi and D. Kim, "Robust head tracking using 3D ellipsoidal head model in particle filter," *Pattern Recognit.*, vol. 41, no. 9, pp. 2901–2915, Sep. 2008.

[21] B. H. P. Prasad and R. Aravind, "A robust head pose estimation system for uncalibrated monocular videos," in *Proc. Indian Conf. Comput. Vis., Graph. Image Process.*, 2010, pp. 162–169.

[22] B. J. Schoch, *Pennsylvania Driver's Manual*. Harrisburg, PA, USA: Pennsylvania Dept. Transp., 1995, p. 28.

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

[24] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Amsterdam, The Netherlands: Elsevier, 2009.

Ralph Oyini Mbouna received the B.S. and M.S. degrees in electrical and computer engineering from Temple University, Philadelphia, PA, USA, in 2012, where he is currently working toward the Ph.D. degree in engineering.

His current research interests include driver monitoring system, 3-D face reconstruction, biometrics, pattern recognition, and computer vision.

Seong G. Kong (SM'12) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1982 and 1987, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1991.

He was an Assistant Professor from 1992 to 1995 and an Associate Professor from 1996 to 2000 with the Department of Electrical Engineering, Soongsil University, Seoul, where he also served as the Department Chair from 1998 to 2000. During 2000–2001, he was a Visiting Scholar with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. From 2002 to 2007, he was an Associate Professor with the Department of Electrical and Computer Engineering, The University of Tennessee, Knoxville, TN, USA. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA, USA. His research interests include pattern recognition, image processing, and intelligent systems.

Dr. Kong received the Best Paper Award from the International Conference on Pattern Recognition in 2004, the Honorable Mention Paper Award from the American Society of Agricultural and Biological Engineers, the Professional Development Award from the University of Tennessee in 2005, and the Most Cited Paper Award from *Computer Vision and Image Understanding* in 2007 and 2008. He served as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, a Guest Editor for a Special Issue of the *International Journal of Control, Automation, and Systems*, and a Program Committee Member of various international conferences. He is a member of the International Society for Optics and Photonics.

Myung-Geun Chun received the B.S. degree in electronics engineering from Pusan National University, Pusan, Korea, in 1987 and the M.S. and Ph.D. degrees in intelligent systems from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 1989 and 1993, respectively.

He was a Senior Researcher with Samsung Electronics. He is currently a Professor with the Department of Electronic Engineering, Chungbuk National University, Cheongju, Korea. His research interests include the design and development of intelligent systems for human–computer interaction based on thermal and visual image processing techniques and the design of biometric systems having the capability for privacy protection.

Dr. Chun served as an Editor for the Joint Technical Committee 1 of the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC JTC 1) SC27 24745 project "Biometric Information Protection."