Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State

Bappaditya Mandal, Liyuan Li, Gang Sam Wang, and Jie Lin

***Abstract*—Driver's fatigue is one of the major causes of traffic accidents, particularly for drivers of large vehicles (such as buses and heavy trucks) due to prolonged driving periods and boredom in working conditions. In this paper, we propose a vision-based fatigue detection system for bus driver monitoring, which is easy and flexible for deployment in buses and large vehicles. The system consists of modules of head-shoulder detection, face detection, eye detection, eye openness estimation, fusion, drowsiness measure percentage of eyelid closure (PERCLOS) estimation, and fatigue level classification. The core innovative techniques are as follows: 1) an approach to estimate the continuous level of eye openness based on spectral regression; and 2) a fusion algorithm to estimate the eye state based on adaptive integration on the multimodel detections of both eyes. A robust measure of PERCLOS on the continuous level of eye openness is defined, and the driver states are classified on it. In experiments, systematic evaluations and analysis of proposed algorithms, as well as comparison with ground truth on PERCLOS measurements, are performed. The experimental results show the advantages of the system on accuracy and robustness for the challenging situations when a camera of an oblique viewing angle to the driver's face is used for driving state monitoring.**

***Index Terms*—Driver monitoring, fatigue detection, fusion, machine learning, percentage of eyelid closure (PERCLOS), spectral regression, video analytics.**

I. Introduction

FATIGUE, drowsiness and sleepiness are often used synonymously in driving state description [1]. Involving multiple human factors, it is multidimensional in nature that researchers have found difficult to define over past decades [2]–[5]. Despite the ambiguity surrounding fatigue, it is a critical factor for driving safety. Studies have shown that fatigue is one of the leading contributing factors in traffic accidents worldwide [6]. It is particularly critical for occupational drivers, such as drivers of buses and heavy trucks, due to the fact that they may have to work over a prolonged duration of the driving task, during the peak drowsiness periods (i.e., 2:00 A.M. to 6:00 A.M. and 2:00 P.M. to 4:00 P.M.), and under monotonous or boredom working conditions [7], [8].

Research to detect driver drowsiness can be classified into three categories: 1) vehicle-based approaches, 2) behavior-based approaches, and 3) physiological-signal based approaches (see [7], [9] for a good review). In physiological approaches, the physiological signals from a body, such as electroencephalogram (EEG) for brain activity, electrooculogram (EOG) for eye movement, and electrocardiogram (ECG) for heart rate, are evaluated to detect driver drowsiness [10]–[15]. Recent studies show that the methods using physiological signals (specially the EEG signal) can achieve better reliability and accuracy of driver drowsiness detection compared to other methods [16]. However, the intrusive nature of measuring physiological signals can hinder driving, especially for prolonged driving periods. Vehicle-based approaches collect signal data from sensors in vehicles to evaluate driver's performance. These methods monitor the variations of steering wheel angle, lane position, speed, acceleration, and braking to predict the driver fatigue [17]–[21]. It is convenient to collect vehicle signals. However, these approaches might be too slow to detect driver drowsiness [7]. Behavior-based approaches depend on vision analysis to monitor driver's behavior, including eye-closure, eye-blinking, yawning, head pose, hand gesture, etc., through a camera directed to driver's face [22]–[27]. The driver is alerted if a drowsiness symptom is detected. The vision-based systems on behavior analysis are attractive to automobile industries since they are non-intrusive to the driver and the measures are effective and reliable to predict driver fatigue [28].

A drowsy driver displays a number of symptoms, including frequent eye-closure, rapid and constant blinking, nodding or swinging head, and frequent yawning [29]. In the last decade, numerous vision systems have been developed to detect such behaviors of drowsiness for driving safety. Most of the existing systems require the installation of a camera directly toward the driver's face to capture high-resolution face images, and some of them employ specifically designed infra-red (IR) cameras [23], [30] or stereo cameras [31], [32]. The vision algorithms are designed for high-resolution front-view face and eye images (e.g., the height of the face is over 60% of image height in input images over 640×480 pixels). This configuration is not applicable for buses and large vehicles. A bus mostly has a large front glass window to let the driver have a wide-field-view of scene for safe driving since it is much wider than cars. Placing a camera on the front glass window is not practical, and that also blocks the drivers' view. If the camera is installed on the frame around the window, the camera is not able to capture the frontal view of driver's face, so that existing vision algorithms are not applicable.

Manuscript received July 15, 2015; revised December 16, 2015 and April 20, 2016; accepted June 8, 2016. Date of publication August 5, 2016; date of current version February 24, 2017. The Associate Editor for this paper was L. M. Bergasa.

B. Mandal, L. Li, and J. Lin are with the Department of Visual Computing, Institute for Infocomm Research (I^2R), Agency for Science, Technology and Research (A∗STAR), Singapore 138632 (e-mail: bmandal@i2r.a-star.edu.sg; lyli@i2r.a-star.edu.sg; lin-j@i2r.a-star.edu.sg).

G. S. Wang is with the Department of Intelligent Transportation Systems, Institute for Infocomm Research (I^2R), Agency for Science, Technology and Research (A∗STAR), Singapore 138632 (e-mail: gswang@i2r.a-star.edu.sg).

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.2016.2582900

Fig. 1. Existing dome camera in the bus and example images of bus drivers captured by the dome cameras in buses. Only an oblique view of low-resolution face images can be captured by the existing dome cameras for normal driving poses. (Images came from open sources of bus service companies with face portions of the drivers being pixelated for keeping anonymity.).

In this paper, we present a novel vision system for bus driver monitoring. It is designed for easy and flexible deployment on existing cameras in buses with no extra hardware cost required. In most existing buses, there are already many dome cameras installed for security purposes. One is mounted in the upper-right or upper-left position with respect to the driver to record the driver behavior on duty, as shown in the left image in Fig. 1. Since this camera is first installed for recording of driving behavior, a wide-view dome camera is used to capture the visible upper body of the driver, as shown in the rest images in Fig. 1. This configuration poses three challenges to vision-based driver monitoring:

1) Oblique-view: In most normal driving poses, only oblique-view face images can be captured by the camera (e.g., $30^\circ \sim 40^\circ$). In this case, the vision algorithms designed for a frontal view of faces are not applicable.
2) Low-resolution: From the images which capture the full visible part of the driver body by a wide-view dome camera, only low-resolution face and eye images can be obtained (in the raw input image, the height of the face is of 80$\sim$110 pixels). In such case, the feature-point based approaches would often fail, and binary classification of eye state is not reliable.
3) Pose variation: The bus driver has to move and turn his upper body and head to look around when driving a large vehicle. In this case, the approaches based on face detection and tracking often fail to locate the head and face in the image.

A novel vision system is proposed to deal with these challenges, which integrates upper-body detection, face detection, eye detection, eye openness estimation, fusion, and symptom measure estimation. The framework is illustrated in Fig. 2. The core innovative algorithms are eye-openness estimation and fusion. We propose a manifold learning algorithm which can learn a mapping from a low-resolution eye image (e.g., 32 $\times$ 24 pixels) to a 1-dimensional continuous level of eye-openness. There are two advantages of this approach. First, there is no need to detect eye feature points for symptom estimation that often fails on low-resolution face images. Second, it avoids classifying the ambiguous partially closed eyes into open or close state for binary state (0/1) classification. This improves the accuracy of identifying "drowsy" state between "normal" and "sleepy" states. To obtain an accurate and robust estimate of eye openness, a novel fusion algorithm is proposed which adaptively integrates the results of eye openness estimations on the multi-model eye detections for both eyes. Based on the innovative techniques, the system achieves robust performance on the challenging scenarios where the existing approaches often fail.

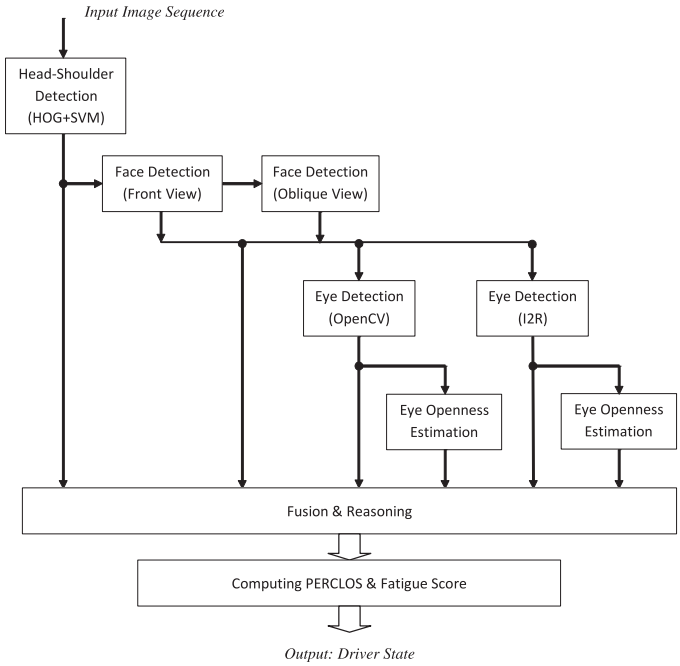


Fig. 2. Framework of the proposed system.

The main contribution of this paper is a novel vision-based system for bus driver fatigue detection which is applicable to low-resolution face images captured from an oblique viewing angle to the driver's face, so that it can share a wide-view camera mounted for driver's full body behavior monitoring. The technological contributions can be summarized as follows:

- A novel framework for vision-based driver fatigue detection which integrates head-shoulder detection, multi-pose face detection, multi-model eye detection, eye openness estimation, fusion, and PERCLOS estimation for driver fatigue detection;
- A manifold learning algorithm to learn a mapping from a low-resolution eye image to a continuous level of eye-openness;
- A fusion algorithm to obtain an accurate and robust eye openness estimate based on adaptive integration on multi-model eye detections on both eyes;
- A refined approach to compute PERCLOS measure based on the continuous levels of eye-openness.

The next subsection discusses some related work. Section II presents the proposed approach and system, including an overview description and details of each module. Experimental results and evaluations are presented in Section III and conclusions are given in Section IV.

A. Related Work

Many vision-based systems have been developed for driver fatigue detection based on driver face monitoring [9], [33], [34]. These systems employ a camera mounted on the dashboard that points directly towards driver's face and focus on detecting the open eyes of the driver in high-resolution face images. They can be categorized on the camera used and approach for facial and eye feature computing. IR cameras and normal color cameras have been employed. On the other hand, both feature-point based approaches and binary classification approaches have been investigated for eye state estimation. A review and evaluation of emerging driver fatigue detection measures and technologies can be found in [35] and testing of commercially available fatigue monitors are reported in [36].

In [22], [23], a specific IR camera is designed to capture the high reflection of eye, i.e., "red eye." Multiple visual cues such as eyelid, gaze, head movements and facial expressions are fed to a trained Bayesian Network for estimating fatigue level of the driver. IR cameras have also been used in several other systems for open eye detection, gaze direction estimation, and head pose tracking for driver fatigue detection [24], [26], [30], [37]. The approaches on IR cameras exploit the high reflection from pupil to detect eyes. They are not applicable if the camera is not placed directly in front of the driver's face.

Using visible light, numerous methods for fatigue detection based on facial feature point detection are reported [32], [38], [39]. These approaches detect facial feature points, e.g., the corners of eye, the upper and lower positions of eyelids, etc., to recognize closed eyes [40]. In [41], a number of facial features, including those of the eyes, mouth, and gaze, are integrated to measure the driver's vigilance level. A robust approach not depending on face detection is proposed in [42], which consists of eye detection, validation, and open eye recognition. An interesting scheme using eye index, pupil activity and head pose is proposed in [25] for continuous monitoring of alertness of a vehicle driver. Using support vector machine (SVM), video segments are classified into alert or non-alert driving events. Their classification results indicate that combining eye and head information achieves a better classification accuracy.

The classification approaches train a binary classifier, e.g., SVM, Adaboost, nearest neighbor (NN), etc., to classify the high-resolution eye image into open or close state. In [30], after eye localization, the eye states are classified as open and close by an SVM. Pradhan *et al.* proposed to learn feature subspaces of eigen-eyes to classify eye states [43]. Various classifiers, such as k-NN and SVM, are evaluated in [39]. A real time vision system is proposed in [26] to monitor drivers during day and night time. Face detection and tracking are performed using Haar-like features and Kalman filtering respectively. Principal Component Analysis (PCA) is used to detect eyes during day-time whereas Local Binary Pattern (LBP) is used for nighttime. Finally, eye state is classified as open or close using SVM. Under binary eye state classification framework, the ambiguous partially closed eyes may cause either too many false alarms or a lower detection rate. To the best of our knowledge, there is no literary on estimating a continuous level of eye openness and applying it to compute PERCLOS measurement for driver fatigue detection.

On the other hand, in computer vision, there are investigations on face detection [44], head pose estimation [45], [46], and eye detection [47] in low-resolution images. However, no work on eye state analysis on low-resolution image is reported, due to the challenges of detecting tiny feature points around the eyes in low-resolution images. Head-shoulder detection has been investigated for video surveillance [48], [49]. However, attempt to employ the technique for driver monitoring has not been found since almost all vision-based driver monitoring start from face detection or eye detection [9].

Most vision-based methods infer the states of fatigue and sleepiness based on the symptoms extracted from the eye regions [9]. Measurements estimated from eye regions include percentage of eyelid closure (PERCLOS) [26], [30], [32], eyelid distance [50], eye blink speed [51], eye blink rate [24] and gaze direction [22]. Face tracking [52], head nodding [24] and head orientation [51] are also used for driver fatigue detection. Based on detailed investigations, PERCLOS is proven to be a reliable measure of driver drowsiness from images [28], [53].

II. The Proposed Method and System

In this Section, first, an overview of the system framework is presented, and then, each module is described. The details are focused on the innovative eye-openness estimation, fusion, and PERCLOS computing.

A. System Overview

The system framework is illustrated in Fig. 2. There are six main steps in the process. First, a head-shoulder detector is applied to detect the presence of a driver and locate roughly the position of the driver's head. Then, two models of face detectors are used to detect a front-view face or an oblique-view face within the region of the head. Third, two eye detection methods are employed to locate the potential eye positions and scales in the image. In the fourth step, our proposed eye openness estimation method is applied to the located eyes by the two eye detectors. Next, a fusion operation is proposed to obtain an accurate and robust estimate of driver's eye openness based on adaptive integration on multi-model eye detections for both eyes. The score of driver's fatigue, i.e., PERCLOS, is computed on the recent records of eye openness over a specified period. Details are described in the following subsections.

B. Head-Shoulder Detector for Driver Detection

Since the camera observes the driver's face from an oblique viewing point, and the bus driver may change his body pose and head direction to check the surrounding situations when approaching an intersection or a bus stop, a system starting

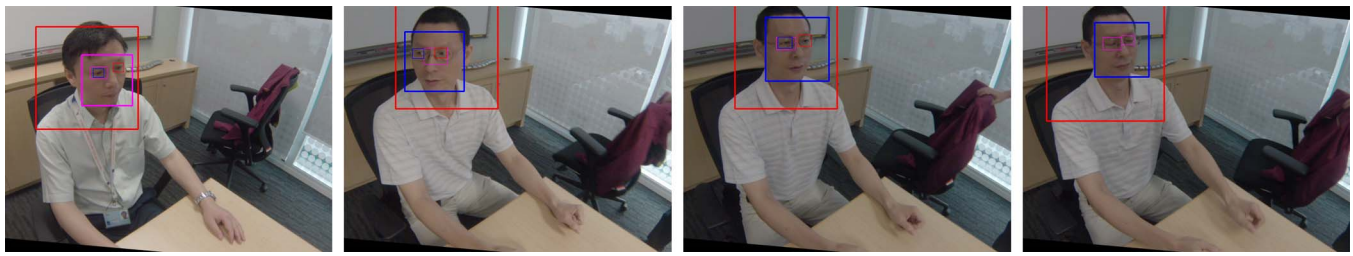


Fig. 3. Examples of eye detections using the OpenCV eye detector (CV-ED) [54] and our previous eye detector (I2R-ED) [55]. The test images are captured under the similar configuration in buses, as shown in Fig. 1. In each image, the large red box indicates head-shoulder detection, the pink box over face indicates the detection by OpenCV face detector, and the blue box indicates the detection by the additional face detector. Around the eyes, the pink small boxes are the detections by the OpenCV eye detector, the blue and red boxes are the pair of the detected left and right eyes by our eye detector [55]. (From left to right) (a) CV-ED detected only one (left) eye and failed to detect the right eye, whereas I2R-ED succeeded for both eyes. (b) CV-ED failed to detect both eyes correctly, whereas I2R-ED succeeded for both eyes. (c) CV-ED mislocated the right eye around the left eye, whereas I2R-ED succeeded for both eyes. (d) CV-ED succeeded for closed eyes, but I2R-ED missed both closed eyes.

from face detection and tracking may fail to localize the driver's head in the input images. In this system, we build a head-shoulder detector to detect the presence of a driver on seat and locate the driver's head position in the incoming image. A HOG (histogram of oriented gradients) descriptor [56] is designed to capture the appearance features of sitting drivers, and an SVM classifier is trained for driver detection. In this application, the scene (the seat of driver in buses) and object (sitting drivers) are limited compared with pedestrian detection in natural scenes [56]. Hence, a small training set, as long as that covers large variation of drivers' appearances and inner scenes in buses, can generate a detector of very good performance. From the result of head-shoulder detection, we can obtain roughly the position and scale of the driver's head in the image.

C. Face Detection

Over the region of head, face detectors are applied to find a face looking towards the front of the bus. First, the OpenCV face detector [54] is applied. It is very robust to find the front-view faces since it is trained with a huge number of samples. However, it mostly fails to detect the oblique-view faces observed by the camera as shown in Fig. 1. In a ten-minute video recording of a bus driver on duty, the OpenCV detector can only detect about 40% of the driver's faces in the normal driving poses. Therefore, we train an additional face detector specially for the oblique-view faces using the OpenCV face detection algorithm, i.e., the algorithm based on Harr features and Adaboost classifier [54]. This face detector will be applied if the OpenCV face detector fails to find a face in the head position. If both face detectors fail to find a face in the head position, it means that the bus driver may have turned his face away, which may indicate an abnormal driving state. Since the face detectors are only applied to the head regions, much less false positives are generated compared with the approach of scanning the whole image for face detection.

D. Eye Detection

Two eye detectors are applied to the rectangular region of the detected face. The OpenCV eye detector [54] performs well to locate the eyes in the front-view faces, even with closed eyes. But it often fails to locate the pair of eyes in the oblique-view faces. Since the driver's left eye is too close to the face boundary in the image, the OpenCV eye detector may miss it, or locate the driver's left eye on the right eye or mouth, as shown in the first 3 examples in Fig. 3. The results of OpenCV eye detection in this application situation can be summarized as: (a) no output, i.e., both eyes are missed; (b) one output, i.e., only the left eye is detected and the right eye closing to the face boundary is missed; (c) two outputs, i.e., two eyes are correctly detected or right eye is wrongly located on the left eye or mouth. We have developed an eye detector for human-robot interaction [55]. It combines both sketch and graph patterns of eye for eye detection with a trained SVM, and employs a Maximum-Likelihood algorithm to locate a pair of eyes on the multiple detections of eyes on face. It was refined for detecting eyes in oblique-view faces. It performs better to detect and locate a pair of open eyes in oblique-view faces than the OpenCV eye detector. But it might fail to find the correct locations of closed eyes, as shown in the last example in Fig. 3. The outputs of our eye detector can be summarized as: (a) no output, i.e., both eyes are missed; (b) a pair of outputs, i.e., two eyes are correctly detected or located at wrong positions due to closed eyes. The two eye detectors perform complementary roles, so that fusion on the two detectors results in a robust performance of eye state estimation as described later in Section II-F. More examples can be found in Fig. 8 in the Section III.

E. Eye Openness Estimation

Drowsy driver may not simply close his/her eyes frequently. He/she may be struggling to keep his/her eyes partially open. Hence, to accurately characterize driver drowsiness based on eye images, it is desirable to measure the eye-openness continuously. As described above, in this application scenario, we can only obtain small, low-resolution images of eyes in images. Feature-based techniques for eye image analysis are not applicable. We develop a novel method for eye-openness estimation based on the full image of the eye so that there is no need to detect feature points and curves around eyes. This method is applicable for low-resolution images of eyes.

From the result of eye detection, we obtain a normalized eye image of 32×24 pixels. We apply Spectral Regression

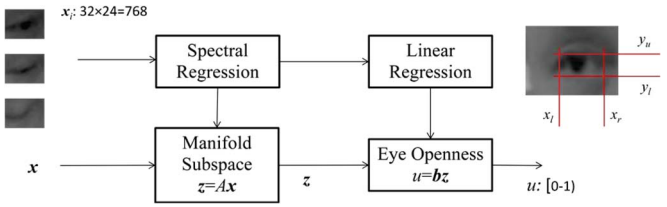


Fig. 4. The block diagram of using Spectral Regression to learn a mapping from raw image vector $\mathbf{x}$ to a continuous measure of eye-openness u, where the upper right picture shows the measurement of eye openness (see text for descriptions).

(SR) Embedding [57] to learn the eye shape model. The idea is illustrated in Fig. 4. Let I_E be a normalized eye image (left or right eye). It can be expressed as a 1D feature vector $\mathbf{x}$ of 768 elements. From the manually labeled four marks as shown by the upper-right picture in Fig. 4, i.e., the left and right corners of the eye (x_l and x_r), and the central positions of the upper and lower eyelids (y_u and y_l), the corresponding level of eye openness can be computed as

$$u = \frac{|y_l - y_u|}{|x_r - x_l|}. \tag{1}$$

Given m training samples $\{\mathbf{x}_i, u_i\}_{i=1}^m$ ($\mathbf{x}_i \subset R^n$ and $n = 32 \times 24$), we wish to find a transformation A to map $\mathbf{x}$ into a d-dimensional manifold subspace $\mathbf{z}$ ($d \ll n$) with SR Embedding (i.e., $\mathbf{x} \to \mathbf{z} = A^T\mathbf{x}$), and then find a mapping $\mathbf{z} \to u$ to estimate the level of eye openness.

Let W be a $m \times m$ symmetric similarity matrix. Its element W_{ij} represents the similarity between samples $\mathbf{x}_i$ and $\mathbf{x}_j$. Different from existing SR approaches which define the similarity weight based on either the distance between sample points (i.e., unsupervised learning on a continuous similarity weight) or sample class labels (i.e., supervised learning on a binary similarity weight), we define the similarity weight based on the ground truth of eye openness. The similarity weight is defined as

$$W_{\text{ij}} = \exp\left(-|u_i - u_j|^2/\sigma_u^2\right) \tag{2}$$

where, if the eye-openness levels of two images are close, the similarity weight W_{ij} is high. Hence, our approach can be considered as a supervised SR learning on a continuous similarity weight. Suppose $\mathbf{y} = [y_1, y_2, \cdots, y_m]^T$ is the map from the graph to the real line. The optimal $\mathbf{y}$ tries to minimize

$$\sum_{i,j}(y_i - y_j)^2 W_{\text{ij}} = 2\mathbf{y}^T L\mathbf{y} \tag{3}$$

where $L = D - W$ and D is a diagonal matrix with $D_{\text{ii}} = \sum_j W_{\text{ji}}$. Formally, it can be expressed as

$$\mathbf{y}^* = \arg\min_{\mathbf{y}^T D\mathbf{y}=1} \mathbf{y}^T L\mathbf{y} = \arg\min \frac{\mathbf{y}^T L\mathbf{y}}{\mathbf{y}^T D\mathbf{y}} = \arg\max \frac{\mathbf{y}^T W\mathbf{y}}{\mathbf{y}^T D\mathbf{y}}.$$

This is equivalent to finding eigenvectors corresponding to the maximum eigenvalues arranged in descending order of eigen-problem $W\mathbf{y} = \lambda D\mathbf{y}$. Solving this problem, we obtain d eigenvectors $\{\mathbf{y}_k\}$ for top eigenvalues except the first one. For mapping for all samples including new test ones, we choose a linear function $y_i = f(\mathbf{x_i}) = \mathbf{a}^T\mathbf{x}'_i$, where $\mathbf{x}'_i = [\mathbf{x}_i^T, 1]^T$ is the augmented feature vector so that the linear mapping function has not to be constrained to pass the center of the feature space. The mapping vectors $\{\mathbf{a}_k\}_{k=1}^d$ can be obtained as the solution of regularized least square problem

$$\mathbf{a}_k = \arg\min_{\mathbf{a}} \left(\sum_{i=1}^m (\mathbf{a}^T\mathbf{x}'_i - y_i^k)^2 + \alpha\|\mathbf{a}\|^2 \right) \tag{4}$$

where $k = 1, \ldots, d$ and y_i^k is the ith element of $\mathbf{y}_k$. Let $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_d]$ as a $n \times d$ transformation matrix. Expressed in matrix form, one can denote

$$Y = [\mathbf{y}_1, \ldots, \mathbf{y}_m] \quad \text{and} \quad X = [\mathbf{x}'_1, \ldots, \mathbf{x}'_m]. \tag{5}$$

Then, the solution of A can be obtained as

$$A^T = YX^T(XX^T + \alpha I)^{-1} \tag{6}$$

where I is the identity matrix. The new input sample $\mathbf{x}$ can be embedded into the d-dimensional manifold subspace by $\mathbf{x} \to \mathbf{z} = A^T\mathbf{x}'$.

Finally, a linear model $u = \mathbf{b}\mathbf{z}'$ with $\mathbf{z}' = [\mathbf{z}^T, 1]^T$ is used to learn the mapping for eye openness measurement, which can be obtained by least square fitting. The shape model maps the 2D normalized eye image to a continuous 1D measurement of eye openness of which a small value (e.g., 0) means a closed eye while a large value (e.g., ≥ 0.5) indicates a wide open eye.

The similarity weight defined in (2) indicates the advantage of using manifold learning for this problem. We want to evaluate a continuous measure of eye openness from an eye image, and to be robust to variations of inter-person differences, head poses, and lightings. According to (2), the eye images of close openness levels have high similarity weights, so that they would be mapped to close positions in the learned manifold feature subspace. However, if two eye images are similar on visual appearance (e.g., the brightness), but the eye-openness levels are different significantly, they would be mapped to well separated positions in the learned manifold feature subspace. Hence, due to the advantage of spectral regression [57], the learned low-dimensional manifold feature subspace of eye-openness representation is smooth and reliable for continuous eye-openness measurement and adaptive to the variations caused by inter-person differences, such as age, gender, race, make-up, wearing glasses, etc., and changes of head poses and lighting conditions.

In this application, due to the oblique viewing angle to the driver's face, the eye shapes of left and right eyes are quite different. Hence, two models are trained separately for the left and right eyes. A few examples of eye openness estimation for left and right eyes are shown in Fig. 5. The openness levels of closed eyes are close to 0 and wide open eyes have higher openness level over 0.5. Fig. 5 shows some examples of eye openness estimation for both left eyes (upper row) and right eyes (lower row), along with ground truth and test values. Detailed evaluation of this model will be reported in Section III.

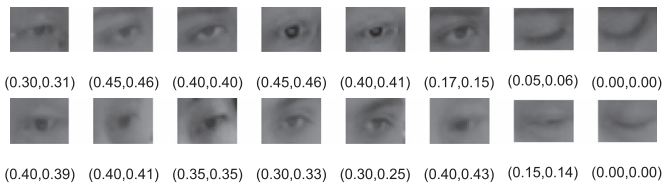


Fig. 5. Examples of eye-openness estimation. Left eyes (upper row) and right eyes (lower row) with continuous openness levels (ground-truth values, estimated values).

F. Fusion

Since two eye detectors (CV-ED and I2R-ED) are applied and each detector generates a pair of detection windows for the left and right eyes, four detection windows of eyes are obtained. Eye openness estimation is very accurate if the eye center is well localized. If the eye detector can locate the center of the detection box at the center of the pupil closely, the computed value of the eye openness level is very close to the ground truth value. It is observed that for the open eyes on the normal front-view and oblique-view faces, both OpenCV and I2R eye detectors can locate the two eyes quite well, among them the I2R eye detector performs slightly better to locate the centers of pupils due to the design of the model [55]. If the eyes are closed or face turns to one side, especially to the left side of the driver (i.e., turn to the right side from the view of camera), the detections of the two eye detectors would not match with each other. Since the driver's left eye is too close to the face boundary in the image, the OpenCV eye detector may miss it, or locate the driver's left eye on the right eye or mouth. I2R eye detector may miss the closed eyes or locate them at the wrong positions on the faces. To obtain an accurate and robust result of eye-openness estimation from the four eye detections, a novel fusion method is proposed, which adaptively integrates the results of eye-openness estimations obtained on the four eye detection windows. The details of the algorithm are described below.

First, let us denote

- $(x^{\mathrm{lc}}_{I2R}, y^{\mathrm{lc}}_{I2R})$: center of the left eye detected by I2R-ED;
- $(x^{\mathrm{rc}}_{I2R}, y^{\mathrm{rc}}_{I2R})$: center of the right eye detected by I2R-ED;
- $(x^{\mathrm{lc}}_{CV}, y^{\mathrm{lc}}_{CV})$: center of the left eye detected by CV-ED;
- $(x^{\mathrm{rc}}_{CV}, y^{\mathrm{rc}}_{CV})$: center of the right eye detected by CV-ED;

where the detected left and right eyes in the face image correspond to driver's right and left eyes, respectively. We define the agreement measures of the two eye detectors for the left and right eyes as

$$a_l = \exp\left[-\left(\frac{\left|x^{\mathrm{lc}}_{I2R} - x^{\mathrm{lc}}_{\mathrm{CV}}\right|^2}{\sigma_x^2} + \frac{\left|y^{\mathrm{lc}}_{I2R} - y^{\mathrm{lc}}_{\mathrm{CV}}\right|^2}{\sigma_y^2}\right)\right] \quad (7)$$

$$a_r = \exp\left[-\left(\frac{\left|x^{\mathrm{rc}}_{I2R} - x^{\mathrm{rc}}_{\mathrm{CV}}\right|^2}{\sigma_x^2} + \frac{\left|y^{\mathrm{rc}}_{I2R} - y^{\mathrm{rc}}_{\mathrm{CV}}\right|^2}{\sigma_y^2}\right)\right] \quad (8)$$

where σ_x and σ_y are the scale parameters which are determined by the scale of face window generated by the face detector. Specifically, let H_F be the height of the detected face window, the parameters are set as $\sigma_x = r_W H_F$ and $\sigma_y = r_H H_F$, where r_W and r_H are selected empirically based on face biometrics. Obviously, if an eye is accurately located by both eye detectors, the agreement measure for the eye is high (i.e., close to 1), otherwise, it is low (i.e., close to 0) which may be caused by poor head pose or closed eye. If the eye is detected just by one eye detector (i.e., CV-ED or I2R-ED), the agreement measure is set as 1.

In normal cases, we can obtain four eye detections from one face. When applying the method of eye openness estimation, one can obtain four estimates of the eye openness. Let us denote them as

- u^l_{I2R}: the estimate of eye openness level from the detection box of left eye generated by I2R eye detector;
- u^r_{I2R}: the estimate of eye openness level from the detection box of right eye generated by I2R eye detector;
- u^l_{CV}: the estimate of eye openness level from the detection box of left eye generated by OpenCV eye detector;
- u^r_{CV}: the estimate of eye openness level from the detection box of right eye generated by OpenCV eye detector.

The value of the estimated eye openness level varies in the range [0,1), where $u = 0$ indicates a closed eye. For a naturally open eye, u is within 0.3 to 0.5. If I2R-ED missed the eyes, mostly due to closed eyes or poor head poses, both u^l_{I2R} and u^r_{I2R} are set as 0. Also, if CV-ED missed the two eyes, u^l_{CV} and u^r_{CV} are set as 0. For an oblique-view face, the right eye on the face image (i.e., the driver's left eye) may be too close to the face boundary. CV-ED may miss the right eye, or locate both eyes overlapped on the left eye, or locate the right eye on the mouth. In these cases, u^r_{CV} is set to be 0. The latter two cases can be expressed as

- IF $(|x^{\mathrm{lc}}_{\mathrm{CV}} - x^{\mathrm{rc}}_{\mathrm{CV}}| < 2\sigma_x)(|y^{\mathrm{lc}}_{\mathrm{CV}} - y^{\mathrm{rc}}_{\mathrm{CV}}| < 2\sigma_y)$, THEN $u^r_{\mathrm{CV}} = 0$;
- IF $(|y^{\mathrm{lc}}_{\mathrm{CV}} - y^{\mathrm{rc}}_{\mathrm{CV}}| > 2\sigma_y)$, THEN $u^r_{\mathrm{CV}} = 0$.

The last condition means that the detected two eyes are largely separated along vertical direction, mostly due to locating the right eye on mouth.

Considering all the above situations, the final eye openness measure (eye state) can be obtained by adaptively integrating all the estimations on the eye detections. The fusion function is defined as

$$\tilde{u} = f(u^l_{I2R}, u^r_{I2R}, u^l_{\mathrm{CV}}, u^r_{\mathrm{CV}}) = k_l a_l(\beta_I u^l_{I2R} + \beta_C u^l_{\mathrm{CV}}) + k_r a_r(\beta_I u^r_{I2R} + \beta_C u^r_{\mathrm{CV}}) \quad (9)$$

where $k_l + k_r = 1$ and $\beta_I + \beta_C = 1$. These parameters are determined according to the reliability of each detector. In this work, the parameters are chosen empirically as $k_l = 0.67$ and $\beta_I = 0.6$. This selection is consistent with the observation that from the oblique viewing angle as shown in the two middle images in Fig. 1 and images in Fig. 3, the detections and estimations on the left eye in the image are more reliable than those from the right eye.

G. Measurement of Fatigue

PERCLOS is an important symptom of driver fatigue detection. PERCLOS is defined as the percentage of eyelid closure over the pupil over a specified time period. Specifically,

PERCLOS calculates the proportion of time within a specified time duration that the eyelid covers over 80% of the pupil [28] (80% and more of eye closure may correspond to 0.1 level and less of eye-openness). PERCLOS is still considered as the most effective measurement of drivers' drowsiness for vision-based non-intrusive approaches [1], [7], [9], [28].

Existing measures of PERCLOS are defined on hard thresholding on 80% of eye closure, which is an ambiguous measurement [58]. This may miss some important characteristics of driver drowsiness, such as slow roving eye movements between open and close, named partially open or partially closed [58]. In this work, since we can obtain a continuous measure of eye openness, we propose a refined definition of PERCLOS measure on soft classification of eye openness. Let $\tilde{u}_i$ denote the eye openness level obtained from (9) at time step i, and Δt_i be the time interval from time step i to $i+1$. At current time step t, the record of the eye opennesses over the most recent time period T can be expressed as $\{\tilde{u}_i\}_{i=t-N_t}^{t}$, where N_t is the minimum value to satisfy $\sum_{i=t-N_t}^{t} \Delta t_i \geq T$. PERCLOS is defined on the 80% of the eyelid closure. Obviously, this is a fuzzy measurement. Human factor research suggest the drowsiness metric on PERCLOS is around 70% to 80% of eye closure [28]. Based on the continuous measure of eye openness, we can characterize the 80% of eye closure for PERCLOS much more naturally than existing methods on hard classification of two eye states (i.e., open and close). The eye openness level is defined as (1). For a fully opened eye, the eye openness level u is around 0.5 according to the videos. Hence, 80% of the eyelid closure corresponds to $u = 0.1$. Based on this, a weight for eyelid closure is defined as

$$w_{\mathrm{ec}}(\tilde{u}_i) = \begin{cases} 1, & \text{when } \tilde{u}_i < 0.07; \\ 1 - \frac{\tilde{u}_i - 0.07}{0.06}, & \text{when } 0.07 \leq \tilde{u}_i \leq 0.13; \\ 0, & \text{when } \tilde{u}_i > 0.13. \end{cases} \quad (10)$$

This weight measure indicates that if the eye openness level $\tilde{u}_i$ is less than 0.1 (80% eyelid closure), there is high chance that the driver is fatigued. Now, the PERCLOS defined on eye openness can be computed as

$$\mathrm{PERCLOS}_t = \frac{\sum_{i=t-N_t}^{t} w_{\mathrm{ec}}(\tilde{u}_i)\Delta t_i}{\sum_{i=t-N_t}^{t} \Delta t_i}. \quad (11)$$

The driving states can be classified as *Normal* and *Fatigue* on PERCLOS values as

$$S_t = \begin{cases} 1, & \text{if } \mathrm{PERCLOS}_t \leq 0.2; \\ 0, & \text{Otherwise;} \end{cases} \quad (12)$$

where the threshold 0.2 corresponds to 80% of eye closures [28], [58]–[60].

III. Experiments and Evaluations

In this paper, we focus on bus driver fatigue detection from an existing surveillance camera in a bus. The datasets used to evaluate existing vision-based approaches to monitor a driver's facial and head behaviors from a camera directly pointing to the driver's face in a car are not applicable for evaluation of the proposed method and system. On the other hand, it is not safe and ethical to make a drowsy person drive a bus on real roads. We have obtained real-world videos of bus drivers on duty from a local company, but it is not allowed to use them for open report. To evaluate the performance of the proposed techniques and system for bus driver fatigue detection under real-world conditions, we capture dozens of simulated bus driving videos using two wide-angle cameras from the viewing point similar to the dome cameras mounted in existing buses. The simulation scenarios of driving bus are performed by 23 people under varying lighting conditions. They are asked to simulate both normal and drowsy driving over ten minutes. The participants include young and middle aged male and female persons, among them 6 people wear transparent glasses. During the simulated bus driving periods, we switch on/off a different number of lights and lower/raise curtains in the lab to simulate lighting changes in real-world conditions. In addition, to evaluate the performance of our system in real-world conditions, we also captured 3 simulated driving videos in moving car.

TABLE I
Average Errors of Eye-Openness Estimation

	TrainSet	TestSet	AutoSet	Phase B
LeftEye	0.029	0.124	0.132	0.137
RightEye	0.033	0.161	0.129	0.143

The system is developed and evaluated in two phases. In the first phase, we collect individual images to train and validate the eye openness estimation model. Images for training and validation are randomly selected from simulated driving sequences by 8 participants. This set of experiments is performed on individual images. In the second phase, we evaluate our system for driver fatigue detection according to PERCLOS measurement on novel sequences of simulated driving. This set of experiments is performed on whole video sequences of another 15 persons.

A. Experiments on Eye Openness Estimation

In the first phase, from 8 image sequences, we first randomly select 1068 images to form a training set (TrainSet), and then, from the rest of the images, we randomly select 337 frames according to even distribution to form a test set (TestSet). In both sets, the eye positions and the 4 marks for each eye (as illustrated by the upper-right picture in Fig. 4) are manually annotated, so that the ground truths of eye openness levels can be obtained. We train our model for eye openness estimation (i.e., the mapping matrices A and $\mathbf{b}$ in Section II-E) using the TrainSet, and then evaluate our model on TestSet. The average error is calculated as the average of the differences between the estimated values and the ground truths of eye openness levels. The average errors for left and right eyes in the images are listed in Table I, where, as for comparison, the average errors on both TrainSet and TestSet are included. We observe from the table that our model learns the mapping from the eye image to eye openness measure successfully on the TrainSet with the average errors less than 0.03 for the left eye and

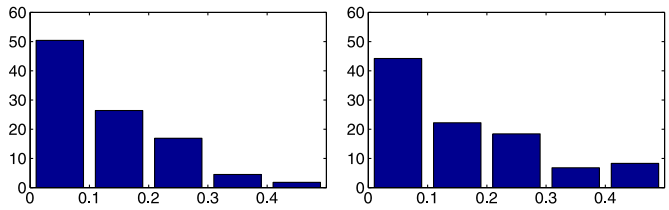


Fig. 6. The distributions of error levels of eye-openness estimations for the left eyes (left) and right eyes (right), where the vertical axis represents the percentage of errors, and the horizontal axis denotes the error segments.

0.033 for the right eye. Considering the low resolution of the eye images, this error level is close to the level of variations of ground truths annotated by human beings. There is a clear increase of errors when the learned model is applied to the novel dataset (TestSet), but the performance is still quite good. If more training samples are added, the performance can be improved further. The performance for the right eye is slightly poorer than that for the left eye because observing the face from the oblique angle, the right eye is smaller than the left eye and the viewing angle to the right eye differs largely from the face orientation compared to the viewing angle to the left eye. Due to that fact a few worse cases of estimations would result in a large increase of average error. For a close examination, we also generate the distributions of errors within 5 segments (i.e., $[0, 0.1)$, $[0.1, 0.2)$, $[0.2, 0.3)$, $[0.3, 0.4)$, $[0.4, \infty)$) for both left and right eyes, as shown in Fig. 6. It can be seen that about half of the errors are less than 0.1.

When applying our model in the proposed system, the eye openness estimation is performed on the eye boxes automatically located by OpenCV and I2R eye detectors. There are shifts of eye center positions and variations of eye box scales. We also build a dataset (AutoSet) which consists of 402 eye samples successfully localized by either OpenCV or I2R eye detectors. The samples are randomly collected from the rest images of the 8 sequences. The average errors of our model on this dataset are also listed in Table I. It can be found that the performance of our model on the set on automatic eye localization is very close to that obtained on manual annotation. From these results, it can be observed that in most cases the errors of eye openness estimations are quite small. The performance is accurate enough to distinguish the open and closed eyes and recognize eyes blinking. To verify the stability of the trained model, we also evaluated it on 150 samples randomly collected from the dataset used in next phase in Section III-B (10 frames from each video of each person). The results are shown in the rightmost column in Table I (Phase B).

The model for eye openness estimation is trained for large variations in lighting conditions. A few examples of training eye images which cover a large range of brightness are shown in Fig. 7. Our model succeeds to capture the distinctive features from opened eyes to closed eyes under varying lighting conditions, which can be observed from the low level of average errors on TrainSet in Table I. However, when applied to novel images, the error level will increase and it becomes worse due to shifts of the eye centers and variations of scales (i.e., the sizes of bounding boxes) automatically generated by two eye detectors. The proposed fusion approach makes the final estimation much more reliable and stable due to the fact that the final result depends on more reliable eye detections. A few examples of eye openness estimations after fusion with respect to good and poor eye detections are shown in Fig. 8.

Under each image in Fig. 8, the ground truth, estimated eye openness levels of both left and right eyes on the detections by I2R eye detector (I2R-ED) and OpenCV eye detector (CV-ED), as well as fusion results are listed. In the first example (a), the detection windows of both eye detectors match well. The next two examples (b)&(c) show the cases with good eye localizations by both eye detectors, where the fusion is a weighted average over four detections. In the fourth example (d), it can be found that the estimation of eye openness on the left eye detected by I2R-ED fails due to that the detection window shifts up quite large, meanwhile, CV-ED fails to detect the right eye. In the fifth example (e), the eye openness estimation on the left eye detected by CV-ED fails due to that the detection window is too large and the center shifts to the lower-left side. In both of such cases (d)&(e), the fusion algorithm can select the reliable ones and filter out failed estimations to generate the eye openness levels corresponding to normal open eyes. In the next two examples (f)&(g), I2R-ED fails for both eyes due to closed or nearly closed eyes, but the fusion algorithm can generate good estimations based on succeeded detections for closed and partially closed eyes. In the last example (h), CV-ED fails on both closed eyes due to a large oblique viewing angle, but I2R-ED succeeds to detect the two closed eyes. The fusion algorithm generates an accurate estimation. From these examples, it can be seen that the final estimations of eye openness after fusion are much more stable, reliable, and robust to failed or poor eye detections.

B. Experiments on Accuracy of PERCLOS Measurement

In the second phase, we evaluate the performance of our system for driver fatigue detection on whole video sequences. The test dataset consists of 15 long videos performed by another 15 different individuals. Each participant is told to simulate bus driving from normal to drowsy and then sleepy states gradually. Each video sequence lasts about 9 minutes.

To compare with state-of-the-art technologies, we implement a baseline method. The baseline method consists of three steps. In the first step, PCA is applied to the raw eye image to reduce the feature dimensionality. In the second step, an SVM (Support Vector Machine) classifier is trained to classify the eye image into two states, i.e., open and closed eyes. The SVM classifier is trained using all samples in TrainSet and TestSet (in which the samples are obtained according to manual annotations). Finally, PERCLOS measures are computed according to the percentage of recognized closed eye within recent 1 minute. This baseline method is similar to the approaches reported in [26], [30], [39].

It is not safe and practical to collect data of drowsy bus driving on real-world roads. To evaluate the effectiveness of our system for drowsy driving detection, we compared the PERCLOS generated by our method with that computed on ground truths of eye openness levels, since the PERCLOS has

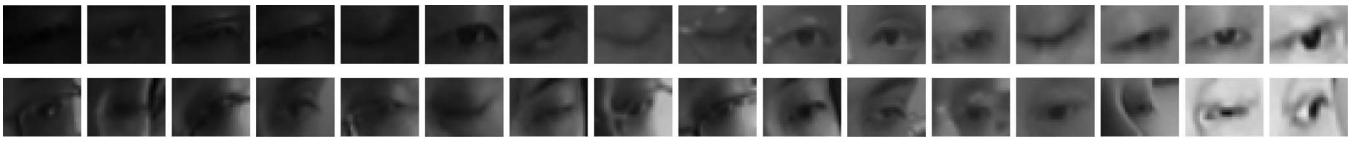


Fig. 7. Left eyes (top row) and Right eyes (bottom row) with illumination variations.

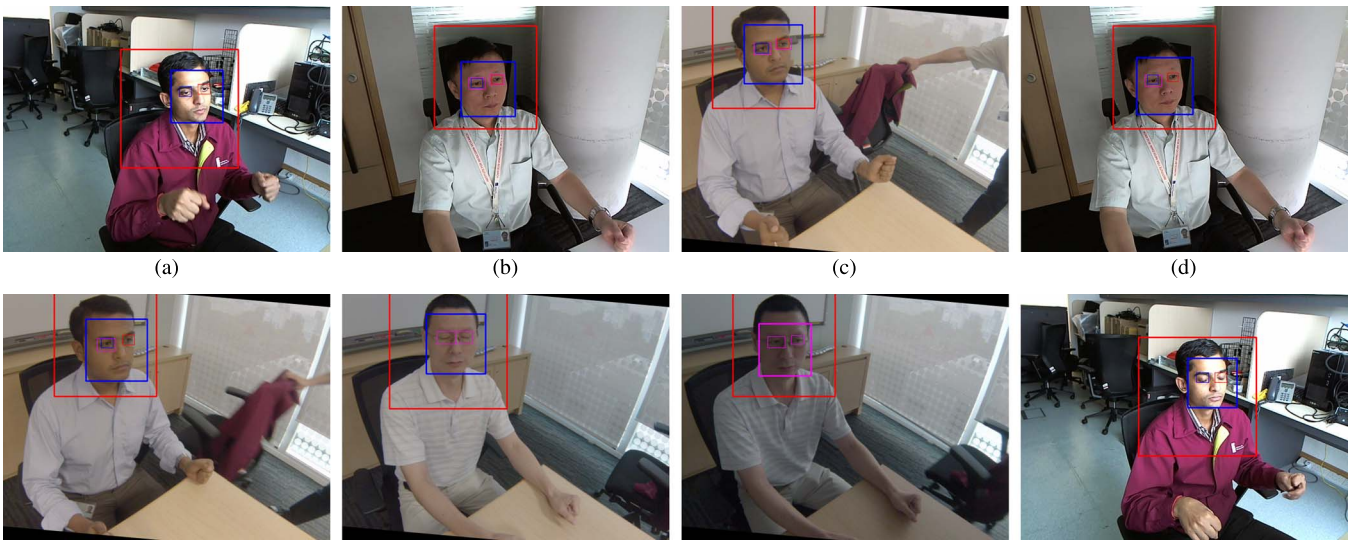


Fig. 8. Some examples to show the stability of eye-openness estimation after fusion, where GT means ground truth. (Best viewed in color). (a) Eye openness estimations: GT L: 0.55, R: 0.48; On I2R-ED L: 0.56, R: 0.49; On CV-ED L: 0.56, R: 0.49; Fusion: 0.53. (b) Eye openness estimations: GT L: 0.35, R: 0.18; On I2R-ED L: 0.31, R: 0.17; On CV-ED L: 0.37, R: 0.17; Fusion: 0.27. (c) Eye openness estimations: GT L: 0.36, R: 0.42; On I2R-ED L: 0.36, R: 0.44; On CV-ED L: 0.37, R: 0.44; Fusion: 0.37. (d) Eye openness estimations: GT L: 0.37, R: 0.28; On I2R-ED L: 0.02, R: 0.33; On CV-ED L: 0.48, R: 0.00; Fusion: 0.19. (e) Eye openness estimations: GT L: 0.48, R: 0.33; On I2R-ED L: 0.56, R: 0.49; On CV-ED L: 0.16, R: 0.00; Fusion: 0.29. (f) Eye openness estimations: GT L: 0.00, R: 0.00; On I2R-ED L: 0.00, R: 0.00; On CV-ED L: 0.00, R: 0.09; Fusion: 0.01.(g) Eye openness estimations: GT L: 0.13, R: 0.10; On I2R-ED L: 0.00, R: 0.00; On CV-ED L: 0.32, R: 0.11; Fusion: 0.10. (h) Eye openness estimations: GT L: 0.00, R: 0.00; On I2R-ED L: 0.00, R: 0.00; On CV-ED L: 0.00, R: 0.00; Fusion: 0.00.

been evaluated as an effective measure of driver fatigue on visual signal [7], [9], [28]. The ground truths of eye openness levels are annotated manually for the test videos.

The plots of the PERCLOS measurements against the number of frames and the corresponding classifications of driver states based on the PERCLOS scores for 4 typical sequences of the 15 test videos are shown in Fig. 9, where the odd plots from top to bottom are the PERCLOS measurements and the even plots are the corresponding classifications of driver states. In a plot of PERCLOS measurements, the blue curve represents the PERCLOS measures computed using the baseline method, the red curve represents the PERCLOS measures by the proposed system, and the green curve indicates the PERCLOS values computed on ground truths. It is observed that different persons may have different understandings and simulated performance of different driving states, i.e., normal, drowsy and sleepy driving states. The first person increases the times of eye blinks very slowly for about 5.7 minutes (at about #850 frame) and then increases the times and durations of eye closures in the rest of time duration quickly like falling asleep in a few minutes. Compared with the first person, the second person blinks slightly more frequently in the first half duration (about 4.5 minutes) and increases the eye blinks and eye closure durations quickly (from #580 frame) as if feeling very tired. As to the third and the fourth persons, they frequently blink their eyes from the beginning like a drowsy driver and finally fall into sleepy state after about 6 minutes (at about #900 frame).

The baseline method fails to detect drowsy and sleepy driving states for the third and fourth persons due to a large number of classification errors for eye states. For the second person, it is too sensitive due to a large number of false detection errors of closed eyes. The blue PERCLOS curve of the first person seems to be able to capture the increases of eye closure durations, but the difference of it with green curve of ground truth is still quite large. The large error rate of SVM classification may be caused by low-resolution of eye images, large oblique angle to the faces, and ambiguity of partial open eyes, as well as the variations of center positions and scales of eye detection windows. From the figure, one can observe that the red and green plots match quite well, especially when the PERCLOS values are small which is critical to distinct normal and fatigue drivers. If we define a distance measure between the PERCLOS curves as

$$D_{\mathrm{PERCLOS}} = \frac{1}{N}\sum_{i=1}^{N}\frac{|P_M(i) - P_G(i)|}{(P_G(i) + 0.2)} \tag{13}$$

where N is the number of images in Set2, $P_M(i)$ is the predicted PERCLOS value and $P_G(i)$ is the ground truth PERCLOS value for the ith image. The denominator is used for normalization (where the constant 0.2 corresponds to

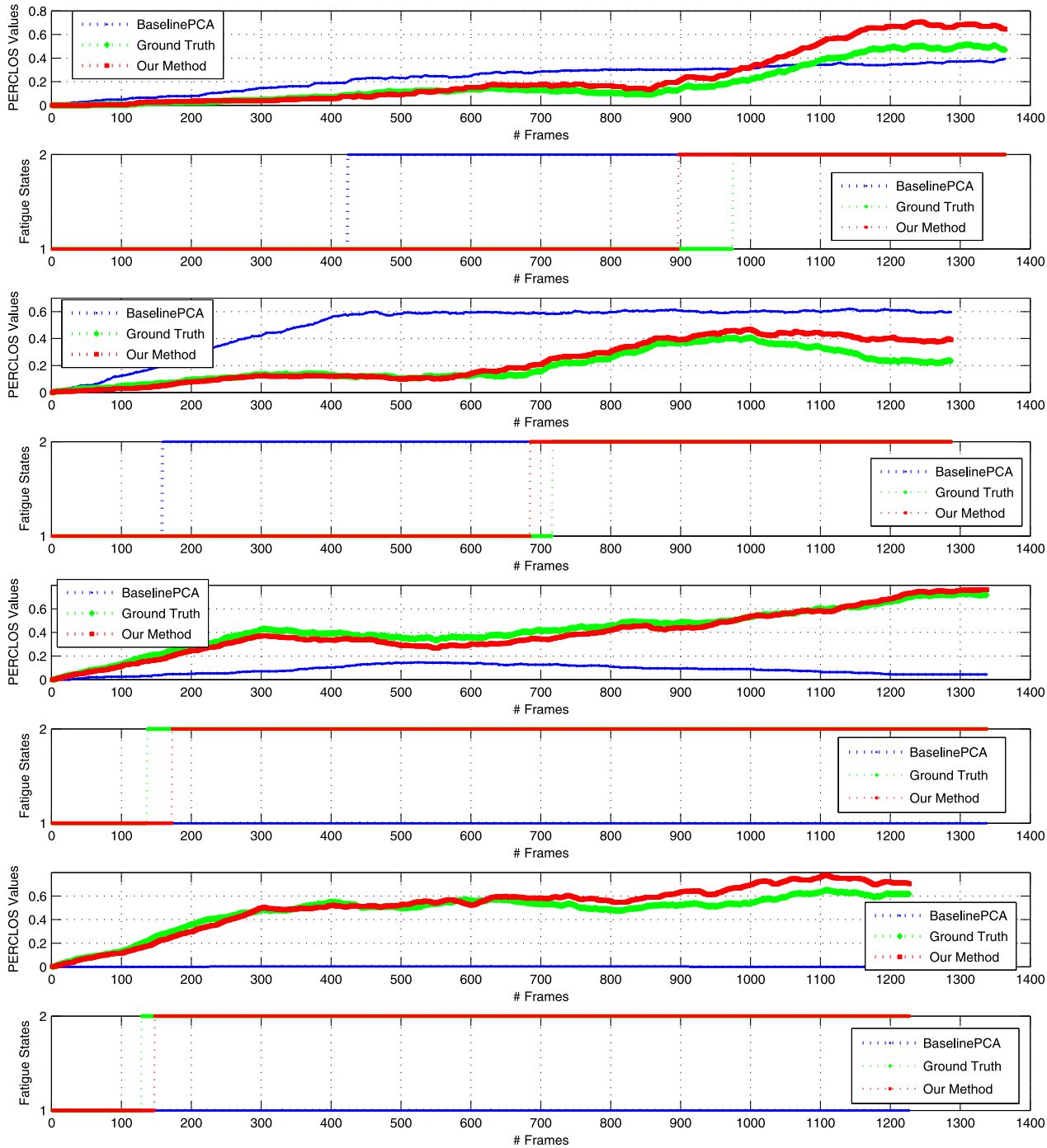


Fig. 9. PERCLOS estimations of four subjects using (10) and (11) (odd plots from top to bottom) and the corresponding state classifications (even plots from top to bottom). Different people perform differently, but, for all the simulated sequences, we can observe the increase of PERCLOS values from the simulated state of normal, drowsy, to sleepy. (Best viewed in color).

80% of eye closure) since a relative large difference for large $P_G(i)$ value has less impact for state classification. Using this measure, the average distance between our method and ground truth is 0.1223 and that for the baseline approach is 0.5308.

In the plots of the corresponding classifications of driver states based on PERCLOS scores, the "blue" plots represent the driving states obtained by the baseline method, the "red" plots represent the driving states generated by our method, and the "green" plots represent the driving states obtained from PERCLOS measures on ground truths. The state 1 represents normal state and the state 2 indicates drowsy and sleepy state. It can be observed that our method matches the ground truth well. If we define a matching rate for kth state as

$$M_k = \frac{\sum_{i=1}^{N}(B(S_i = k) \wedge B(S_i^g = k))}{\sum_{i=1}^{N} B(S_i^g = k)} \tag{14}$$

where S_i is the predicted state and S_i^g is the ground truth state for the ith image in Set2, and $B()$ is a Boolean logic function. The matching rates between our method and the

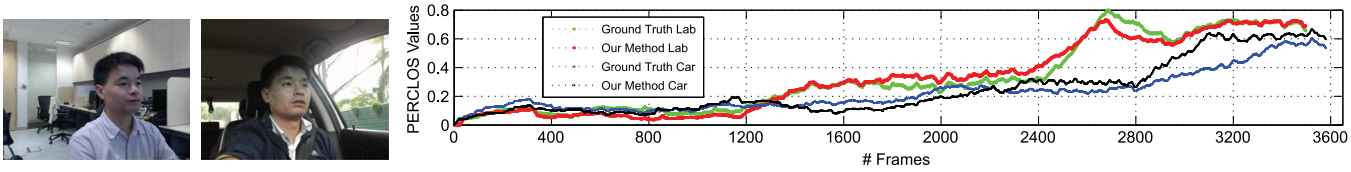


Fig. 10. The comparison of performance on videos in the laboratory and in the car. (Left) Test images of the same person in the laboratory and in the car. (Right) PERCLOS curves on the test videos captured in the laboratory and in a moving car.

ground truth for normal and fatigue states are 85.02% and 95.18%, respectively. The matching rates of baseline method and ground truth are 63.50% and 59.52% for normal and fatigue states, respectively.

The simulated drowsiness may not be same as true drowsiness in driving. The participants were asked to perform eye opening, blinking, and eye closure with different frequencies and durations to simulate drowsiness and sleepy. From the curves in Fig. 9, one can observe that our datasets cover a wide range of eye states and movements from opening, blinking slightly and frequently, and closed eyes for a while. The statistics on Set2 indicates that, the classified driver states based on the computed PERCLOS measures using our method match those based on the PERCLOS measures on ground truths in a very high percentage, which shows that our method is effective for bus driver fatigue detection when using existing monitoring cameras in buses.

Our system detects the driver first using the head-shoulder detector, and then face detection and eye localization are performed on head regions. The head-shoulder detector is extended from pedestrian detector [56] which is robust to background and lighting changes. Later, our processing will focus on face and eye regions which are separated from background in the image. Hence, our approach is robust to background changes. To validate this, we collected 3 more videos of simulated driving in moving cars. The test images of the same person in lab and in car are shown in Fig. 10. We make a comparison of our methods performance on videos of same participant in lab and in car. As an example for visual examination, the PERCLOS curves on videos of one person in lab and car are shown in Fig. 10. For the three participants, the average distance between PERCLOS curves of our method and the ground truth on the videos in lab is 0.1232, and that on videos in car is 0.1190. It can be found that the performances are similar for the compared videos of the same person in different conditions.

C. Computational Complexity

We implemented the proposed system in C++ Visual Studio 2008 on a Windows 7. On a Dell desktop of Intel 2.80GHz CPU (single core) with 4 GB of memory, head shoulder detection takes about 20 ms, face detection takes about 49 ms, both the eye detections takes about 110 ms, fusion, PERCLOS measure estimation and fatigue state classification takes about 8 ms. Overall, the operation for a single frame takes about 0.19 second, that means our system can run at about 5 frames per second (fps). Since no tracking is required in our system, the computations on consecutive frames can be performed in parallel. Hence, using a common quad-core machine, our system can run at about 20fps. In addition, the delay of half or one second will not affect the PERCLOS based fatigue detection in real-world applications.

IV. Conclusions and Future Work

In this paper, we presented a vision-based method and system towards bus driver fatigue detection using existing dome cameras in buses. Our approach starts with the detection of head-shoulders of the figure in the image, followed by face and eye detections and eye openness estimation. Finally, a multi-model fusion scheme is designed to infer eye state and a PERCLOS measure on the continuous measure of eye openness is computed to predict driver's attention state, i.e., normal or fatigue driving state. Experimental results show that our proposed method is able to distinguish the simulated drowsy and sleepy states from the normal state of driving on the low-resolution images of faces and eyes observed from an oblique viewing angle. Hence, our system might be able to effectively monitor bus driver's attention level without extra requirement for cameras. Our approach could extend the capability and applicability of existing vision-based techniques for driver fatigue detection. In the next work, we will investigate to apply it for other vehicles like car, van, minibus and lorry for easy and cheap deployment. One more issue for future work is how to exploit the fatigue detection to improve driver safety in the drowsiness situations [34]–[37].

References

[1] J. May and C. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 12, no. 3, pp. 218–224, 2009.
[2] S. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, no. 3, pp. 173–194, 2001.
[3] E. Hitchcock and G. Matthews, "Multidimensional assessment of fatigue: A review and recommendations," in *Proc. Int. Conf. Fatigue Manage. Transp. Oper.*, Seattle, WA, USA, Sep. 2005.
[4] A. Williamson, A. Feyer, and R. Friswell, "The impact of work practices on fatigue in long distance truck drivers," *Accident Anal. Prevent.*, vol. 28, no. 6, pp. 709–719, 1996.
[5] W. Dement and M. Carskadon, "Current perspectives on daytime sleepiness: The issues," *Sleep*, vol. 5, no. S2, pp. S56–S66, 1982.
[6] L. Hartley, T. Horberry, N. Mabbott, and G. Krueger, "Review of fatigue detection and prediction technologies," Nat. Road Transp. Commiss., Melbourne, Vic., Australia, Tech. Rep., 2000.
[7] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, pp. 16 937–16 953, 2012.
[8] S. Kee, S. Tamrin, and Y. Goh, "Driving fatigue and performance among occupational drivers in simulated prolonged driving," *Global J. Health Sci.*, vol. 2, no. 1, pp. 167–177, 2010.

[9] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, and M. Fathy, "A review on driver face monitoring systems for fatigue and distraction detection," *Int. J. Adv. Sci. Technol.*, vol. 64, pp. 73–100, 2014.

[10] S. Kar, M. Bhagat, and A. Routary, "EEG signal analysis for the assessment and quantification of drivers fatigue," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 13, no. 5, pp. 297–306, 2010.

[11] C. Zhang, H. Wang, and R. Fu, "Automated detection of driver fatigue based on entropy and complexity measures," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 168–177, Feb. 2014.

[12] A. Kokonozi, E. Michail, I. Chouvarda, and N. Maglaveras, "A study of heart rate and brain system complexity and their interaction in sleep-deprived subjects," in *Proc. Conf. Comput. Cardiol.*, 2008, pp. 969–971.

[13] S. Hu and G. Zheng, "Driver drowsiness detection with eyelid related parameters by support vector machine," *Exp. Syst. Appl.*, vol. 36, pp. 7651–7658, 2009.

[14] M. Kurt, N. Sezgin, M. Akin, G. Kirbas, and M. Bayram, "The ANN-based computing of drowsy level," *Exp. Syst. Appl.*, vol. 36, pp. 2534–2542, 2009.

[15] M. Patel, S. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Exp. Syst. Appl.*, vol. 38, pp. 7235–7242, 2011.

[16] M. Akin, M. Kurt, N. Sezgin, and M. Bayram, "Estimating vigilance level by using EEG and EMG signals," *Neural Comput. Appl.*, vol. 17, pp. 227–236, 2008.

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

[18] C. Liu, S. Hosking, and M. Lenne, "Predicting driver drowsiness using vehicle measures: Recent insights and future challenges," *J. Safety Res.*, vol. 40, no. 4, pp. 239–245, Aug. 2009.

[19] P. Forsman, B. Vila, R. Short, C. Mott, and H. van Dongen, "Efficient driver drowsiness detection at moderate levels of drowsiness," *Accid. Anal. Prevent.*, vol. 50, pp. 341–350, 2013.

[20] J. Wang, S. Zhu, and Y. Gong, "Driving safety monitoring using semisupervised learning on time series data," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 728–737, Sep. 2010.

[21] B.-F. Wu, Y.-H. Chen, C.-H. Yeh, and Y.-F. Li, "Reasoning-based framework for driving safety monitoring using driving event recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1231–1241, Sep. 2013.

[22] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imaging*, vol. 8, no. 5, pp. 357–377, 2002.

[23] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1052–1068, Jul. 2004.

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

[25] R. Mbouna, S. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1462–1469, Sep. 2013.

[26] A. Dasgupta, A. George, S. Happy, and A. Routray, "A vision-based system for monitoring the loss of attention in automotive drivers," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1825–1838, Dec. 2013.

[27] E. Ohn-Bar and M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2368–2377, Dec. 2014.

[28] D. Dinges, M. Mallis, G. Maislin, and J. Powell, "Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management," Nat. Highway Traffic Safety Admin. (NHTSA), Washington, DC, USA, Tech. Rep., 1998.

[29] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Proc. ICCV Workshop HCI*, 2007, pp. 6–8.

[30] M. Flores, J. Armingol, and A. de la Escalera, "Driver drowsiness warning system using visual information for both diurnal and nocturnal illumination conditions," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 1–20, Mar. 2010.

[31] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc. Conf. Comput. Vis. Pattern Recog.*, 2003, pp. 451–458.

[32] T. V. Jan, T. Karnahl, K. Seifert, J. Hilgenstock, and R. Zobel, "Don't sleep and driver—VW's fatigue detection technology," in *Proc. Int. Tech. Conf. Enhanced Safety Veh.*, 2005, pp. 1–12.

[33] Q. Wang, J. Yang, M. Ren, and Y. Zheng, "Driver fatigue detection: A survey," in *Proc. World Congress Intell. Control Autom.*, 2006, pp. 8587–8591.

[34] M. Blanco *et al.*, "Assessment of a drowsy driver warning system for heavy vehicle drivers," Nat. Highway Traffic Safety Admin. (NHTSA), U.S. Dept. Transp., Washington, DC, USA, Tech. Rep., Final Rep. DOT-HS-811-117, 2009.

[35] L. Barr, S. Popkin, and H. Howarth, "An evaluation of emerging driver fatigue detection measures and technologies," Nat. Highway Traffic Safety Admin. (NHTSA), U.S. Dept. Transp., Washington, DC, USA, Final Rep. FMCSA-RRR-09-005, Tech. Rep., 2009.

[36] N. L. Haworth and P. Vulcan, "Testing of commercially available fatigue monitors," Monash Univ. Accid. Res. Centre, Clayton, Vic., Australia, no. 15, 1991.

[37] E. Aidman, C. Chadunow, K. Johnson, and J. Reece, "Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance," *Accid. Anal. Prevent.*, vol. 81, pp. 8–13, 2015.

[38] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[39] R. Senaratne, B. Jap, S. Lal, A. Hsu, S. Halgamuge, and P. Fischer, "Comparing two video-based techniques for driver fatigue detection: Classification versus optical flow approach," *Mach. Vis. Appl.*, vol. 22, no. 4, pp. 597–618, 2011.

[40] T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, "Passive driver gaze tracking with active appearance models," in *Proc. World Congr. Intell. Transp. Syst.*, 2004, pp. 1–12.

[41] K. Yao, W. Lin, C. Fang, J. Wang, S. Chang, and S. Chen, "Real-time vision-based driver drowsiness fatigue detection system," in *Proc. IEEE Veh. Technol.*, 2010, pp. 1–5.

[42] T. Orazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection," *Pattern Recognit.*, vol. 40, pp. 2341–2355, 2007.

[43] T. Pradhan, A. Bararia, and A. Routray, "Measurement of PERCLOS using Eigen-Eyes," in *Proc. Int. Conf. Human Comput. Interact.*, 2012, pp. 1–4.

[44] J. Zheng, G. A. Ramírez, and O. Fuentes, "Face detection in low-resolution color images," in *Image Analysis and Recognition*. New York, NY, USA: Springer-Verlag, 2010, pp. 454–463.

[45] N. Robertson and I. Reid, "Estimating gaze direction from low-resolution faces in video," in *Proc. 9th IEEE Eur. Conf. Comput. Vis.*, Graz, Austria, 2006, pp. 402–415.

[46] N. Gourier, J. Maisonnasse, D. Hall, and J. L. Crowley, "Head pose estimation on low resolution images," in *Multimodal Technologies for Perception of Humans*. New York, NY, USA: Springer-Verlag, 2007, pp. 270–280.

[47] Y. Wu and Q. Ji, "Learning the deep features for eye detection in uncontrolled conditions," in *Proc. IEEE 22nd ICPR*, 2014, pp. 455–459.

[48] M. Li, Z. Zhang, K. Huang, and T. Tan, "Rapid and robust human detection and tracking based on omega-shape features," in *Proc. IEEE 16th ICIP*, 2009, pp. 2545–2548.

[49] R. Hu, R. Wang, S. Shan, and X. Chen, "Robust head-shoulder detection using a two-stage cascade framework," in *Proc. IEEE 22nd ICPR*, 2014, pp. 2796–2801.

[50] D. Wenhui, Q. Peishu, and H. Jing, "Driver fatigue detection based on fuzzy fusion," in *Proc. Chin. Control Dec. Conf.*, Shandong, China, 2008, pp. 2640–2643.

[51] J. Batista, "A drowsiness and point of attention monitoring system for driver vigilance," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Seattle, WA, USA, Oct. 2007, pp. 702–708.

[52] M. Sigari, M. Fathy, and M. Soryani, "A driver face monitoring system for fatigue and distraction detection," *Int. J. Veh. Technol.*, vol. 2013, 2013, Art. no. 263983.

[53] W. Wierwille, "Historical perspective on slow eyelid closure: Whence PERCLOS?" in *Proc. Ocular Meas. Driver Alertness Tech. Conf.*, Herndon, VA, USA, 1999, pp. 130–143.

[54] OpenCV, *Open Source Computer Vision*, 2014. [Online]. Available: http://opencv.org/

[55] X. Yu, W. Han, L. Li, J. Shi, and G. Wang, "An eye detection and localization system for natural human and robot interaction without face detection," in *Proc. TAROS*, 2011, pp. 54–65.

[56] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE CVPR*, 2005, vol. 1, pp. 886–893.

[57] D. Cai, X. He, and J. Han, "Spectral regression: A unified subspace learning framework for content-based image retrieval," in *Proc. ACM Multi-Media*, 2007, pp. 403–412.

[58] U. Trutschel, B. Sirois, D. Sommer, M. Golz, and D. Edwards, "PERCLOS: An alertness measure of the past," in *Proc. Int. Driving Symp. Human Factors Driver Assess., Train. Veh. Des.*, 2011, pp. 172–179.

[59] M. Mallis *et al.*, "Biobehavioral responses to drowsy driving alarms and alerting stimuli," U.S. Dept. Transp., Nat. Highway, Traffic Safety Admin., Washington, DC, USA, Tech. Rep., DOT HS 809 202, 2000.
[60] R. Grace, "Drowsy driver monitor and warning system," in *Proc. Int. Driving Symp. Human Factors Driver Assess., Train. Veh. Des.*, 2001, pp. 64–69.

Bappaditya Mandal received the B.Tech. degree in electrical engineering from Indian Institute of Technology (IIT), Roorkee, India, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University (NTU), Singapore, in 2003 and 2008, respectively.

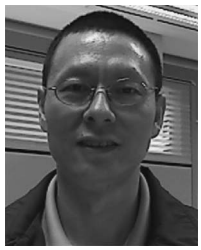
He is a Scientist with Cognitive Vision Laboratory, Visual Computing Department, Institute for Infocomm Research, Agency for Science, Technology and Research (A∗STAR), Singapore. His research interests include subspace learning, feature extraction and evaluation, computer vision, image and signal analysis, and machine learning.

Dr. Mandal was the recipient of the Best Biometric Student Paper Award at the 19th International Conference on Pattern Recognition, 2008 and the recipient of the full Research Scholarship Award from NTU, between 2004 and 2008. In 2001, he was the recipient of the Summer Undergraduate Research Award from IIT Roorkee. He is a member of the IEEE Signal Processing Society.

Liyuan Li received the B.E. and M.E. degrees from Southeast University, Nanjing, China, in 1985 and 1988, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2001.

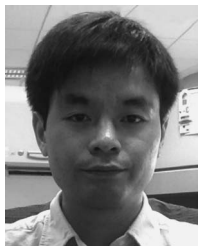
From 1988 to 1999, he was with the faculty of Southeast University, where he was an Assistant Lecturer (1988–1990), Lecturer (1990–1994), and Associate Professor (1995–1999). Since 2001, he has been a Research Scientist with the Institute for Infocomm Research, Singapore, where he is currently a Senior Scientist and the Head of Cognitive Vision Laboratory in the Visual Computing Department. His research interests include cognitive vision, egocentric vision understanding, memory and vision, mobile vision, video surveillance, human–robot interface, multiclass object detection and tracking, event and behavior understanding, image classification, etc.

Dr. Li is also a member of the IEEE Signal Processing Society and IEEE Computer Society.



Gang Sam Wang received the B.S. and M.S. degrees in computer science from Wuhan University of Technology, Wuhan, China.

Since 1997, he has had 16 years of experience of software system development in Singapore and 1.5 years in Silicon Valley, USA. Since 2011, he was a Senior Software Engineer with the Institute for Infocomm Research, Agency for Science, Technology and Research (A∗STAR), Singapore. His research interests include system design and implementation, video analysis, multitask learning for multiclass and multiview object detection, and parallel programming.



Jie Lin received the Ph.D. degree from the School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China, in 2014.

He was a Research Engineer with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. He is currently a Research Scientist with the Institute of Infocomm Research, Singapore. His research interests include mobile visual search, large-scale image/video retrieval, and deep learning. His work on low-complexity feature compression has been adopted as the core contribution to the MPEG-7 Compact Descriptors for Visual Search standard.