Abnormal Driving Behavior Detection Using Sparse Representation

Chien-Yu Chiou, Pau-Choo Chung
Dept. of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan
puchung@ee.ncku.edu.tw

Chun-Rong Huang
Dept. of Computer Science and Engineering
National Chung Hsing University
Taichung, Taiwan
crhuang@nchu.edu.tw

Ming-Fang Chang
Automotive Research and Testing Center
Changhua, Taiwan
neil_cmf@artc.org.tw

Abstract—To reduce the chance of traffic crashes, many driver monitoring systems (DMSs) have been developed. A DMS warns the driver under abnormal driving conditions. However, traditional approaches require enumerating abnormal driving conditions. In this paper, we propose a novel DMS, which models the driver’s normal driving statuses based on sparse reconstruction. The proposed DMS compares the driver’s statuses with his/her personal normal driving status model and identifies abnormal driving statuses that greatly change the driver’s appearances. The experimental results show good performance of the proposed DMS to detect variant abnormal driver conditions.

Keywords— Driver monitoring system, sparse reconstruction, normal model, anomaly detection

I. INTRODUCTION

Recent statistical reports from National Highway Traffic Safety Administration (NHTSA) of U.S.A. [1] show that more than 5 million traffic accidents occur annually in U.S.A. including about 30 thousand fatal crashes. The reports also show that more than half of the crashes are caused by drivers. Thus, driver monitoring systems (DMSs) are proposed to monitor driver’s driving statuses. When a dangerous driving behavior is detected, DMSs will alarm the driver to avoid possible accidents.

The DMSs can be separated to three kinds of systems, including visual based, physiology signal based, and mechanics based DMSs. Visual based DMSs monitor the driver’s appearance with cameras set in the car. The driver’s statuses are estimated from the changes of the driver’s facial regions. Recently, DMSs focus on detecting drowsy driving and distracted driving, which are two well-known causes of fatal crashes. The NHTSA report [2] shows that the percentage of eye closure time is related to drowsiness. To detect driver’s drowsy driving behavior, many methods search the driver’s eye region and detect eye closure. For example, Bergasa et al. [3] detect the eye regions with the reflection of the pupils of eyes under near infrared light. The opened and closed eye states are then estimated with the shapes of the pupils. Jo et al. [4] combine adaptive template matching and blob detection to detect and track the driver’s eyes and use a support vector machine (SVM) to estimate the eye states. Jayanthi and Bommy [5] detect the driver’s eyes based on colors and eye blinks with template matching. Mbouna et al. [6] consider both the eyelid height and closed eyes to identify the level of eye closure. These methods achieve decent performance on detecting the driver’s drowsiness.

Nevertheless, the eye closure and the behavior of the drowsy driving are variant for different drivers.

Besides the drowsy driving, the distracted driving is also a common abnormal driving behavior. Similar to the behavior of the drowsy driving, the behavior of the distracted driving for different drivers is variant. To reduce the difficulty of detecting the distracted driving, the distracted driving is usually defined as the behavior that the driver is not looking at the front of the vehicle. Estimating eye gazes or head poses provides a naïve evidence to detect distracted driving. To estimate the head poses, some methods directly use features from 2-D images of the driver’s face. For example, Murphy-Chutorian et al. [7] adapt localized gradient orientation histograms with support vector regression to estimate the driver’s head orientations. Fu et al. [8] use principal component analysis and Gabor wavelet transformation with particle filters to estimate the head poses.

Compared to 2-D based approaches, 3-D based methods reconstruct a 3-D face model to estimate head poses. In [9], an ellipsoidal 3-D head model is reconstructed based on the shapes of the driver’s face images. Similar to [9], the ellipsoidal 3-D head model is also used in [10]. Instead of using shapes of the driver’s face, [11] and [12] apply key facial landmarks to 3-D face models and estimate head poses.

Compared to visual based methods, physiology signal based methods monitor the physiological activities of the driver. Some methods apply heart rate variability (HRV) based on electrocardiogram (ECG) or photoplethysmogram (PPG) signals to monitor the driver’s heart beats and detect drowsiness [13][14][15]. Some methods measure the driver’s brain activities with electroencephalograph (EEG) signals to detect drowsiness [16] or distraction [17]. Although physiology signals directly reflect the driver’s body conditions, additional sensors and devices are required to gather the signals. Wearing the devices may make drivers feel uncomfortable and interfere their driving. Moreover, these signals are very sensitive to noise caused by the vibrations of vehicles and the driver’s movement.

Different from visual based and physiology signal based methods, mechanics based methods measure the driver’s behavior from the signals of the vehicle. These signals include the variation of vehicle velocity, the acceleration, the distance to lane lines, the angle of the steering wheel, and specific vehicle operation patterns to detect abnormal driving conditions [18][19][20]. However, vehicle signals are often affected by the road and traffic conditions and cannot truly reflect the abnormal driving behaviors.
In summary, recent DMSs focus on detecting drowsy driving and distracted driving. They define these two specific abnormal driving behaviors, collect training data, and build the corresponding abnormal driving behavior detectors. Although DMSs aim to reduce the chance of traffic accidents, recent systems are hard to identify various abnormal or dangerous driving behaviors. Moreover, different drivers have different appearances and show various driving behaviors even under similar situations. Thus, it is hard to enumerate and collect sufficient training data for all possible conditions. As a result, traditional approaches are not very effective to monitor the driver’s driving statuses and detect abnormal driving conditions. New driver monitoring systems that can resolve the aforementioned problems still need to be developed.

Although abnormal driving behaviors vary with respect to drivers, driver’s appearances under abnormal driving conditions will be different from the normal one. Therefore, different from traditional systems, we propose an innovative DMS that models the appearances of a driver’s eyes regions under normal driving conditions. For each driver, a personal normal driving model is built from the driver’s videos during normal driving. By comparing the driver’s current driving status to the normal model, the proposed system can identify abnormal driving conditions.

The paper is organized as follows. Section 2 describes the driver’s normal driving model and the descriptor, which is used to represent the driver’s statuses. The experimental results in Section 3 show the performance of the normal driving model. Finally, Section 4 gives the conclusion and the future work.

II. METHOD

A. Overview

Abnormal driving conditions have a large variety. It is not only difficult to enumerate and define all possible conditions, but also hard to collect training data. Moreover, detection of multiple abnormal conditions needs to build detection models for each anomaly and compares with each model separately, which increases the computational load. Thus, we build a personal normal driving status model of a driver from the driver’s normal driving videos. By comparing to the driver’s personal normal driving status model, we can identify those abnormal driving statuses that greatly change the driver’s appearance without pre-defining abnormal conditions.

Figure 1 shows the flow chart of the proposed method. The driver’s video is captured with a camera set in front of the driver as shown in Figure 2. Descriptors are then computed with the image of the driver’s eyes regions to represent the driver’s statuses. During training, descriptors computed from normal driving statuses are first used to build the driver’s personal normal driving status models. The normal driving status models encode the descriptors by using sparse representation [21][22]. After building the normal models, the system compares the descriptors of the driver’s driving statuses with the normal models. When the descriptors are similar to the models, the system decides that the driver is under normal driving statuses. If the descriptors differ from the models, the driver may be under abnormal driving statuses, and the system will send a notice to the driver. With the proposed normal driving status models, the system is able to detect various anomalies, such as drowsy, talking on phone or eating, without predefining these dangerous driving behaviors and collecting training data of such behaviors.

B. Face/Eye Detection, and Driver’s Status Description

A face detector [23] and a facial landmarks regression tree [24] are applied on each frame \( t \) captured from the camera to detect the face region and estimate locations of facial landmarks. Then, we extract the left eye region \( FR_{left}(t) \) and right eye region \( FR_{right}(t) \) around the corresponding facial landmarks to represent the driver’s status in frame \( t \).

During driving, the environment lighting may change rapidly. The descriptor should not only be able to represent the driver’s appearances but also be invariant to illumination changes. To reduce the lighting effects, we use contrast context histogram (CCH) descriptor [25] to describe the driver’s eye regions. CCH is a contrast-based descriptor, which is known to reduce the effects of light variations.

Given a region \( R \), CCH divides \( R \) into four sub-regions \( R_i \), where \( i \in \{1, 2, 3, 4\} \), and then computes the histogram of contrast values between all pixels \( p \) in \( R \) and the center pixel \( p_c \) of \( R \), as shown in Figure 3. The intensity \( I(p) \) of \( p \) is obtained by averaging the intensities of the four center pixels of \( R \), which are marked by orange in Figure 3. Let the intensity contrast between a pixel \( p \) and the center point \( p_c \) be \( \delta(p, p_c) = I(p) - I(p_c) \), where \( I(p) \) is the intensity value of \( p \). Then, we can compute the intensity contrast histogram of a sub-region and the center point. To increase the discriminability of descriptors, the histogram is computed with the image of the driver’s eyes regions to represent the driver’s statuses. During training, descriptors computed from normal driving statuses are first used to build the driver’s personal normal driving status models. The normal driving status models encode the descriptors by using sparse representation [21][22]. After building the normal models, the system compares the descriptors of the driver’s driving statuses with the normal models. When the descriptors are similar to the models, the system decides that the driver is under normal driving statuses. If the descriptors differ from the models, the driver may be under abnormal driving statuses, and the system will send a notice to the driver. With the proposed normal driving status models, the system is able to detect various anomalies, such as drowsy, talking on phone or eating, without predefining these dangerous driving behaviors and collecting training data of such behaviors.
computed in a positive bin and a negative bin. The positive histogram bin value $H^+(R_i)$ of $R_i$ is defined as follows:

$$H^+(R_i) = \frac{\sum \{ \delta(p, p_c) | p \in R_i \text{ and } \delta(p, p_c) \geq 0 \}}{\#^+ R_i},$$

(1)

where $\#^+ R_i$ is the number of pixels in $R_i$ that has higher intensity value than $p_c$. Similarly, the negative bin value $H^-(R_i)$ of $R_i$ is defined as follows:

$$H^-(R_i) = \frac{\sum \{ \delta(p, p_c) | p \in R_i \text{ and } \delta(p, p_c) < 0 \}}{\#^- R_i},$$

(2)

where $\#^- R_i$ is the number of pixels in $R_i$ that has lower intensity value than $p_c$.

Then, the contrast histogram $H(R)$ of region $R$ is obtained by connecting histogram bins of every sub-region $R_i$ as follows:

$$H(R) = \{H^+(R_1), H^-(R_1), H^+(R_2), H^-(R_2),$$

$$H^+(R_3), H^-(R_3), H^+(R_4), H^-(R_4)\}.$$  

(3)

The dimension of $H(R)$ is 8 (2 for positive and negative values, 4 for quartered sub-regions). To reduce the effect of lighting changes and solve the scale problem of values, the histogram vector is normalized to a unit length vector. In this way, we can efficiently describe the appearance changes of the driver’s face with a rather low dimensional descriptor.

C. Personalized Driver’s Status Estimation

With the descriptors of the driver’s eyes, the driver’s personalized normal driving status model can be built. By comparing with the driver’s normal driving status model, we can estimate if the driver’s driving status is normal or abnormal. In addition, to identify abnormal conditions and remind the driver in time, the comparison should be done in real-time.

We consider building normal driving status models using sparse reconstruction of descriptors of the driver’s normal driving statuses. Sparse reconstruction can represent vectors using a few atoms in the dictionary. If the vector belongs to the spans of the dictionary, the reconstruction error will be small. In other words, if the driver’s status is similar to the normal model built by the sparse representation, the distance between the descriptor and the reconstructed vector will be small. Therefore, we can use sparse reconstruction to compare the driver’s status with the normal driving status model. We build the model directly with the normal driving statuses descriptors.

Let $k \in \{\text{left eye, right eye}\}$ indicate the regions of the driver’s eyes, and $H(R_k(t))$ be the descriptor of eye $k$. We compose a dictionary $D_k$ for each eye region $k$ with $N_D$ descriptors as follows

$$D_k = \{H(R_k(t_1)), \ldots, H(R_k(t_{N_D}))\},$$

where $t_1$ is the frame index used to build the normal model and $N_D$ is the number of descriptors used in $D_k$.

After we build the dictionaries for each eye, we compare the driver’s driving status with these normal driving status models by sparse reconstruction [21][22]. The dictionary of respective region is used to reconstruct the descriptor, by solving the following equation:

$$\text{Err}(H(R_k(t)), D_k) = \min_{\alpha_1(t)} \|H(R_k(t)) - D_k \cdot \alpha_1(t)\|_2 \text{ s.t. } \alpha_1(t) \geq 0$$

where $\alpha_1(t)$ is the reconstruction coefficient. Since we directly use the original descriptors to form the dictionary, the reconstructed descriptor should be a linear combination with non-negative coefficient. Thus, we constrain $\alpha_1(t)$ to be non-negative. If the driver’s driving status is similar to the normal dictionary, the descriptor should be close to the span of the dictionary. The reconstruction error $\|H(R_k(t)) - D_k \cdot \alpha_1(t)\|_2$ should be small. When the reconstruction error is large, it implies that the driver’s face is different from his normal faces, i.e. abnormal driving behavior occurs. If the reconstruction error of any region $i$ in frame $t$ is larger than a threshold, we mark frame $t$ as an abnormal frame.

III. Experiments

In this section, we evaluate the performance of the proposed method. However, there are no state-of-the-art DMS methods that can detect unspecific abnormal driving behaviors for comparison. For comparison, we use [21] to build generic models for all of the testing videos. The proposed method is implemented in Visual C++ 2013, open source libraries OpenCV 2.4.9 [26] and Dlib 18.14 [27] for image processing, and SPAMS [21][22] to solve the sparse reconstruction problem. The experiments are run on an Intel i7 3.4 GHz computer with 16 GB RAM of Windows 7 operating system.

A. Datasets

To evaluate the performance of the proposed method, we collect a driving video dataset. Drivers simulated different kinds of abnormal driving conditions. The dataset contains 20 subjects recorded in car. For each subject, two videos are captured. One video is the subject looking toward the camera without face expression, which is used to build the personal normal driving status model of that subject. The other video contains different kinds of abnormal driving conditions of the subject, such as eating, laughing, and sleeping, etc. Between two different abnormal driving conditions, the subject is required to recover to the normal status for a short period to separate the two events.
TABLE 1. PERFORMANCE OF ABNORMAL DRIVING CONDITION DETECTION

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>0.713</td>
<td>0.800</td>
<td>0.799</td>
</tr>
<tr>
<td>Generic model [21]</td>
<td>0.539</td>
<td>0.456</td>
<td>0.586</td>
</tr>
</tbody>
</table>

TABLE 2. COMPUTATION TIME OF MODEL BUILDING

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Frame #</th>
<th>Time(seconds)</th>
<th>FPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face/Eye Detection</td>
<td>11,478</td>
<td>110.78</td>
<td>103.61</td>
</tr>
<tr>
<td>Driver’s Status Descriptor</td>
<td></td>
<td>5.54</td>
<td>2072.22</td>
</tr>
<tr>
<td>Build Model</td>
<td></td>
<td>2.55</td>
<td>4510.02</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>118.86</td>
<td>96.56</td>
</tr>
</tbody>
</table>

TABLE 3. COMPUTATION TIME OF MONITORING

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Frame#</th>
<th>Time(second)</th>
<th>FPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face/Eye Detection</td>
<td>61,296</td>
<td>684.28</td>
<td>89.58</td>
</tr>
<tr>
<td>Driver’s Status Descriptor</td>
<td></td>
<td>64.66</td>
<td>947.99</td>
</tr>
<tr>
<td>Model Comparison</td>
<td></td>
<td>83.15</td>
<td>737.19</td>
</tr>
<tr>
<td>Decision</td>
<td></td>
<td>11.37</td>
<td>5393.40</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>843.45</td>
<td>72.67</td>
</tr>
</tbody>
</table>

B. Abnormal Driving Condition Detection

The recall, precision and F-measure scores of the proposed method and generic models using [21] are shown in Table 1. The proposed method can detect large changes of the driver’s appearances when the driver is under abnormal driving conditions. Therefore, the proposed method can effectively detect unspecified abnormal driving conditions. Also shown in Table 1, generic models [21] achieve worse results. The possible reason is that there are large variations of the appearances and behaviors among different drivers. Using generic model may not cover all of possible conditions. Therefore, the performance of abnormal driving condition detection with generic model is worse than that of the proposed model.

Figure 4 and Figure 5 show the results of the proposed method and generic model method. When the driver closes eyes or looks aside, his/her appearances at eye regions will change. Thus, the proposed normal model can detect the differences. Because the generic model also models the eyes, detection of the closing eyes and eye gaze direction can also be achieved for the generic model method. However, in the last column of Figure 4 and Figure 5, the driver takes down the glasses, which is a very dangerous conditions during driving. The proposed method can successfully identify the appearance changes of eyes and detect such an abnormal event. The generic model fails to detect such conditions because the training contains drivers without wearing glasses. Such results show that the personal normal driving model is more suitable for every driver compared to a pre-built generic model.

The second column of Figure 6 shows the failure results of our method and generic model method. The driver is talking on phone. Nevertheless, the driver is looking forward with eyes opened and the driver’s expression does not change. Therefore, both methods are unable to detect the event. Such conditions may be solved by including more regions of the driver’s face into the personal model. In the third column of Figure 6, a cigarette partially occludes the driver’s eyes region. The proposed personal model can detect the change, while the generic model fails.

Figure 7 shows the more results. False alarms raise when the driver is under normal driving conditions for the generic model method. When the driver falls asleep, the generic model method still considers that the driver is under normal conditions. The reason may be that the driver’s appearance is different from the drivers of the generic model. Therefore, the generic model cannot effectively estimate the driver’s driving statuses. The results again show that the proposed personal normal driving model outperforms traditional detection methods that using generic model.

Table 2 and Table 3 show the computation time of the proposed method to build personal normal driving status model for a driver and monitor the driver’s driving status, respectively. The average FPS is around 96 for building the model, and the average FPS for estimating the driver’s status is around 72. The proposed method achieves the real-time performance (FPS > 30) in both training and testing stages. The experimental results show that the proposed method can effectively and efficiently detect unspecified abnormal driving conditions.

IV. CONCLUSION AND FUTURE WORKS

In this paper, we propose a novel DMS to monitor the driver’s driving statuses. Instead of enumerating and modeling different abnormal driving conditions, we propose to model the normal driving statuses. With the personal normal driving status model, our method can successfully detect undefined abnormal driving conditions that change the driver’s appearance greatly. Compared to recent DMSs, the proposed normal driving status model provides a novel concept to the DMS development. It shows that instead of building the abnormal model, it can be more efficient and effective to build the normal driving models to identify the driver’s driving statuses.

Currently, the proposed method makes decision independently in every frame. However, the driver’s motions are continuous in the video. Temporal information may help to better represent the driver’s driving status. In the future, we will extend our work to propagate the temporal information of the driver for the system to make more accurate decisions. We will also improve the face tracking algorithm to make the face detection and face part detection become more stable and time coherent.

ACKNOWLEDGMENT

This work was supported in part by Ministry of Science and Technology, Taiwan under MOST 104-2221-E-005-027-MY3, 103-2923-E006-001-MY3 and Ministry of Education, Taiwan for the Top University Project to the National Cheng Kung University.
REFERENCES


