Detecting Drivers' Mirror-Checking Actions and its Application to Maneuver and Secondary Task Recognition

Nanxiang Li, *Student Member,* Carlos Busso, *Senior Member, IEEE,*

Abstract—This study explores the feasibility of detecting drivers' mirror-checking actions using non-invasive sensors. Checking the mirrors is an important primary driving action that allows drivers to maintain their situational awareness, especially when they are planing to turn or change lanes. Recognizing when drivers are checking the mirrors can facilitate the detection of hazard scenarios by considering contextual information (e.g., turning without checking mirrors, lack of mirror-checking actions signaling cognitive distractions, or distinction between gazes due to primary or secondary tasks). This study analyzes drivers' mirror-checking actions under various real driving conditions. We analyze the drivers' mirror-checking actions under normal conditions, and when the drivers are engaged in secondary tasks such as tuning the radio or operating a cellphone. We also compare mirror-checking behaviors observed during different maneuver actions: driving straight, turning and switching lanes. The study reveals statistically significant differences in mirror-checking actions among most of the comparisons. The results suggest that mirror-checking actions can be useful indicators in recognizing drivers engaged in secondary tasks, and in detecting driving maneuvers. We propose to detect mirror-checking actions using features extracted from multiple noninvasive sensors (CAN-Bus and cameras facing the driver and the road). We consider three machine learning algorithms for unbalanced datasets, achieving F-score of 91%. The recognized mirror-checking actions are used as additional features to improve the performance of secondary task detection and maneuver recognition. These promising results suggest that it is possible to detect mirror-checking actions, providing contextual information to improve new driver monitoring systems.

Index Terms—Driver distraction, driving maneuver, mirror-checking detection, binary classification, imbalanced classification.

1 Introduction

DRIVERS rely on frequent peripheral visual scanning and mirror-checking actions to keep their situation awareness. However, under high mental workload, secondary tasks, or the influences of fatigue, drivers tend to reduce the range of peripheral scanning, neglecting mirror-checking actions. As a result, they are detached from the surrounding traffic, increasing the risk of accidents [1], [2]. Therefore, an *advanced driver assistance system* (ADAS) should be aware of the drivers' mirror-checking behaviors, which can be valuable in assessing the attention level of the drivers while performing primary (e.g., turning, switching lane) and secondary (e.g., tuning radio, using a cellphone) tasks. We propose that drivers' mirror-checking action is a unique sub-driving task that can be automatically detected using multimodal features extracted from noninvasive sensors. The automatically detected mirror-checking actions can be used to estimate when the driving behaviors deviates from normal patterns.

There is a close relationship between mirror-checking actions and drivers' behaviors. Primary driving tasks such as visual scanning, turning and switching lanes require mirror-checking actions [3]–[6]. Secondary tasks such operating a mobile device and tuning the radio also affect mirror-checking actions [7]. As the cognitive load increases, drivers reduce the frequency and duration in mirror-checking actions affecting their situational awareness [8], [9]. Therefore, understanding mirror-checking actions in the context of ADAS can have a number of benefits. An ADAS is designed to increase drivers' situation awareness by providing additional sensory information. For example, a *lane departure warning* (LDW) should not warn the driver if he/she wants to execute a lane change. A driver monitoring system should distinguish between glances related to primary or secondary driving tasks. This study focuses on detecting mirror-checking actions, and its role in recognizing both drivers engaged in secondary tasks, and vehicle maneuvers.

After manually annotating mirror-checking actions, we evaluate the deviation from normal behaviors observed when drivers are engaged in secondary tasks. The analysis relies on real driving recordings using the UTDrive platform [10]. We observe clear differences indicating that secondary tasks reduce frequency and duration of mirror-checking behaviors, affecting the driver's situational awareness. We also compare the mirror-checking behaviors when the drivers were turning, driving straight and switching lanes. We find significant differences across these maneuver actions. Based on these findings, we propose to detect mirror-checking actions using noninvasive sensors, achieving an F-score of 0.91 using multimodal features extracted from

- *N. Li and C. Busso are with the Department of Electrical Engineering, University of Texas at Dallas, Dallas, TX, 75080.*
E-mail: nxl056000@utdallas.edu, Busso@utdallas.edu

the *controller area network* (CAN)-bus signal and video cameras facing the driver and the road.

To demonstrate the benefits of detecting mirror-checking actions, we use the detected actions as features in two separate problems: identify drivers engaged in secondary tasks; and recognize driving maneuver. In both cases, the classification results improve when we use the mirror-checking actions. The findings from this study demonstrates that it is feasible to detect drivers' mirror-checking behaviors and use this information to improve driving behavior modeling. It illustrates the benefit of leveraging high level semantic features, containing important information about the driver behaviors, complementing low level features. This study opens new directions in the field of monitoring driving behaviors to improve the next generation of ADASs.

The paper is organized as follows. Section 2 summarizes the finding from studies related to this work. Section 3 describes the multimodal corpus recorded in real-world driving conditions, the feature extraction procedure, and the manual annotation of mirror-checking actions in the database. Section 4 analyzes drivers' mirror-checking behaviors under various driving conditions, including primary and secondary driving tasks. Section 5 presents our approach to detect the drivers' mirror-checking actions using multimodal features. Section 6 demonstrates the benefits of using mirror-checking action in maneuver recognition and secondary task detection. Section 7 concludes the paper with final remarks and our future research directions.

2 Related Work

Studies on driving maneuver indicated that mirror-checking behavior is closely related to actions such as turning and switching lanes [4]–[6]. Georgeon et al. [4] found that lane switch maneuver can be detected from the changes in the vehicle speed and the drivers' left mirror-checking actions. Olsen [5] suggested that identifying glances to the left outside mirror provides earlier prediction of left turn and lane switch. Itoh et al. [6] used increases in mirror-checking frequency to estimate intention of truck drivers to change lanes.

Studies have reported the close relationship between mirror-checking behaviors and different driving conditions [3], [7], [11]. In 1972, Robinson et al. [11] suggested that mirror-checking behaviors can indicate drivers' visual search workload (busy traffic versus no traffic). This observation becomes more relevant nowadays due to the various in-vehicle infotainment systems and handheld devices used by modern drivers. Several studies on driver distraction have mentioned the changes in mirror-checking behaviors under different driving conditions [1], [2], [7]–[9], [12]–[15]. Harbluk et al. [8], [9] observed that drivers are likely to reduce the mirror-checking frequency and duration under fatigue or high cognitive workload. Brookhuis and de Waard [2] suggested that this behavioral change is produced by drivers allocating workload resources to complete tasks while driving. As the task demand increases, drivers reduce the rearview mirror-checking frequency.

In addition to indication of high mental workload and driving maneuvers, other related studies have used mirror-checking behaviors for different purposes. Murray et al. [16] studied the driving patterns of people driving either for work or personal purposes. They concluded that people who drive for work purposes engage in less safe driving practices. One of the factors considered in the analysis was to overtake without checking the mirrors. Lei et al. [17] used the *driver behavior questionnaire* (DBQ) to investigate the skill and style of drivers. The self-reported survey explicitly includes cases when the driver failed to check his/her rear-view mirror before conducting maneuver such as changing lanes. Underwood et al. [18] used mirror-checking behaviors to evaluate the differences in skills between novice and experienced drivers on different road types. They found that novice drivers checked the external door mirror less often than experienced drivers in lane change tasks while driving in dual-carriageways.

Despite the clear relationship between mirror-checking actions and driver behaviors, current ADASs do not explicitly consider this primary driving task. The studies mentioned in this section extract mirror-checking actions using manual annotations, or self-reported questionnaires. To the best knowledge of the authors, the automatic detection of mirror-checking actions has not been explored. Furthermore, related studies have not considered mirror-checking actions as features in driving classification problems (normal versus tasks, maneuver recognition). This study takes up the challenge of detecting mirror-checking actions using noninvasive sensors. The detected actions are used as contextual features in maneuver and secondary task recognitions.

3 Data collection and Preprocessing

This study relies on recording collected in real-world driving conditions using the UTDrive platform. The corpus provides realistic driving behaviors, preserving the computer vision and signal processing challenges existing in real world applications (i.e., vibrations, noise, adverse illumination, complicated road dynamics, behaviors from other vehicles).

3.1 Data Collection

We collected a multimodal database from people driving the UTDrive platform in real roads (see Fig. 1). The details of the corpus are given in Li et al. [19]. The UTDriver car is a 2006 Toyota RAV 4 equipped with multiple sensors: a microphone array, a camera facing the road, a camera facing the driver, and two pressure sensors on the gas and brake pedals [10]. We extract multiple information describing the vehicle performance from the CAN-Bus signal (e.g., speed and steering wheel). The UTDrive offers a perfect platform to simultaneously

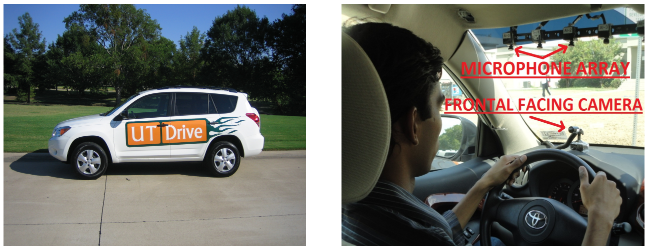


Fig. 1. UTDrive car and placement of the sensors.

study the vehicle dynamic, environment condition and driver performance. We have used this corpus to study the driver' visual and cognitive distractions [19]–[22].

As shown in Figure 2, we define a 5.6-mile route comprising both city and residential roads with various traffic signs (e.g., traffic light, stop signs). Twenty students/employees (10 male and 10 female) from the University of Texas at Dallas participated in the recording. The average and standard deviation of the participants' age are 25.4 and 7.03, respectively. Each participant drove along the predefined route twice. During the first lap, the drivers operated the vehicle while performing different secondary tasks during specific segments in the route. The colors in Figure 2 indicates the route segments corresponding to each secondary tasks. The description of these tasks is given below.

- Operating the in-built car radio (Fig. 2, red route): The driver is asked to tune the radio to the preselected stations.
- Operating and following instructions from the GPS (Fig. 2, green route): A pre-defined address is given to the driver, who is asked to enter it into the GPS. Then, they are asked to follow the GPS instructions to reach the destination. This task is subdivided into *GPS - Operating* and *GPS - Following* (preliminary results suggested that driver behaviors are different for these two activities [23]).
- Operating and talking on the phone (Fig. 2, navy blue route): The driver is asked to call an airline automatic flight information system (toll-free) to retrieve flight information between two given US cities, using a cellphone. This task is also subdivided into *Phone - Operating* and *Phone - Talking* for similar reasons as above.
- Describing pictures (Fig. 2, orange route): This task requires the driver to look and describe randomly selected pictures which are held out by a passenger seated beside the driver. The purpose of this task is to simulate the task of looking at objects outside the car, such as billboards, sign boards and shops.
- Conversation with a passenger (Fig. 2, black route): The last task is a spontaneous conversation between the driver and a second passenger in the car.

Fig. 2. The subjects drove this route twice (5.6 miles): first lap, performing a series of secondary tasks: *Radio* (red), *GPS-Operating* and *GPS-Following* (green), *Phone-Operating* and *Phone-Talking* (navy blue), *Picture* (orange) and *Conversation* (black) (Table 1); second lap, driving normally without getting involved in any other task.

TABLE 1
Secondary tasks in the corpus. The tasks phone and GPS were split, since we observed different driver behaviors while operating and using the devices.

Tasks	Duration [s]		Mean Speed [km/hr]	
	Normal	Task	Normal	Task
Radio	1748	1635	50.56	54.04
GPS-Operating	1113	975	48.30	55.46
GPS-Following	3286	3050	35.13	37.09
Phone-Operating	476	478	51.76	53.46
Phone-Talking	2403	2546	40.76	37.81
Picture	1564	1573	51.79	52.34
Conversation	1648	1618	43.75	45.37

The secondary tasks, approved by our *institutional review board* (IRB), are chosen from common activities that drivers daily engage while driving. They span broad range of activities with different level of difficulty, inducing different types of distractions (visual, manual and cognitive). We do not consider activities such as texting, for safety reasons. For the second lap, the drivers drove the same route without performing any secondary tasks. This protocol to record the database follows the experience from other research groups [10], [24]–[26], where we fix the order of the tasks over predefined route segments. Therefore, we can collect a reliable baseline for normal driving behavior, in which most of the other variables are kept fixed (e.g., route, traffic, and street signals). The differences in behaviors observed for normal and secondary task conditions can be primarily associated with the effect induced by secondary tasks – not the route.

Table 1 lists the average duration and speed for each task route segment for both normal and task driving conditions. All experiments are conducted under good weather conditions, with a short break between the two laps. The entire corpus comprises approximately 12 hours of multimodal data.

TABLE 2
Multimodal Features. (*CAN* = CAN-Bus signal; *MI* = microphone; *RC* = road camera; *DC* = driver camera)

	Multimodal Features	
CAN	Vehicle Speed (Speed)	Brake Pressure (Brake)
	Steering Wheel Angle (Steering)	Acceleration (Acceleration)
	Steering Wheel Jitter (Jitter)	Gas Pedal Pressure
	Brake Pedal Pressure	
MI	Energy	
RC	Road Optical Flow	
	Road Intensity	
DC	Head Yaw Angle (Yaw)	Chin Raiser (AU17)
	Head Pitch Angle (Pitch)	Lip Stretcher (AU20)
	Head Roll Angle (Roll)	Cheek Raiser (AU6)
	Inner Brow Raiser (AU1)	Lip Tightener (AU7)
	Outer Brow Raise (AU2)	Lip Puckerer (AU18)
	Brow Lowerer (AU4)	Lip Tightener (AU23)
	Upper Lid Raiser (AU5)	Lip Pressor (AU24)
	Nose Wrinkler (AU9)	Lips Part (AU25)
	Upper Lip Raiser (AU10)	Jaw Drop (AU26)
	Lip Corner Puller (AU12)	Lip Suck (AU28)
	Dimpler (AU14)	Eye Openness (AU45)
	Lip Corner Depressor (AU15)	
	Statistics Derived for each Analysis Window	
	Mean, standard deviation, maximum, minimum, range, inter-quatile range, skewness and kurtosis	
	Gaze features derived from the driver's camera	
	Eyes-Off-the-Road Duration (EOR Dur.)	
	Eyes-Off-the-Road Frequency (EOR Freq.)	
	Longest Eyes-Off-the-Road Duration (LEOR Dur.)	
	Eye Blink Frequency (Blink Freq.)	

3.2 Feature Extraction from Noninvasive Sensors

This study considers features extracted from the CAN-Bus signal, pedal sensors, an array of five microphones, and two cameras, one facing the driver and another facing the road. The signals from the CAN-Bus can be directly used to describe driving behaviors. Table 2 lists the signals extracted from the CAN-Bus and pedal sensors. The feature *Brake pressure (Brake)* is extracted from the CAN-Bus and it measures the actual pressure from the brake sensor in the vehicle. The feature *Brake Pedal Pressure* is extracted from the brake pedal sensor, capturing activity even when the driver steps on the pedal without pressing it. The other modalities require preprocessing steps, which are described next.

We estimate the energy from the central microphone of the array. This feature is useful to characterize driver behaviors involving speech or background music (e.g., *Conversation*, *Radio* – see Table 1). We rely on the *computer expression recognition toolbox* (CERT) [27] to extract head and facial features from the video facing the drivers. CERT estimates frame-by-frame both the head pose and facial *action units* (AUs). AUs are part of the *facial action coding system* (FACS) and describe facial muscle activity [28]. Although the main purpose for CERT is on expression recognition, it has been used for driver drowsy detection [29]. Table 2 lists the AUs and head pose estimation angles extracted from CERT. The toolkit occasionally fails to detect the driver face due to adverse illumination conditions and extreme head movement. For these cases, we interpolate the values of head pose and AUs if more than two third of the frames from the analysis window are available. Otherwise, we do not use facial features from these segments.

For the video facing the road, we calculated two quantities in each frame to represent the road dynamics: optical flow energy and image intensity. The former feature characterizes the traffic condition (i.e., the overall objects movement). It is estimated by summing the magnitude of the optical flow for each frame. The latter feature characterizes the driving condition (i.e., illumination, road condition). It is estimated by summing the raw pixel intensity values in each frame. Despite the simplicity of these features, our previous study demonstrated their effectiveness in detecting driver distractions [22].

In our previous study, we have successfully modeled driver behaviors using statistics within a moving window from the features listed on Table 2 [19], [21], [22]. The same method is adopted here (see Sections 5 and 6). We estimate eight statistics for each signal: mean, standard deviation, maximum, minimum, range, inter-quartile range, skewness and kurtosis. The window size in the study is set to five seconds, as a trade-off between lag in driver behavior detection and reliable statistic estimation. The lowest sampling rate among all signals is the vehicle speed, which is 2Hz. Therefore, we can have at least 10 samples per window to estimate the statistics. Also, mirror-checking behavior is usually completed during short period, usually within two seconds, occurring at relatively long intervals (20 - 40 seconds) [30]. Therefore, the window size should be short enough to capture the characteristic behaviors associate with mirror-checking actions. The promising mirror-checking detection results described in Section 5 suggest that a five-second window is appropriate for detecting this behavior.

To capture the drivers' eye movement, we derive four additional gaze related features estimated over the analysis window using the head pose and eye openness estimation from CERT. They are eye-off-the-road duration (*EOR Dur.*), eye-off-the-road frequency (*EOR Freq.*), the longest eye-off-the-road duration (*LEOR Dur.*) and eye blink frequency (*Blink Freq.*) (see Table 2). The details of these four features are given in Li and Busso [22].

Overall, we extract a 268D feature vector for each five second window (33 signals $\times$ 8 statistics + 4 gaze features). To improve time resolution, we shift the analysis window every one second producing overlapped windows (i.e., four seconds overlap). Therefore, we have a feature vector every one second for the entire recording, excepting the first four seconds of the recordings.

3.3 Mirror-Checking Behavior Annotation

We manually annotate mirror-checking actions for the entire database. Two external observers were asked to annotate mirror-checking actions by watching synchronized recordings displaying the driver and the road.

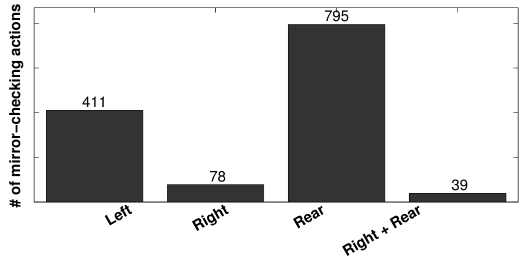


Fig. 3. Number of mirror-checking actions in the corpus. Overall, there are 1323 mirror-checking actions.

Occasionally, primary driving actions such as road glancing and traffic sign checking are confused with mirror-checking actions. In these cases, the video showing the road is useful to clarify the ambiguities, improving the accuracy of the annotation. Each observer annotates half of the recordings.

The annotators identified four types of mirror-checking actions. They include left mirror-checking, rear mirror-checking, right mirror-checking, and combined right and rear mirror-checking (referred to as "Right+Rear"). The later actions usually happens when the driver is about to turn right or change lanes. Each mirror-checking action follows these steps: the driver initiates the mirror-checking action by moving his/her head towards the target mirror; the driver fixes his/her gaze on the target mirror for a short period; finally, the driver gazes back to the road. The annotators marked the starting and ending frames for each mirror-checking action. Figure 3 summarizes the number of mirror-checking actions identified in the corpus. Looking at the rear mirror is the most common mirror-checking action, followed by checking the left mirror. A possible explanation for the distribution of mirror-checking actions is that drivers frequently use the rear mirror to be aware of the driving environment, while the side mirror are mainly used for changing lanes or turning. Figure 4 shows sample frames for mirror-checking actions.

4 Mirror-checking Behavior Analysis

As discussed in Section 2, previous studies have observed changes in mirror-checking behaviors when the drivers are under different driving conditions. Most of these changes are observed in terms of frequency and duration of mirror-checking actions [1], [2], [7], [8], [12]. Following these studies, this section analyzes how drivers' mirror-checking behaviors change in terms of these two measurements, while performing secondary tasks and driving maneuvers.

4.1 Secondary Tasks and Mirror-Checking Actions

This section analyzes differences on duration and frequency of mirror-checking actions between normal and secondary task conditions. We only compare the behaviors from normal driving recording collected over the corresponding route segment for each secondary task (see Sec. 3). This setting is particularly important for studying mirror-checking behaviors, since different routes require different glance actions. We calculate the duration of mirror-checking actions using the starting and ending frames from the manual annotations for each highlighted action. We estimate the frequency by counting the total number of mirror-checking actions occurred within non-overlap 5-second window (referred to as "5s frequency").

Figure 5 shows the histogram of the mirror-checking duration and the 5s frequency under both normal and secondary task driving conditions. Most of the mirror-checking actions have durations between 0.5 and 1 second. The figure also shows the low frequency of mirror-checking actions. We rely on the Wilcoxon signed-rank test [31] to assess whether these measurements are different under normal and task conditions. This is a nonparametric statistical hypothesis test commonly used when the data is not normal distributed, as in this case (common statistical tests such as t-test and ANOVA assume that the data is normal distributed).

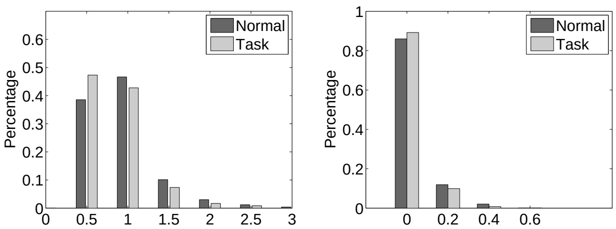
Table 3 shows the test results for each of the seven secondary tasks, where we assert significance when $p < 0.05$. A "1" in the table indicates that the differences between normal and task conditions are significantly different (otherwise, we list a "0"). The Right-tailed Wilcoxon signed rank test indicates that the mirror-checking duration is significantly higher for normal driving condition for the secondary tasks *Radio*, *GPS - Operating*, *GPS - Following* and *Phone - Operating*. Drivers perform shorter mirror-checking actions while engaged in these secondary tasks. The mirror-checking frequency is significantly higher under normal driving condition for the secondary tasks *GPS - Operating*, *Phone - Operating*, *Phone - Talking* and *Picture*. This result indicates that both the duration and frequency of mirror-checking actions are affected by secondary tasks.

The only secondary task where both the duration and the 5s frequency features are not significantly different from normal driving behaviors is *Conversation*. In some of the recordings, the second passenger sat on the back seat to control the equipments. For the task *Conversation*, we noticed that drivers used the rear mirror to establish eye contact with the passenger who is sitting in the back seat. Therefore, the mirror-checking actions were induced by the secondary task, affecting the duration and frequency of the behaviors. This is the only secondary task where some of the mirror-checking actions are induced by the task. Notice that this is not an artifact of the corpus, but an actual approach used by drivers to establish conversation with another passenger. In this case, mirror-checking actions do not increase the drivers' situation awareness. Instead, the passenger in the back seat takes away the driver attention from the road.

Overall, the results from this section agree with previous studies showing that mirror-checking behaviors are different when drivers are performing secondary tasks. These results suggest that mirror-checking behavior is a

(a) Left mirror-checking (b) Rear mirror-checking (c) Right mirror-checking

Fig. 4. Sample frames identified by external observer as mirror-checking actions.



(a) Duration [sec] (b) Frequency

Fig. 5. Histogram of the mirror-checking duration and frequency under normal and task conditions.

TABLE 4
Summary of driving maneuver annotation (T - turns, LS - lane switch).

Maneuver	Number			Duration [sec]		
	Normal	Task	All	Normal	Task	All
Left T	139	138	277	2687	2573	5260
Right T	120	119	239	1328	1221	2549
Left LS	36	49	85	297	293	590
Right LS	30	35	65	125	185	309
Straight	425	419	844	10877	11402	22279
Stop	87	80	167	1665	1523	3188

useful indicator to detect when drivers are engaged in secondary tasks. We explore these ideas in Section 6.

4.2 Driving Maneuvers and Mirror-Checking Actions

This section explores the differences in the drivers' mirror-checking behaviors while they are performing different driving maneuvers. We manually annotate the entire corpus with the six most common driving maneuvers: left turn, right turn, left lane switch, right lane switch, stop and driving straight [32]. Since the vehicle's movements is the result of the driving maneuvers, we conduct the annotation by observing the road camera video. We annotate the beginning of the turning maneuvers when the vehicle starts deviating from the driving lane. We annotate the ending point when the vehicle finished turning and get to the center of the new driving lane. For lane switch annotations, we label the beginning point as two seconds before the vehicle starts moving towards the target driving lane, to capture the behaviors displayed while the drivers prepare to complete the maneuver. The ending point is determined when the vehicle starts moving along the new driving lane. For stop maneuvers, the annotation starts when the vehicle is approaching either the stop sign or a red light. The annotation ends when the vehicle starts moving again. Unlike other driving maneuver studies [32], [33], we do not distinguish between curving and driving straight, since both maneuvers present similar behaviors in our corpus. We annotate these two cases as straight driving maneuver. Table 4 summarizes the driving maneuver annotation results.

TABLE 3
Right-tailed Wilcoxon signed rank test for mirror-checking duration and 5s frequency between normal and task driving conditions. We assert significance when $p<0.05$ (1 - significant, 0 - not significant)

Task	Duration	5s Frequency
Radio	1	0
GPS Operating	1	1
GPS Following	1	0
Phone Operating	1	1
Phone Talking	0	1
Picture	0	1
Conversation	0	0

For simplicity in the analysis, we ignore the difference between left and right maneuvers and combine them as one maneuver: left and right turns are regarded as turns, and left and right lane switch are regarded as lane switch. Notice that clustering the maneuvers assumes similar mirror-checking behaviors between left and right maneuvers in terms of frequency and duration. In addition, we focus our analysis on the mirror-checking behaviors while the vehicle is moving, thus stop maneuver is not considered in this study. As a result, the analysis focuses on 3 driving maneuvers: driving straight, turning and switching lanes. Using the mirror-checking annotation, we estimate the mirror-checking frequency and duration associated with each driving maneuver. The duration is calculated using a similar approach as before. The frequency is estimated by normalizing the number of mirror-checking actions by the associated maneuver duration.

Following the approach used in Section 4.1, we used the Wilcoxon signed-rank test to compare the duration and frequency during different driving maneuvers. We assert significance when $p<0.05$. The analysis considers

TABLE 5
Right-tailed Wilcoxon signed rank test with 5% significance level for mirror-checking duration and frequency under different driving maneuvers: T - turns, S - straight, LS - lane switch, (1 - significant, 0 - not significant).

	Duration			Frequency		
Maneuvers	**T**	**S**	**LS**	**T**	**S**	**LS**
Turns (**T**)	-	1	0	-	0	0
Straight (**S**)	0	-	0	0	-	0
Lane switch (**LS**)	0	1	-	1	1	-

pair-wise comparison between the driving maneuvers. Table 5 shows the comparison results for duration and frequency of mirror-checking actions for the three maneuvers (1 - significant, 0 - not significant). The mirror-checking duration for turn and lane switch maneuvers are significantly longer than the mirror-checking duration observed for the straight maneuver. For mirror-checking frequency, we observe that lane switch maneuver induces significantly higher number of mirror-checking actions than turn and straight maneuver.

Overall, the results suggest that drivers have different mirror-checking behaviors while performing different driving maneuvers. This conclusion motivates the idea of using mirror-checking behaviors as features for maneuver detection, as discussed in Section 6.

5 Mirror-Checking Action Detection

An important goal of this study is to detect the drivers' mirror-checking actions using noninvasive sensors. Few studies have addressed the intrinsic relation between mirror-checking actions and traffic, vehicle dynamic and driver behaviors. Cai et al. [34] observed that less mirror-checking actions is associated with poor lane control. When traffic density increases, Wierwille et al. [35] showed that drivers focus more on the traffic scene, affecting their peripheral awareness. These findings suggest that it is feasible to detect driver mirror-checking behaviors using multimodal features describing the vehicle dynamic, road conditions, and driver behaviors.

We segment the corpus into five second videos with four seconds overlap resulting in 36593 segments (see Fig. 6). As described in Section 3.2, we form a feature vector using statistics extracted from multiple sensors over these segments. We hypothesize that the audio does not provide discriminative information about mirror-checking actions. Therefore, this feature is excluded, forming a 260D feature vector. There are 7894 data segments where CERT failed to extract facial features. We create two sets of segments: *All Data*, including all of the 36593 segments; and, *Reduced Set*, including segments where CERT successfully detected facial features (28699 segments). The proposed process for estimating mirror-checking actions introduces a delay of 5 seconds.

We model the detection of mirror-checking actions as a binary classification problem, where the positive class has mirror-checking actions, and the negative class has all other driving behaviors. The data for the two classes is labeled based on the manual annotations of mirror-checking actions (Section 3.3). The videos are labeled as a positive class if they include more than 80% of a labeled mirror-checking action within the period (see Fig. 6). Otherwise, the videos are labeled as non mirror-checking actions. Notice that the 80% threshold refers to the percentage of the mirror-checking action included in the 5s window. In Section 5.2, we demonstrate that changing this threshold does not affect the performance of the proposed classifier (see Table 6). With this approach, 4141 videos are labeled as mirror-checking action (14.4%) and 24558 videos (85.6%) are labeled as non mirror-checking actions (*Reduced Set*). Notice that we do not distinguish between types of mirror-checking actions since the classification problem would be even more unbalanced, especially for right mirror-checking events (see Fig. 3). For applications that require to know the target mirror, we could implement a second classification stage that categorizes the type of mirror-checking actions of the detected videos.

5.1 Imbalanced Classification Problem

Normal machine learning algorithms attempt to minimize the overall error function. Applying these algorithms to this highly imbalanced problem (14.4% versus 85.6%) would result in a heavily skewed classification outcome where the overall error would be dominated by the error of the overrepresented class. One way to address this problem is to modify the number of samples per class before training using up- or down-sampling mechanisms. Another way is to modify the training process such that the minority class receives more weights. These methods include cost sensitive learning and minority class based recognition learning. This section highlights three machine learning algorithms that are suitable for this task.

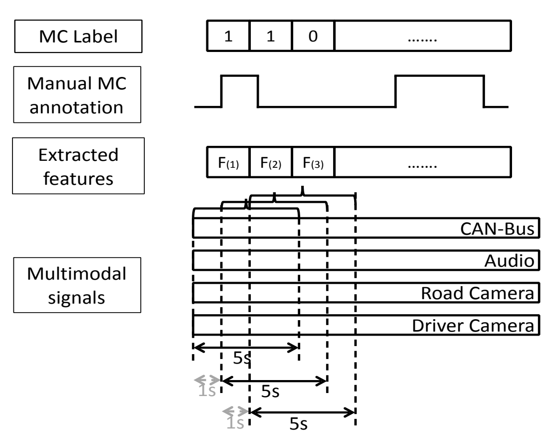


Fig. 6. The multimodal data is segmented into 5-second segments for mirror-checking binary classification. This figure illustrates the process of creating positive and negative classes (see MC label).

5.1.1 Modified Support Vector Machine

The soft margin *support vector machine* (SVM) constructs a separation hyperplane in the feature space $X = \{x_1, x_2, \ldots, x_n\}$ to maximize the margin between the classes $y_i \in \{1, -1\}$. The margin is defined as

$$y_i \left(x_i^T w + b \geq 1 - \xi_i \right) \tag{1}$$

where the penalty term $\xi_i > 0$ is used for dealing with non-separable data. Any data point that falls within the margin on the correct side of the separating hyperplane ($0 < \xi_i < 1$), or on the wrong side of the separating hyperplane ($\xi_i > 1$) will have a penalty cost $C\xi_i$ associated with it. In this case, the maximum margin is found by formulate the well-posed optimization problem (using $l1$ regularization):

$$\begin{array}{ll} min & \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \\ s.t. & y_i \left(x_i^T w + b \geq 1 - \xi_i \right) \quad \xi_i > 0 \end{array} \tag{2}$$

Although the soft margin SVM has been successfully applied to many challenging machine learning problem, it has very limited success when applied to imbalanced datasets. The standard SVM classifier assumes uniform prior probabilities, which is invalid when the classes are not balanced. To address the unbalanced classes, a modified SVM can be used where the weights between the majority and minority classes are adjusted [36]. In particular, we rescaled the cost factor C to $\frac{N}{2N_i}$ for class i, where N_i is the number of samples in class i and $N = \sum N_i$. The formulation for SVM becomes:

$$\begin{array}{lll} min & \frac{1}{2} w^T w + C_i \sum_{i=1}^{n} \xi_i & \\ s.t. & y_i \left(x_i^T w + b \geq 1 - \xi_i \right) & \xi_i > 0 \\ & C_i = \frac{N}{2N_i} & N_i = \sum (y = y_i) \end{array} \tag{3}$$

This rescale increases the penalty cost on mistakes made on the minority class (C_i is higher when N_i is small). As a result, this version of the SVM classifier is less affected by unbalanced data sets.

5.1.2 1 Nearest Neighbor

Previous studies have used k-NN in imbalanced classification problems [37], [38]. Unlike other classification algorithms that construct a general internal model such as *Linear discriminant Classifier* (LDC) and *Support Vector Machine* (SVM), k-NN classifier is a non-parametric method that simply stores the instances of the training data. For a given test data x_t, it looks for its k nearest neighbors in the training data, measured by a distance metric $\phi(x_t)$. The majority vote among the labels of the k selected samples in the training set is used to determine the label of the test sample. It is formulated as:

$$y_t = \arg \max_{l \in \{-1, 1\}} \sum_{x_i \in \phi(x_t)} I(y_i = l) \tag{4}$$

where $I()$ is an indicator function. For the unbalanced dataset, the majority class has higher chances to appear in $\phi(x_t)$, the k-neighborhood of x_t, especially near the class boundary region. Thus the output of the classifier favors the majority class. One simple solution is to use 1-NN classifier (the nearest neighbor classifier) where the label of the closest training sample is used to label the test data. The output is only determined by the nearest sample, reducing the effect of a dominant class.

k-NN is sensitive to overfitting when the feature dimension increases (e.g., feature vector with thousands of features [39]). However, the mere number of dimensions does not necessarily affect k-NN, since relevant additional dimensions can also increase the contrast between the classes [40]. Only irrelevant dimensions reduce the discrimination. As discussed in Section 5.2, we do not observe overfitting problems when using 1-NN in our detection problem.

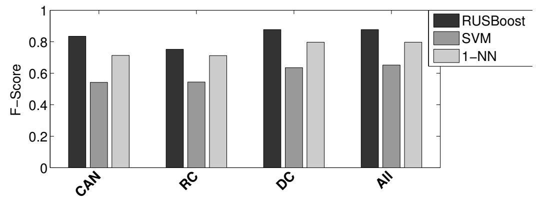
5.1.3 Random Under-Sampling Boost

The *random under-sampling boost* (RUSBoost) classifier combines both the data level and training level methods to address the imbalanced classification problem: *random under-sampling* (RUS) and boosting. The RUS step consists in applying random under-sampling over the majority class to form a balanced dataset between classes. Therefore, many classifiers can be trained using the balanced datasets. Since each classifier is trained with limited training data, they are considered as weak learners. By applying boosting algorithm (standard Adaboost method), these weak classifiers are combined into a boosted classifier with better performance. The idea of boosting is especially appealing for imbalanced classification problems since the samples from the minority class are likely to be misclassified. Increasing the weights for these samples is equivalent to increasing the cost of misclassifying the minority samples. The detailed description of the RUSBoost algorithm can be found in Seiffert et al. [41].

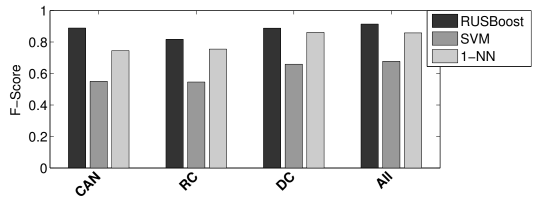
5.2 Mirror-Checking Detection Result

We consider the three classifiers described in Section 5.1 to evaluate their performance in the mirror-checking detection problem (260D features). For the RUSBoost algorithm, we use deep trees as the weak learners, which is achieved by setting the minimum observations for each leaf (5 in this case). The depth of a weak learner tree makes a difference for training time, memory usage, and predictive accuracy. We set the learning rate to 0.1, and constructed 500 such deep trees to form the ensemble. This setting has been widely used in other studies [42]–[44]. We implement SVM with linear kernel using the formulation presented in Section 5.1, where we change the cost factor C according to Equation 3. We derive the solution using quadratic programming optimization.

We use a 20-fold driver independent cross-validation approach to evaluate the classifiers, where in each fold the data from one driver is used for testing and the



(a) All Data (36593 samples)



(b) Reduced Set (28699 samples)

Fig. 7. Comparison of mirror-checking detection using different modalities and classifiers. The reduced set ignores samples when CERT fails to extract facial features (*CAN* = CAN-Bus signal; *RC* = road camera; *DC* = driver camera; *All* = 260D feature vector).

data from other drivers is used for training. This cross-validation approach guarantees that the data for training and testing are completely different, coming from different drivers, collected during different days. Given the unbalanced dataset, we measure performance in term of F-score as follows. First, we compute the precision (i.e., positive predictive value) and recall (i.e., sensitivity value) rates for each of the two classes. For a given class, the precision rate is the fraction between the number of samples that are correctly classified and the total number of recognized samples. The recall rate is the fraction between the number of samples that are correctly classified (for a given class) and the total number of samples for that class. We estimate the average precision and recall rates across both classes. Then, we estimate the F-score by combining both metrics (see Eq. 5). Given this definition, the F-score is 0.5 for random performance, regardless of how unbalanced the dataset is.

$$F = 2 * \frac{precision * recall}{precision + recall} \tag{5}$$

We evaluate the performance over the *Reduced Set* set (28699 samples where 7894 data samples are excluded due to unreliable CERT feature extraction) and *All Data* samples (36593 samples). This comparison allows us to quantify the performance when features from the driver camera cannot be robustly estimated.

Figure 7 shows that the best performance is achieved with RUSBoost (F score of 0.877 for *All Data*, and 0.914 for *Reduced Set*). We believe that RUSBoost outperforms other classifiers since it effectively combines the data level, and training level methods to deal with unbalanced classes. The 1-NN classifier also provides high F-score (F score of 0.797 for *All Data*, and 0.858 for *Reduced Set*). To evaluate potentially problems associated with using a high dimensional feature vector in k-NN, we train a 1-NN classifier using the first 30 features identified by the *forward feature selection* (FFS) algorithm. The performance is very similar to the F-score achieved with the 260D feature vector (F score of 0.811 for *All Data*, and 0.873 for *Reduced Set*), suggesting that 1-NN does not present overfitting problems in our study. The modified SVM does not have good performance (F score of 0.652 for *All Data*, and 0.677 for *Reduced Set*). We believe this is mainly due to the limitation of the linear kernel. The fact that the 1-NN and RUSBoost have better F-scores than the modified SVM suggests a highly non-linear separation hyper plane in the features space between the mirror-checking and non mirror-checking behaviors.

To understand the contribution of individual modalities, we also compare the results of mirror-checking detection using features extracted only from the CAN-Bus, road camera or driver camera (See Fig. 7). As expected, the driver camera contain rich information about drivers' head movement. Therefore, the performance is close to the one achieved using all the modalities. We also observe the discriminative power of CAN-Bus features. These results validate the close relation between driving maneuvers and mirror-checking actions. Overall, the results in Figure 7 suggest that when one modality is not available, one can rely on different modalities to detect drivers' mirror-checking actions.

Figure 8 shows the mirror-checking detection performance under both normal and task conditions per route segment (normal refers to the recording over the second lap from the corresponding route segments – see Fig. 2). We observe consistent performance of mirror-checking behaviors for normal and secondary tasks, which implies that secondary tasks do not affect the performance of the proposed mirror-checking detection system (no statistical differences between normal and task conditions at 1% significant level using pairwise t-test). Even for *Conversation*, the difference is not statistically significant, where the objective of checking the mirror may be to establish eye contact with the passenger in the back seat. We observe that the route have an effect on the performance, especially for the route associated with GPS following. This route segment has more intersections. Even for this route, the drop in performance is consistent across normal and secondary tasks.

As mentioned before, we form the binary classes (positive versus negative) by evaluating if 80% of the duration of a mirror-checking action was included in the 5-second window. The labels for the classes depend on this threshold (set to 80%). To validate our assumption that the effect of this threshold is small, we evaluate the mirror-checking detection with different thresholds (from 50% to 100%). Table 6 reports the classification results achieved by the RUSBoost classifiers for the *Reduced Set* in term of F-score. It also reports the accumulative confusion matrix across folds for each condition (i.e., it includes all samples in the *Reduced Set*). When the threshold is set to 80%, the F-score is 0.914. The accuracy

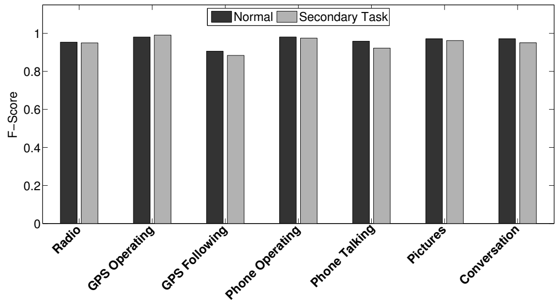


Fig. 8. Mirror-checking detection result for each road segment under both normal and secondary task conditions.

TABLE 6
Confusion matrix for the detection of mirror-checking actions using RUSBoost (*Reduced Set*). When the threshold to set the classes is set to 80%, the classifier achieves a F-score=0.914, and accuracy = 95.9%.

Threshold F-score	50% 0.9101		60% 0.9115		70% 0.912	
	N-MC	MC	N-MC	MC	N-MC	MC
N-MC	23790	352	23939	333	24060	341
MC	973	3584	941	3486	896	3402
Threshold F-score	**80% 0.9132**		**90% 0.914**		**100% 0.912**	
	N-MC	MC	N-MC	MC	N-MC	MC
NMC	24232	321	24256	302	24452	314
MC	863	3283	872	3269	833	3100

is 95.9%, which is very high given the challenging problem of estimating features from real driving recordings. The table shows similar results across different values of this threshold. The results suggest that the definition of the classes is not very sensitive to this threshold. Most mirror-checking actions have duration less than 2s (see Fig. 5(a)). Therefore, only a small portion of the data will have different labels if the threshold is changed.

These promising results suggest that mirror-checking actions are highly detectable behaviors using multimodal signals. It provides a solid foundation for the second purpose of this study which is to use mirror-checking behaviors as contextual information to detect driver distraction and driving maneuver. The following evaluations consider the *Reduced Set* with the RUSBoost-based mirror-checking detection system.

6 Maneuver and Secondary Task Detection with Mirror-Checking Actions

Various recent driver distraction studies have considered the gaze behavior as a useful indicator [45]–[49]. The goal of this study it to evaluate the effect of mirror-checking actions on detecting drivers engaged in primary (i.e., vehicle maneuvers) and secondary (i.e., tuning radio) tasks. To the best of our knowledge, this is the first study that uses automatically detected mirror-checking actions as feature to recognize these problems.

As described in Section 5, the RUSBoost detector processes five-second segments, providing a sequence of binary label indicating whether the segments contain mirror-checking actions. Since the segments have four seconds overlap, detection is performed every one second. We refer to this value as *mirror-checking* (MC) feature and we use it to classify driving maneuvers and secondary tasks. In addition, we estimate two extra features derived from this binary sequence to characterize mirror-checking duration and frequency. These features are motivated by the analysis in Section 4, which showed significantly differences in mirror-checking duration and frequency across secondary tasks and driving maneuvers. Since these two features need to be estimated within a time window, we limit the window size to be small to localize the effect of mis-detection. Specifically, we use two and five seconds windows (i.e., two and five values from the mirror-checking binary sequence, respectively). To approximate mirror-checking duration, we calculate the ratio between the frames detected as mirror-checking actions and the duration of the window analysis. This feature is referred to as *mirror-checking duration* (MCD). Likewise, we estimate the *mirror-checking frequency* (MCF) by counting the number of transition between mirror-checking actions and non-mirror-checking actions over the window analysis. This number is divided by the duration of the window analysis. We also generate these mirror-checking features (MC, MCD and MCF) directly from the manual annotations to study the effects of errors on the mirror-checking detections.

If we consider the windows used for estimating mirror-checking actions (5s) and the mirror-checking features (2s or 5s), the proposed approach has a delay of either 6s or 9s (the first frame in the window analysis is estimated including the last four seconds). This delay is needed to estimate accurate estimation of these mirror-checking features, which may prevent real-time implementation of the approach in some applications.

6.1 Secondary Task Detection

We investigate the effectiveness of using mirror-checking features (MC, MCD and MCF) to detect when drivers engage in secondary tasks. We estimate the mirror-checking detection over five-second windows with four seconds overlap. With this setting, the numbers of segments for *GPS-Operating* and *Phone-Operating* are very limited, since drivers can complete these tasks in seconds. Therefore, this section consider the tasks *Radio*, *GPS-Following*, *Phone-Talking*, *Picture* and *Conversation*.

The classification evaluation follows the approach proposed in our previous studies consisting in binary classification problems between normal and each of the secondary tasks conditions [19], [23] (i.e., normal versus *Radio*, normal versus *Phone-Talking*). The secondary task data (first lap) and the normal task data (second lap) are derived from the corresponding route segment, reducing the dependency on the route. We compare the performance of classifiers trained with and without the proposed mirror-checking features. For sake of simplicity, we use *linear discriminant classifiers* (LDCs) for the evaluation. Following the methodology used in Section 5, we implement the classification evaluation with a 20-fold driver-independent cross-validation scheme.

We train the baseline classifiers as following: First, we estimate the 268D multimodal features for each segment. Then, we reduce the feature set for each binary classifier using *forward feature selection* (FFS). We use the deviance of the linear regression as the criterion, where the independent variables are features and the dependent variable is the driving condition label (0 - task, 1 - normal). The features are added to the model as long as the the change in deviance is more than a threshold, which is defined by a chi-square distribution with one degree of freedom. The classification evaluations in Li et al. [19] showed that the best performance was achieved when approximately 15 features were selected. Starting from an empty set, we add features until we get a 15D feature vector (without including mirror-checking features). The column "Baseline" in Table 7 reports the F-score for these classifiers. Overall, the classification performances are similar or better than the ones reported in Li et al. [19] with SVM. The main difference between the experimental settings is the increased number of features considered in this study (AUs and features extracted from the road camera).

We use the following approach to evaluate the contribution of the proposed mirror-checking features for these binary classification problems. Starting from the mirror-checking features, we add new features from the 268D set using FFS using the same criteria used for the baseline. We stop the process when we reach a 15D feature vector (including the mirror-checking features). This approach forces the classifiers to use the mirror-checking features, maintaining the dimension of the feature vector. Therefore, we can directly compare the classification performance of the classifier trained with and without the mirror-checking features. The evaluation separately considers MCF, MCFR, and MCDR. It also considers using these three features together.

Table 7 reports the results. We highlight the best performance per row. Using mirror-checking features derived from manual annotation improves the F-score for the five secondary tasks (3.7% relative improvement on average). Extracting features from the automatic mirror-checking detector does not significantly affect the performance. With the exception of *Phone-Talking*, the automatically estimated mirror-checking features improve the detection performance of secondary tasks (3.1% relative improvement on average). All the mirror-checking features contribute to improve the classification performance for some of the secondary tasks. Given that the considered secondary tasks cover various types and levels of distractions (see Li and Busso [20]), the results suggest the robustness of the proposed mirror-checking features against different types of distractions. For example, the duration feature helps more when the secondary task involves more cognitive distraction (*Phone-talking*, *Picture* and *Conversation*). Drivers reduce the mirror-checking duration when they are engaged in a high cognitive secondary task.

We study the selected features to understand better the contribution of mirror-checking features. We compare the selected features before and after including the mirror-checking features across all secondary tasks. Figure 9 shows the percentage of removed (left) and added (right) features per modality. The removed features provide redundant information when we consider the proposed mirror-checking features. As expected, most of them comes from the driver camera (69%). The features that are added after including mirror-checking features provide complementary information. Although the majority of them come from the driver camera, the percentage significantly drops to 37%, allowing features from other modalities to be included. Each modality contains rich information capturing different driver behaviors. For example, the driver face camera not only provides information related to eye and head movement, but also includes lip and eyebrow movement. The different statistical measurements estimated from these signals also provide different information. For example, the mean and standard deviation from *Speed* provide cues to understand the car dynamics and the road conditions. Due to these reasons, it is possible for features from the same modality, or even the same signal, to be added or removed when we incorporate mirror-checking features.

We also estimate the most frequently added/removed features when the mirror-checking features are included across the 20 cross-validation folds. Figure 10 shows the results where black bars correspond to the most frequently added features, and gray bars correspond to the most frequently removed features. The feature AU1 Max (driver camera) is frequently added to the mirror-checking features. Our previous study showed that this feature is closely related to cognitive distractions [22]. Another interesting observation is that features related to the steering wheel movement (jitter mean, jitter STD and jitter max – CAN-Bus features) are among the most frequently removed features. This suggests a close relationship between mirror-checking actions and subtle jitter driving maneuvers. Notice that the only signals repeated are AU1 and RC intensity. Although these

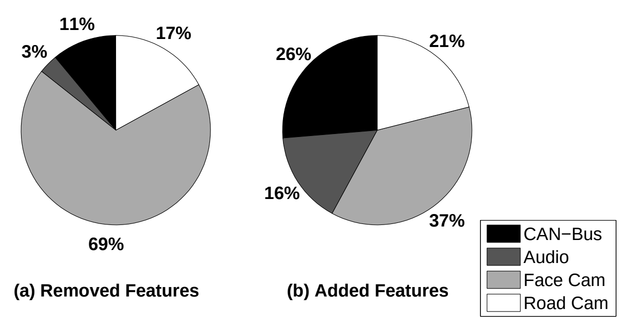


Fig. 9. Percentage of features for each modality that are either (a) removed or (b) added when mirror-checking features are used.

TABLE 7
Secondary task detection with and without mirror-checking features. The table highlights in bold the best F-score per row (MC - mirror-checking detection, MCF - mirror-checking frequency, MCD - mirror-checking duration).

	Baseline	Mirror-Checking Feature Extraction	MC	MCF		MCD		All	
			5s	2s	5s	2s	5s	2s	5s
Radio	0.833	**Manual annotations**	0.821	0.830	0.817	0.831	0.823	0.843	**0.849**
		RUSBoost detector	0.817	0.829	0.810	0.819	0.836	0.832	**0.844**
GPS-Following	0.694	**Manual annotations**	0.699	0.711	**0.738**	0.736	0.723	0.718	0.717
		RUSBoost detector	0.725	0.712	0.733	0.741	0.735	**0.754**	0.717
Phone-Talking	0.817	**Manual annotations**	0.818	0.773	0.789	**0.828**	0.794	0.818	0.795
		RUSBoost detector	**0.816**	0.776	0.802	0.802	0.792	0.792	0.776
Picture	0.869	**Manual annotations**	0.895	0.875	0.870	**0.906**	0.874	0.860	0.866
		RUSBoost detector	0.865	0.878	0.868	0.848	**0.882**	0.852	0.876
Conversation	0.655	**Manual annotations**	**0.685**	0.659	0.647	**0.685**	0.666	0.673	0.655
		RUSBoost detector	0.662	**0.681**	0.658	0.64	0.606	0.652	0.68

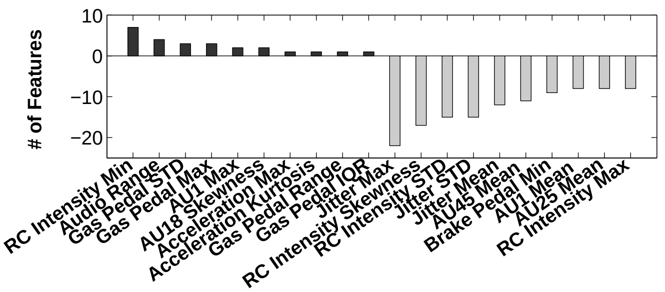


Fig. 10. Top 10 added (black bars) and removed (gray bars) features when mirror-checking features are included in the secondary task classifiers across 20 folds.

signals exist in both the removed and added features, the features are not the same (different statistics).

6.2 Driving maneuver detection

We follow the same methodology to compare the performance of maneuver classifiers trained with and without mirror-checking features (feature selection, classifiers, baselines). The only difference in this evaluation is the initial feature set. Instead of using the 268D feature vector, we consider only the features derived from the CAN-Bus signal (48 features), since driving maneuvers are highly related to the vehicle dynamics. We use FFS to reduce the set to 15 features. We evaluate three binary classifiers (one maneuver versus all the others).

Table 8 reports the results, which show improvements for all maneuver detection tasks when we consider mirror-checking features derived from the manual annotations (8.3% relative improvement on average). When we use mirror-checking features from the RUSBoost detector, the performance increases for all maneuvers, with the exception of straight maneuver detection (4.2% relative improvement on average). We achieve over 20% and 12% improvements for lane switch detection with mirror-checking features derived from manual annotation and RUSBoost detector, respectively. The improvement in performance is expected since mirror-checking features provide complementary information that can not be captured by the CAN-Bus signals. For example, they provide relevant behaviors, as the driver prepares to perform the maneuver (e.g., checking mirrors before turning). These promising results suggest that these mirror-checking features can be used as contextual cues to identify drivers changing lanes without signaling.

7 Conclusions and Discussion

The study explored the differences in mirror-checking behaviors observed when the drivers are engaged in primary and secondary tasks. We found statistically significant differences in mirror-checking duration and frequency under different driving conditions. We presented a framework to detect mirror-checking behaviors, using the RUSBoost algorithm, overcoming the highly imbalanced machine learning problem. In spite of the challenging task, the RUSBoost detector achieved an F-score equals to 0.914 using features derived from multiple noninvasive sensors. Finally, we demonstrated the benefits of detecting mirror-checking actions by conducting classification problems for secondary task recognition, and driving maneuver detection. We extracted measurements describing mirror-checking duration and frequency from the detected mirror-checking detection. Overall, we observed improved classification performance for both task driving detection and maneuver detection when we use mirror-checking features. These promising results validate the idea of automatically detecting mirror-checking actions to provide contextual information for other classification problems related to driver behaviors.

One interesting extension of this study is to consider driver-dependent mirror-checking behavior models. Previous work on driver behavior suggest that driving patterns vary according to the experience and driving style of the drivers [17], [50]. The annotations of mirror-checking actions suggest that some drivers checked the mirror more frequently than others, even when they were driving without performing any secondary tasks. We expect that models that are adapted to the target drivers will outperforms the general models proposed in this study. Another challenging problem is to improve the mirror-checking detection. Given the unique behaviors associated with mirror-checking actions, we believe that collecting more data from various drivers will increase the robustness of the detectors. We are planning

TABLE 8
Driving maneuver detection with and without mirror-checking features. The table highlights in bold the best F-score per row (MC - mirror-checking detection, MCF - mirror-checking frequency, MCD - mirror-checking duration).

	Baseline	Mirror-Checking Feature Extraction	MC	MCF		MCD		All	
			5s	2s	5s	2s	5s	2s	5s
Turns	0.759	**Manual annotations**	0.760	0.759	**0.762**	0.761	**0.762**	0.760	0.761
		RUSBoost detector	**0.759**	0.758	0.758	**0.759**	0.758	0.757	0.757
Lane Switch	0.529	**Manual annotations**	0.603	0.528	0.528	0.533	0.533	0.603	**0.653**
		RUSBoost detector	0.557	0.530	0.529	0.534	0.536	0.578	**0.596**
Straight	0.692	**Manual annotations**	0.694	0.691	0.692	0.690	0.690	0.699	**0.700**
		RUSBoost detector	**0.692**	**0.692**	0.691	0.689	0.674	0.691	0.682

to modify the recording protocol to minimize potential learning effects. For example, we can randomize and balance the order of normal and task conditions for the first and second laps. We can also record other cues from the CAN-Bus such as control signals from radio, smartphone and GPS, which can help to robustly detect when the drivers operate these devices. Finally, the current implementation of the approach is off-line. We are planning to develop a real-time implementation, which will be an important instrument to detect deviation from normal driving behaviors. Improving the accuracy of the mirror-checking detectors will help vehicle manufactures to incorporate this technology in future ADASs.

Acknowledgments

The authors would like to thank Dr. John Hansen for his support with the UTDrive Platform, and the MPLab at UCSD for providing the CERT software.

References

[1] M. Recarte and L. Nunes, "Mental workload while driving: Effects on visual search, discrimination, and decision making," *Journal of Experimental Psychology: Applied*, vol. 9, no. 2, pp. 119–137, June 2003.
[2] K. A. Brookhuis and D. de Waard, "Assessment of drivers' workload: Performance and subjective and physiological indexes," in *Stress, workload, and fatigue*, ser. Human Factors in Transportation, P. Hancock and P. Desmond, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Associates Inc., November 2000, pp. 321–333.
[3] C. Ahlstrom, K. Kircher, and A. Kircher, "A gaze-based driver distraction warning system and its effect on visual behavior," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 965–973, June 2013.
[4] O. Georgeon, M. J. Henning, T. Bellet, and A. Mille, "Creating cognitive models from activity analysis: A knowledge engineering approach to car driver modeling," in *International Conference on Cognitive Modeling (ICCM 2007)*, Ann Arbor, MI, USA, July 2007.
[5] E. C. B. Olsen, "Modeling slow lead vehicle lane changing," Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA, September 2003.
[6] M. Itoh, K. Yoshimura, and T. Inagaki, "Inference of large truck driver's intent to change lanes to pass a lead vehicle via analyses of driver's eye glance behavior in the real world," in *Proceedings of SICE Annual Conference*, Takamatsu, Japan, September 2007, pp. 2385–2389.
[7] N. Li and C. Busso, "Driver mirror-checking action detection using multi-modal signals," in *The 6th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems*, Seoul, Korea, September-October 2013, pp. 101–108.
[8] J. Harbluk, Y. Noy, P. Trbovich, and M. Eizenman, "An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance," *Accident Analysis and Prevention*, vol. 39, no. 2, pp. 372–379, March 2007.
[9] J. L. Harbluk, M. Eisenman, and Y. I. Noy, "The impact of cognitive distraction on driver visual behaviour and vehicle control," Transport Canada, Ottawa, Ontario, Canada, Tech. Rep. TP# 13889 E, February 2002.
[10] P. Angkititrakul, M. Petracca, A. Sathyanarayana, and J. Hansen, "UTDrive: Driver behavior and speech interactive systems for in-vehicle environments," in *IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, June 2007, pp. 566–569.
[11] G. H. Robinson, D. J. Erickson, G. L. Thurston, and R. L. Clark, "Visual search by automobile drivers," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 14, no. 4, pp. 315–323, August 1972.
[12] T. Lansdown, "Visual allocation and the availability of driver information," in *Traffic and Transport Psychology. Theory and Application*, T. Rothengatter and E. Carbonell Vaya, Eds. Bingley, UK: Emerald Group Publishing Limited, September 1997, pp. 215–224.
[13] S. Fairclough, M. Ashby, and A. Parkes, "In-vehicle displays, visual workload and usability evaluation," in *Vision in Vehicles IV*, ser. International Conference on Vision in Vehicles, A. Gale, I. Brown, C. Haslegrave, H. Kruysse, and S. Taylor, Eds. Leiden, The Netherlands: North Holland, February 1993, pp. 245–254.
[14] D. Crundall, C. Shenton, and G. Underwood, "Eye movements during intentional car following," *Perception*, vol. 33, no. 8, pp. 975–986, August 2004.
[15] M. Schweigert and H. Bubb, "Eye movements, performance and interference when driving a car and performing secondary tasks," in *Vision in Vehicles IX*, ser. International Conference on Vision in Vehicles, A. Gale, J. Bloomfield, G. Underwood, and J. Wood, Eds. Brisbane, Australia: Loughborough University, August 2001, pp. 19–22.
[16] W. Murray, S. Newnam, and B. Watson, "A comparison of the factors influencing the safety of work-related drivers in work and personal vehicles," in *The Road Safety Research, Policing and Education Conference*, Adelaide, Australia, November 2002, pp. 488–495.
[17] Z. Lei, W. Jianqiang, Y. Furui, and L. Keqiang, "A quantification method of driver characteristics based on driver behavior questionnaire," in *IEEE Intelligent Vehicles Symposium (IVS 2009)*, Xi'an, Shaanxi, China, June 2009, pp. 616–620.
[18] G. Underwood, P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall, "Visual attention while driving: sequences of eye fixations made by experienced and novice drivers," *Ergonomics*, vol. 46, no. 6, pp. 629–646, May 2003.
[19] N. Li, J. Jain, and C. Busso, "Modeling of driver behavior in real world scenarios using multiple noninvasive sensors," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1213–1225, August 2013.
[20] N. Li and C. Busso, "Using perceptual evaluation to quantify cognitive and visual driver distractions," in *Smart Mobile In-Vehicle Systems – Next Generation Advancements*, G. Schmidt, H. Abut, K. Takeda, and J. H. L. Hansen, Eds. New York, NY, USA: Springer, January 2014, pp. 183–207.
[21] ——, "Analysis of facial features of drivers under cognitive and visual distractions," in *IEEE International Conference on Multimedia and Expo (ICME 2013)*, San Jose, CA, USA, July 2013.
[22] ——, "Predicting perceived visual and cognitive distractions of drivers with multimodal features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 51–65, February 2015.
[23] J. Jain and C. Busso, "Analysis of driver behaviors during common tasks using frontal video camera and CAN-Bus information," in *IEEE International Conference on Multimedia and Expo (ICME 2011)*, Barcelona, Spain, July 2011.
[24] K. Takeda, J. Hansen, P. Boyraz, L. Malta, C. Miyajima, and H. Abut, "International large-scale vehicle corpora for research on driver behavior on the road," *IEEE Transactions on Intelligent*

Transportation Systems, vol. 12, no. 4, pp. 1609–1623, December 2011.

[25] H. Abut, H. Erdoğan, A. Erçil, B. Çürüklü, H. Koman, F. Taş, A. Argunşah, S. Coşar, B. Akan, H. Karabalkan *et al.*, "Real-world data collection with "UYANIK"," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, H. Erdoğan, J. Hansen, and H. Abut, Eds. New York, NY, USA: Springer, 2009, pp. 23–43.

[26] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 97 – 120, March 2005.

[27] M. Bartlett, G. Littlewort, T. Wu, and J. Movellan, "Computer expression recognition toolbox," in *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008)*, Amsterdam, The Netherlands, September 2008.

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

[29] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Human- Computer Interaction*, ser. Lecture Notes in Computer Science, M. Lew, N. Sebe, T. Huang, and E. Bakker, Eds. Berlin, Germany: Springer Berlin / Heidelberg, December 2007, vol. 4796, pp. 6–18.

[30] M. Sodhi, B. Reimer, and I. Llamazares, "Glance analysis of driver eye movements to evaluate distraction," *Behavior Research Methods, Instruments, & Computers*, vol. 34, no. 4, pp. 529–538, November 2002.

[31] S. Siegel, *Nonparametric statistics for the behavioral sciences*. McGraw-Hill Companies, 1956.

[32] A. Sathyanarayana, P. Boyraz, and J. Hansen, "Driver behavior analysis and route recognition by hidden markov models," in *IEEE International Conference on Vehicular Electronics and Safety (ICVES 2008)*, Columbus, OH, USA, September 2008, pp. 276–281.

[33] D. Johnson and M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *IEEE Conference on Intelligent Transportation Systems (ITSC 2011)*, Washington, DC, USA, October 2011, pp. 1609–1615.

[34] H. Cai, Y. Lin, and R. R. Mourant, "Study on driver emotion in driver-vehicle-environment systems using multiple networked driving simulators," in *Driving Simulation Conference North America (DSC 2007)*, Iowa City, IA, USA, September 2007.

[35] W. W. Wierwille, "Visual adaptation of the driver to high-demand driving situations while navigating with an in-car navigation system," in *Vision in Vehicles III*, ser. International Conference on Vision in Vehicles, A. Gale, I. Brown, I. Moorhead, C. Haslegrave, and S. Taylor, Eds. Aachen, Germany: North-Holland, March 1991, pp. 79–87.

[36] K. Veropoulos, C. Campbell, and N. Cristianini, "Controlling the sensitivity of support vector machines," in *International joint conference on artificial intelligence (IJCAI 1999)*, Stockholm, Sweden, July-August 1999, pp. 55–60.

[37] J. Zhang and I. Mani, "knn approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of Workshop on Learning from Imbalanced Datasets ()*, Washington, DC, August 2003.

[38] S. Tan, "Neighbor-weighted K-nearest neighbor for unbalanced text corpus," *Expert Systems with Applications*, vol. 28, no. 4, pp. 667–671, May 2005.

[39] L. Torresani and K.-C. Lee, "Large margin component analysis," in *Advances in neural information processing systems (NIPS 2007)*, Vancouver, B.C., Canada, December 2007, pp. 1385–1392.

[40] M. Houle, H. K. P. Kröger, E. Schubert, and A. Zimek, "Can shared-neighbor distances defeat the curse of dimensionality?" in *Scientific and Statistical Database Management*, ser. Lecture Notes in Computer Science, M. Gertz and B. Ludäscher, Eds. Berlin, Germany: Springer-Verlag Berlin Heidelberg, August 2010, vol. 6187, pp. 482–500.

[41] C. Seiffert, T. Khoshgoftaar, J. V. Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 40, no. 1, pp. 185–197, January 2010.

[42] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, July-December 1999.

[43] Y. Liu and X. Yao, "Simultaneous training of negatively correlated neural networks in an ensemble," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 6, pp. 716–725, December 1999.

[44] M. van Heeswijk, Y. Miche, T. Lindh-Knuutila, P. A. Hilbers, T. Honkela, E. Oja, and A. Lendasse, "Adaptive ensemble models of extreme learning machines for time series prediction," in *Artificial Neural Networks (ICANN 2009)*, ser. Lecture Notes in Computer Science, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds. Limassol, Cyprus: Springer Berlin Heidelberg, September 2009, vol. 5769, pp. 305–314.

[45] L. Yekhshatyan and J. Lee, "Changes in the correlation between eye and steering movements indicate driver distraction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 136–145, March 2013.

[46] P. Jiménez, L. Bergasa, J. Nuevo, N. Hernández, and I. Daza, "Gaze fixation system for the evaluation of driver distractions induced by IVIS," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1167–1178, September 2012.

[47] R. Oyini Mbouna, S. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1462–1469, September 2013.

[48] F. Tango and M. Botta, "Real-time detection system of driver distraction using machine learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 894–905, June 2013.

[49] C. Ahlstrom, T. Victor, C. Wege, and E. Steinmetz, "Processing of eye/head-tracking data in large-scale naturalistic driving data sets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 553–564, June 2012.

[50] J. Duncan, P. Williams, and I. Brown, "Components of driving skill: experience does not mean expertise," *Ergonomics*, vol. 34, no. 7, pp. 919–937, June 1991.

Nanxiang Li (S'2012) received his B.S. degree (2005) in Electrical Engineering from Xiamen University, Fujian, China. He received his M.S. degree (2009) in Electrical Engineering from the University of Alabama, Tuscaloosa, Alabama, USA. In 2015, he received his Ph.D. degree in Electrical Engineering at the University of Texas at Dallas (UTD), Richardson, Texas, USA. He received the third prize IEEE ITSS Best Dissertation Award 2015. He is current a research scientist at Honda Research Institute at Mountain View, California, USA. His research focuses on understanding driver behavior and intention using information from various sensor signal streams. He has also worked on multimodal interfaces with emphasis on gaze estimation, and human tracking and recognition using gait information.

Carlos Busso (S'02-M'09, SM'13) received the BS and MS degrees with high honors in electrical engineering from the University of Chile, Santiago, Chile, in 2000 and 2003, respectively, and the PhD degree (2008) in electrical engineering from the University of Southern California (USC), Los Angeles, in 2008. He is an associate professor at the Electrical Engineering Department of The University of Texas at Dallas (UTD). He was selected by the School of Engineering of Chile as the best electrical engineer graduated in 2003 across Chilean universities. At USC, he received a provost doctoral fellowship from 2003 to 2005 and a fellowship in Digital Scholarship from 2007 to 2008. At UTD, he leads the Multimodal Signal Processing (MSP) laboratory [http://msp.utdallas.edu]. He is a recipient of an NSF CAREER Award. In 2014, he received the ICMI Ten-Year Technical Impact Award. He also received the Hewlett Packard Best Paper Award at the IEEE ICME 2011 (with J. Jain). He is the co-author of the winner paper of the Classifier Sub-Challenge event at the Interspeech 2009 emotion challenge. His current research includes the broad areas of in-vehicle active safety system, affective computing, multimodal human-machine interfaces, modeling and synthesis of verbal and nonverbal behaviors, sensing human interaction, and machine learning methods for multimodal processing. He is a member of ISCA, AAAC, and ACM, and a senior member of the IEEE.