Journal of Systems Architecture 121 (2021) 102319

Contents lists available at ScienceDirect

Journal of Systems Architecture

journal homepage: www.elsevier.com/locate/sysarc

A survey on vision-based driver distraction analysis

Wanli Li [a], Jing Huang [a], Guoqi Xie [a,*], Fakhri Karray [b], Renfa Li [a,*]

[a] College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, Hunan, China
[b] Centre for Pattern Analysis and Machine Intelligence, University of Waterloo, Waterloo, N2L 3G1, Ontario, Canada

ARTICLE INFO

Keywords:
Driver distraction
Vision-based methods
Driver inattention monitoring systems

ABSTRACT

Motor vehicle crashes are great threats to our life, which may result in numerous fatalities, as well as tremendous economic and societal costs. Driver inattention, either distraction or fatigue, is the major cause among human factors responsible for the crashes. Distracted driving has been getting increasingly severe, and has caused many more crashes than drowsy driving, while the latter has been more extensively studied. Therefore, we are motivated to present a comprehensive survey on vision-based approaches for driver distraction analysis. In the paper, we firstly provide an overview of driver distraction, then introduce the available datasets and explore the various cues for driver behavior distraction analysis. After that, two forms of driver behavior distraction (visual distraction and manual distraction) are analyzed separately. Lastly, we conclude the evolvement and future directions of driver distraction analysis for safe driving. To the best of our knowledge, we are the first to propose driver behavior distraction and analyze it in a hierarchical way.

1. Introduction

With the development of science and technology, especially in the field of information technology, the drivers are inclined to be distracted, either actively or passively. Their distracted behaviors tend to be diverse due to more kinds of secondary tasks such as using smartphones and operating the in-vehicle information systems (IVISs), or the more disturbing information on the road like the appealing billboards.

From long-time observations and driving experiences, we see that it is not easy for drivers to pay full attention to driving all the time. This happens more frequently when the drivers are on the way home after working for the whole day without a sound rest, or when they have to drive continuously for hours on the highway. It is quite common for the drivers to eat or drink, talk to passengers, make a phone call or deal with an unexpected coming call, and makeup sometimes during the driving period. Driver distraction and fatigue are the two main categories of driver inattention [1]. When the drivers are in the state of slight fatigue or drowsiness, they are likely to find secondary tasks to do. It is not a bad way to help the drivers go against fatigue and stay active. However, in such case they fall into another types of inattention, converting from fatigue to distraction.

From the retrieved information and statistical analysis, we become more aware that the **distracted driving** problem is not a trivial one. According to the statistical findings by National Highway Traffic Safety Administrator (NHTSA), in 2018 there were 2841 people killed and an estimated additional 400,000 people injured in motor vehicle crashes involving distracted drivers in the United States [2]. Compared to the corresponding number in 2015 (deaths of 3477 people and injuries to 391,000 people), distracted driving is still a serious threat to road safety.

The situations have been much more severe in the developing countries, such as China, India, and Brazil. In fact, the rate of road traffic deaths in middle-income countries is nearly three times higher than that in high-income countries [3]. According to the World Health Organization (WHO) survey, deaths from road traffic crashes have increased to 1.35 million a year. That is nearly 3700 people dying on the world's roads every day [3]. Klauer et al. [4] perform an analysis using the 100-car naturalistic driving study data, so as to correlate driver inattention with crash and near-crash involvement. The results indicate that secondary-task distraction contributed to over 22 percent of all crashes and near-crashes, and glances totaling more than 2 s for any purpose increase near-crash/crash risk by at least two times that of normal, baseline driving.

The crashes can not only cause **deaths and injuries**, but also leave tangible and intangible impacts on the individual and the society, resulting in enormous **economic and societal costs**. The potential costs can involve property damage, medical costs, legal and court costs, congestion costs, emergency services costs, productivity losses, as well as quality-of-life. In 2010, the total economic cost of motor vehicle

* Corresponding authors.
E-mail addresses: liwanli@hnu.edu.cn (W. Li), jingh@hnu.edu.cn (J. Huang), xgqman@hnu.edu.cn (G. Xie), karray@uwaterloo.ca (F. Karray), lirenfa@hnu.edu.cn (R. Li).

https://doi.org/10.1016/j.sysarc.2021.102319
Received 27 June 2021; Received in revised form 29 September 2021; Accepted 18 October 2021
Available online 30 October 2021

crashes in the United States was $242 billion, while crashes in which at least one driver was identified as being distracted cost $40 billion (roughly 16 percent). When quality-of-life valuations are considered, the total value of societal harm from motor vehicle crashes was $836 billion, of which roughly 15 percent was caused by distracted driving crashes [5].

As distracted driving is such a severe and growing threat to road safety, what we can do to prevent the traffic accidents and save lives? One of the ideal solutions can be **autonomous driving**. However, due to the technological obstacles and social challenges, it is still a long way to make the dream of full automation come true. Many researchers focus on automated driving and they have made some inspiring achievements, but even the up-to-date automated vehicles cannot get rid of drivers.

There were some fatalities related to automated driving drawing great attention worldwide. On 20 January 2016, the first fatal crash of a Tesla with Autopilot took place in Hubei, China. Shortly after that, another widely known fatal accident involving Tesla Autopilot occurred in Florida, USA. On 7 May 2016, a Tesla Model S engaged in Autopilot mode hit an 18-wheel tractor-trailer, resulting in the death of the occupant. On 18 March 2018, a pedestrian was killed after being hit by an Uber self-driving car in Arizona, USA. In view of these fatalities, an attentive driver is needed for the latest self-driving vehicles to take control in case of emergency, which makes it necessary to monitor the attention level of the driver. The crashes could have prevented if the safety drivers were not distracted and took the right actions in time.

According to the standard SAE J3016 [6], there are 6 levels of driving automation, including no automation (Level 0), driver assistance (Level 1), partial automation (Level 2), conditional automation (Level 3), high automation (Level 4) and full automation (Level 5). Starting from Level 3, the systems are responsible for monitoring the driving environment [7]. But the ultimate Level 5 is far out of our reach, and the Level 4 is still restricted to highly limited scenarios. Despite of the difficulties, we are about to solve the problems step by step. We believe that **collaborative driving** for Level 3 is a promising and more realistic way to reduce traffic crashes and improve road safety. At level 3, the drivers are expected to respond appropriately to a request to intervene, so it is essential to monitor the driver attention and measure the readiness to take over [8,9]. For Level 2 or below, the drivers have to monitor the driving environment, which makes it necessary to monitor the attention of drivers.

The driver inattention monitoring systems mainly focus on driver distraction and fatigue [10]. The retrieved results within Web of Science show that, researches on driver inattention monitoring appear to have a bias on driver fatigue or drowsiness detection. However, driver distraction detection has draw increasing attention worldwide [11], especially for the last five years. Besides, the number of fatalities in distraction-affected crashes is much greater than that involving a drowsy driver based on NHTSA's statistics for recent years. Thus, it is of great value to summarize the researches on driver distraction monitoring. Given that the vision-based methods are user-friendly with excellent performances for such tasks, we are about to provide a comprehensive survey of those approaches exploiting discriminative visual features.

The **contributions** of the survey are summarized as follows: (1) In the view of machine vision, we firstly propose driver behavior distraction, which includes visual distraction and manual distraction. (2) We introduce the publicly available datasets for vision-based driver distraction analysis, in the hope of attracting more researchers to address the problem and facilitating the process. (3) We summarize the related researches in a hierarchical way. As shown in Fig. 1, the analysis begins with the datasets, then cues such as driver face, head pose, hands can be extracted as a lower level information. With this, we can get into the high-level stages and devote to driver visual or manual distraction detection separately. (4) Lastly, we summarize and discuss the evolvement and trends of driver behavior distraction analysis to help understand and solve this problem better.

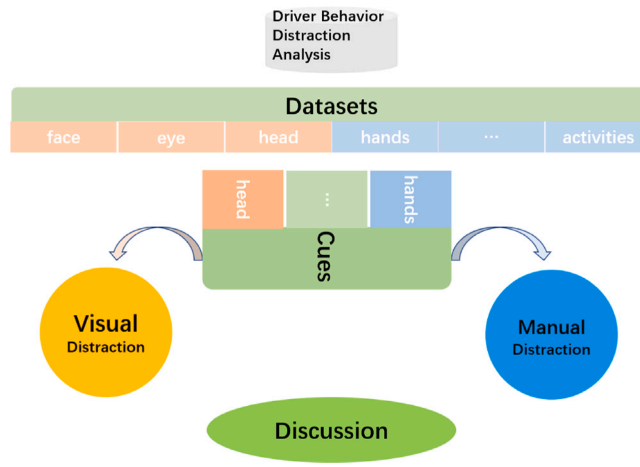


Fig. 1. Overview of our hierarchical structure for driver behavior distraction analysis.

The rest of the paper is **organized** as follows: Section 2 gives an overview of driver distraction. Section 3 and 4 present datasets and cues for detecting driver behavior distraction separately. And the visual distraction detection and manual distraction detection of driver are discussed respectively in Section 5 and 6. Lastly, Section 7 discusses the research trends and Section 8 concludes our work.

2. Driver distraction overview

In this section, firstly we will provide the definition of driver inattention, then introduce some basic concepts of a typical form of inattention, viz. the definition and categories of driver distraction. With this, we zoom in and come to the topic of this article—driver behavior distraction. Interested in applying computer vision to detect distracted driving, we focus on vision-based approaches for analyzing driver behavior distraction.

In [4], driver inattention is broadly defined as "any point in time that a driver engages in a secondary task, exhibits symptoms of moderate to severe drowsiness, or looks away from the forward roadway". From the definition, we see that driver inattention refers to a broader scope of driver behaviors, including secondary task distraction, drowsiness, driving-related inattention to the forward roadway, and non-specific eyeglance away from the forward roadway [4]. After reviewing various definitions and taxonomies of driver inattention, Regan et al. [12] conclude that **driver inattention** means "insufficient, or no attention, to activities critical for safe driving". It is worth noting that, after revisiting the taxonomy they proposed, Regan et al. clarify and elaborate on the categories of inattention, characterize the attentional processes within each category, and make recommendations for several refinements and research avenues [13]. But in this paper, what we are interested in is the analysis of driver distraction, a sub-category of driver inattention, so we are not about to tangle too much with the taxonomy of driver inattention.

Driver distraction is usually taken as a form of driver inattention. Regan et al. propose that a category of inattention called Driver Diverted Attention (DDA) is synonymous with driver distraction. They define driver distraction as "the diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving" [12]. The definition proposed here is almost identical to that coined for driver distraction by Lee et al. [14]. Furthermore, different from Regan et al. [12] deriving taxonomy from crash data, Engström et al. propose driver inattention taxonomy from a review of attentional

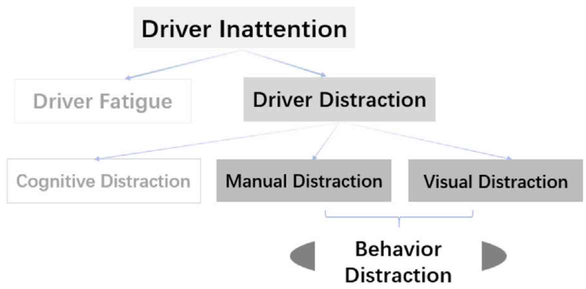


Fig. 2. Driver behavior distraction.

literature in [15], where they define driver distraction as "the driver allocates resources to a non-safety critical activity while the resources allocated to activities critical for safe driving do not match the demands of these activities". These definitions are just like all roads lead to Rome.

As the sources of distraction may take many forms, The NHTSA characterizes driver distraction briefly as "any activity that takes a driver's attention away from the task of driving", and proposes to examine distraction in terms of four distinct categories [16]:

- visual distraction (e.g., looking away from the roadway);
- auditory distraction (e.g., responding to a ringing cell phone);
- biomechanical distraction (e.g., manually adjusting the radio volume);
- cognitive distraction (e.g., being lost in thought).

To put it another way, distraction refers to anything that takes your eyes off the road (visual distraction), your mind off the road (cognitive distraction), or your hands off the wheel (manual distraction) [17]. To tell from cognitive distraction, herein we name these two forms of driver distraction as driver behavior distraction (as shown in Fig. 2), and give its definition as follows: ***Driver behavior distraction*** *refer to any behavior that can visually reflect the diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving.*

As mentioned above, driver fatigue and distraction detection are the two main topics of driver inattention monitoring. Systems designed for the analysis and detection of driver drowsiness can be broadly divided into two categories: visual features based and non-visual features based. Techniques using visual features take advantage of computer vision approaches for the detection of drowsiness [10]. Similarly, we can exploit **visual features** for monitoring driver distraction. Due to the non-intrusive nature, vision-based methods are more use-friendly. With discriminative visual features, vision-based methods are able to achieve high performance. Accordingly, we mainly concentrate on vision-based driver distraction detection in this paper, that is, driver visual distraction and manual distraction.

We must acknowledge that the comprehensive analysis of driver distraction is a complex task, which is likely to involve human factors such as driver physiological indicators, and the driving context such as road conditions, surrounding vehicles, and road vulnerable users. From this point of view, driver distraction is a kind of driver behaviors more difficult to analyze than driver fatigue. Herein we mainly pay attention to the driver and focus on vision-based approaches for detecting driver behavior distraction. There are some reviews involving driver behaviors and road safety. Some reviews focus on driver face [18], head pose [19], or mobile phone usage [20,21]. While some of them zoom out and look at this topic from a wide-angle viewpoint, including not only the humans [22], but also the vehicles and the roads [23,24]. Chhabra et al. [25] classify the various methods for driver behavior monitoring into real-time and non-real-time techniques, and make a comparative study in terms of methodology, advantages, and disadvantages. However, the driver behaviors they discussed are more of driving maneuvers, driving style, as well as risky driving related to the drunk or drowsy driver, whereas the non-trivial driver distraction is a rather small portion of their topics.

On the one hand, from [1,10,26,27], we could see that literature on driver distraction detection are far less than that of driver drowsiness detection. Besides, many of them are kinda out of date. Meanwhile, there have been emerging researchers realizing the urgent need to solve the distracted driving issue and devoting themselves to the tasks for the past decade. On the other hand, it is true that head pose and gaze estimation serves as an effective indicator to detect driver distraction. However, topics related to driver distraction such as driver posture recognition and driver gaze zones estimation are not involved in [1,10,28]. Recent reviews such as [29] addresses more on sensors for driver distraction detection, [28] aims at the new generation of driver monitoring systems within the context of Internet of Cars by introducing the concept of integrated safety. Therefore, we are **motivated** to review the recent researches on driver behavior distraction.

There could be various **challenges** for detecting distraction of driver. As for detecting driver hands, researchers are likely to deal with the hand variations, highly occlusions, low-resolution, strong lighting conditions, and varied in shape and viewpoint, as well as blurring of colors due to hand movement, skin tone variation in recorded videos due to camera quality. Regarding driver gaze zones estimation, problems such as the unpredictability of the environment, presence of sunglasses occluding the eye, rapid changes in ambient lighting including situations of extreme glare resulting from reflection, partial occlusion of the pupil due to squinting, vehicle vibration, image blur, poor video resolution are on the way, we may also need to take care of different subjects, perspectives and scale etc. To address such challenges, we start from introducing the databases and cues for detecting driver behavior distraction.

3. Datasets for detecting driver behavior distraction

Before it comes to the methods for detecting driver behavior distraction, it is necessary to describe the datasets which lie in the foundation for the mission. There were numerous datasets available for human face detection and generic head pose estimation, whereas the datasets focusing on drivers were once almost unavailable. The early researchers cannot but collect datasets by themselves to detect driver distraction, which is really time-consuming, and sometimes rather difficult due to the limited experimental conditions. Besides, it is difficult to make comparisons among experimental results from different custom datasets. Thanks to the hard work of groups of selfless researchers, there are a few publicly available datasets on driver face, head pose and eye gaze, driver hands, and driver posture as well. In this section, we will introduce the publicly available datasets which can contribute to driver behavior distraction analysis, hoping to attract more researchers to address the problem, thereby improving road safety.

3.1. Driver face

The generic face detection has been widely studied with lots of publicly available datasets such as [30,31]. But in this subsection, we are to introduce two face detection datasets specific to drivers.

The one is developed by Diaz-Chito et al. [32], a relatively simple dataset called DrivFace. The dataset is collected over different days from 4 drivers under real driving scenarios, including 606 images (640 × 480 pixels each). Each image is labeled with one of 3 possible classes in terms of gaze direction: looking-right, frontal, and looking-left.

The other one is the face challenge collected by Martin et al. [33], as one part of the Vision for Intelligent Vehicles and Applications (VIVA). The VIVA Face dataset [33,34] consists of selected images from 39 naturalistic driving video sequences, either collected by LISA-UCSD or on the YouTube. Challenges highlighted include harsh lighting

Fig. 3. An instance from the VIVA Face dataset [33].

conditions, facial occlusions, and varying viewpoints. The number of faces in each image ranges from one to four, due to the viewpoint changing from observing the driver only to the entire vehicle cabin. And Fig. 3 is an instance from the VIVA Face dataset.

3.2. Driver head pose and eye gaze

The aforementioned driver face detection datasets are also available for head pose estimation. Next we will introduce some more datasets for estimating driver head pose, and eye gaze as well.

Schwarz et al. [35] introduce a dataset called DriveAHead for head pose estimation in real driving conditions, including parking maneuvers, driving on the highway and through a small town. The DriveAHead dataset is composed of one million depth and infrared images from 20 subjects, whose head orientation and position are recorded with a motion capture system as reference pose measurements. They provide labels for head pose, occlusions, glasses and sunglasses, as well as 2D and 3D data aligned at the pixel level using the Kinect v2.

Borghi et al. [36] collect a new challenging dataset, called Pandora. The dataset has been specifically created for driver head center localization, head pose and shoulder pose estimation. Pandora is collected from 22 subjects (10 male and 12 female actors) with five times for each subject, resulting in 110 annotated sequences. To simulate the viewpoint of camera from the dashboard inside the vehicle, they use a frontal fixed Microsoft Kinect One to acquire the upper body part of the subjects. There are some driving-like actions performed by the subjects, such as grasping the steering wheel, looking to the rear-view or lateral mirrors, shifting gears and so on.

Furthermore, DD-Pose [37] by Roth and Gavrila, the Daimler TU Delft Driver Head Pose Benchmark, is a large-scale and diverse benchmark for image-based head pose estimation and driver analysis. A large-scale dataset AutoPOSE [38] is collected in a car simulator from two different camera positions, where the dashboard view camera provides about 1.1M IR images, and another center mirror view Kinect v2 provides RGB, Depth, and IR images (315K for each). Ribeiro and Costa [39] present a Driver Gaze Zone Dataset With Depth Data in real driving conditions.

3.3. Driver hands

The analysis of driver hands are important for detecting distracted driving, as hands are involved in various driver activities, either driving-related tasks such as operating the shift gear, or secondary tasks such as adjusting the radio. The following are two datasets on driver hands.

Das et al. [40] develop the VIVA hand detection dataset, which is an extensive video-based dataset for driver hand detection. They collect over 2000 annotated images from each of three testbed vehicles, and over 2000 images from YouTube to further diversify the dataset. In each frame, they annotate the hand bounding boxes, as well as left/right, driver/passenger, and number of hands on the wheel. The dataset highlights the possible challenges in naturalistic driving settings, including different background complexities, illumination settings, users, and viewpoints. They denote the datasets as two levels in terms of difficulty. The level-l (L1) serves as the easier evaluation setting which contains only back view imagery and larger instances (above 70 pixels in height), while the level-2 (L2) includes imagery from all viewpoints (including the images from the L1 setting) and instances greater than 25 pixels in height.

Borghi et al. [41] propose a new dataset Turms, which composed of infrared images of driver's hands. The Turms dataset include 14k frames, and the spatial resolution for each frame is 640×240. Due to the stereo capabilities and wide view-angle, the Leap Motion device is used to record 7 subjects (twice for each) from the back of the steering wheel.

3.4. Driver postures

The earlier datasets on driver postures concentrated on only limited set of distractions, and many of them are not publicly available. In 2016, State Farm insurance company held a competition named State Farm Distracted Driver Detection [43] on Kaggle, seeking for the answers of such a question: Can computer vision spot distracted drivers? They released a dataset of 2D dashboard camera images, challenging the competitors to classify driver behaviors. It attracted more than 1400 teams with over 1600 competitors to take part in the competition.

Despite that the State Farm dataset is publicly available, it was limited to the purpose of the competition [44]. Inspired by this, Abouelnaga et al. created a similar dataset named AUC Distracted Driver Dataset [45] and made it publicly available. The dataset is collected using an ASUS ZenPhone (Model Z00UD) rear camera with 31 participants from 7 different countries, of which 22 were males and 9 were females. They extract 17,308 frames from videos which are shot in 4 different cars. The frames are divided into 10 different classes (illustrated in Fig. 4).

To tackle with the unavailability of quality dataset for studying distracted driving, Billah et al. [46] used a camera mounted on the front windshield facing the driver inside the vehicle to develop the EEE BUET Distracted Driving (EBDD) Video Database [47]. They took video clips of a number of drivers on city roads and university campus in Dhaka, Bangladesh, so that the dataset is diverse in landscape, illumination, type of vehicle and road condition (smooth or bumpy), as well as of the age or experience of the drivers. There are four types of distracted driving in the dataset, including talking on cell phone, texting on cell phone, eating, and operating cabin equipments.

3.5. Driver behaviors

Recently, a number of datasets for driver behavior/action recognition emerged, which enable us to go deep into analyzing distractive activities of drivers. Yang et al. [48] present a dataset named FDU-Drivers, which includes 20000 driving images of 100 different drivers under various real driving environments. Saad et al. [49] introduce a low lighting support dataset (shorted as LoLi in the table) for driver distraction recognition. Regarding driver distraction detection as an "open set recognition" problem instead of classifying a set of pre-defined anomalous actions, Köpüklü et al. introduce the video-based Driver Anomaly Detection (DAD) dataset [50].

Martin et al. [51] introduce a multi-view multi-modal dataset Drive&Act for fine-grained driver behavior recognition, in which up to 83 classes of driver secondary activities are labeled hierarchically. Ortega et al. create the Driver Monitoring Dataset (DMD) [52] from 3 cameras with 3 streams per camera. Different from the Drive&Act dataset, this dataset aims at the context of automated driving SAE L2-3

Fig. 4. Ten classes of driver postures from AUC Distracted Driver dataset [42].

for driver monitoring systems. There is another multi-modal multi-view dataset by Jegham et al. named MDAD [53] for in-vehicle driver action recognition which includes 16 kinds of actions. Taking nighttime into consideration, they upgrade the dataset as 3MDAD (Multiview, Multimodal and Multispectral Driver Action Dataset) [54].

In this section, we have introduced some publicly available datasets for vision-based driver distraction analysis, ranging from driver face, head pose and eye gaze, and driver hands, to driver postures and driver behavior. We summarize the datasets mentioned above for driver behavior distraction analysis in Table 1. Next we will start with exploring how cues from local parts of a driver, the head and the hands for example, contribute to driver distraction detection.

4. Cues for detecting driver behavior distraction

Detecting and analyzing different parts of a driver deliver helpful information to understand the behavior and state of the driver. For example, from the head pose of a driver, we can infer roughly where the driver is looking. And detecting a driver's hands could shed light on what the driver is doing. Thus, in this section, we will analyze cues from the head, cues from hands, and "more cues" to see how they facilitate detecting driver behavior distraction.

4.1. Cues from the head

When it comes to the face, eyes, and head pose of a driver, surely we can take it for granted that they are different cues. However, due to the close relationship among driver face detection, head pose estimation, and eye state analysis, sometimes it is not easy and does no good to tell them apart. Therefore, try as we might to discuss the cues from the driver head separately, sometimes it is not a surprise to see that they are mixed together.

Face detection and tracking are basic but crucial steps in computer vision systems for driver monitoring. Yuen et al. [55] propose a system based on AlexNet and stacked hourglass network for face detection, landmark localization, and landmark-based head pose estimation. The system consists of three modules: Face Detector Module, Landmark Module, and Head Pose Estimator Module. Subsequently, they [56] present a system for Facial Landmark Localization and Occlusion Estimation using Occluded Stacked Hourglass , which could output 68 heat maps to derive landmark location, occlusion levels and a refined face detection score. To track face for driver alertness monitoring, Nuevo et al. propose a system based on Active Appearance Models [57], and another system based on simultaneous modeling and tracking [58].

In [59], Hu et al. switch the emitter pattern of a Kinect sensor to obtain both infrared images and depth information. The former is used for driver facial features detection and tracking, while the latter is employed for face region detection and face orientation estimation. Jo et al. [60] propose a driver monitoring system to detect both driver distraction and fatigue. Inspired by PERCLOS for measuring driver fatigue level, the authors define PERLOOK as the percentage of time spent not looking ahead during a certain time interval to measure driver distraction level. Face detection and head orientation detection are applied first, and the PERLOOK is calculated when the yaw angle of the driver is out of [−15, 15] degree. Sigari et al. [61] propose a fuzzy expert system to detect driver distraction and fatigue both based on the features from face and eye regions. The head rotation extracted by template matching and the eye closure rate extracted by horizontal projection are used for distraction detection.

Eye detection and eye state analysis are critical for driver fatigue monitoring. Unlike fatigue detection, it is usually not sufficient to determine whether the driver is distracted or not on the basis of eye features alone. However, they are still of importance for distraction detection, and will be more useful when combined with head pose information. Sabet et al. [62] use face orientation derived from eye position to detect driver visual distraction. Braunagel et al. [63] propose an approach for driver-activity recognition base on head- and eye-tracking without utilizing gaze estimation. Next we will see how cue from eyes are used for estimating driver visual attention.

Rezaei and Klette [64] propose a system to monitor driver awareness under five different scenarios. Herein three cascade classifiers for detecting face, open-eye, and close-eye work in parallel. In the case that nothing but only one eye is detected, and if this last for more than one second, the system will raise an alarm for the sign of distraction. After that, they propose a framework integrating the adaptive detection module and the tracking module for eye detection and tracking under challenging lighting conditions [65]. It turns out to be fast and effective for driver monitoring as the two modules can boost each other recursively. In their following work [66], the authors propose a method for eye gaze estimation under various lighting conditions and in low resolution images. The region of interest is generated and narrowed, then eye pupil and eye corner are detected, afterwards the distances related to pupil in both horizontal and vertical direction are calculated for eye gaze estimation.

Ahlstrom et al. [67] propose data refinement algorithms for naturalistic eye-tracking data. Eye-off-road estimation is performed to illustrate the effectiveness of the quality handling and signal enhancement techniques. Realizing that higher percentages of eyes-off-road times are associated with an increased likelihood of the occurrence of safety critical events, Ahlstrom et al. [68] develop an eye-tracking-based distraction warning system called AttenD, and investigate its effects on driver glance behavior. The authors conclude that slight behavioral changes were observed and that they were more apparent on the local level than on the global level. Ahlstrom and Dukic [69] compare the performance of a one-camera system with a three-camera one to investigate the potential of a single perspective for driver state monitoring. The results show that both systems work well when the driver is looking straight ahead, but the multi-camera system achieves better results when the driver is looking at a peripheral gaze target. Such finding can also be of great value for driver head pose estimation.

Head pose is a discriminative cue for driver visual attention estimation by nature. Bergasa et al. [70] fuse head movements and facial

Table 1
Publicly available datasets for driver behavior distraction analysis.

Dataset	Year	Target	Scenarios	#Cameras	#Subjects(F/M)	Size
DrivFace [32]	2016	Driver face, head pose	Naturalistic	1	4 (2/2)	606 images
VIVA-Face [33]	2016	Driver face, head pose	Naturalistic	1	N/A	39 video sequences
DriveAHead [35]	2017	Driver head pose	Naturalistic	1	20	21 video sequences
Pandora [36]	2017	Driver head pose, upper body	simulated	1	22 (12/10)	110 video sequences
DD-Pose [37]	2019	Driver head pose	Naturalistic	2	27 (6/21)	2 × 330k images
AutoPOSE [38]	2020	Driver head pose, eye gaze	Simulated	2	21 (10/11)	21 sequences
DG-Unicamp [39]	2019	Driver gaze zone	Naturalistic	1	45 (10/35)	1M images
VIVA-Hands [40]	2015	Driver hands	Naturalistic	1	N/A	11k images
Turms [41]	2018	driver hands	Naturalistic	1	7 (2/5)	14k frames
StateFarm[a] [43]	2016	Driver postures	Naturalistic	1	N/A	22,424 images
AUC[b] [42]	2019	Driver postures	Naturalistic	1	44 (15/29)	14,478 frames
EBDD [47]	2018	Driver postures	Naturalistic	1	13	2 × 312 video segments
FDUDrivers [48]	2020	Driver attention	Naturalistic	1	100	20,000 images
LoLi[c] [49]	2020	Driver distraction	Naturalistic	1	70	52,350 frames
DAD [50]	2021	Driver anomaly	Simulated	2	31	783 min videos
Drive&Act [51]	2019	Driver behavior	Naturalistic	6	15 (4/11)	9.6M frames
DMD [52]	2020	Driver behavior	Naturalistic, simulated	3	37 (10/27)	41h videos
MDAD [53]	2019	Driver action	Naturalistic	2	50 (12/38)	2 × 2 × 800 video sequences
3MDAD [54]	2020	Driver action	Naturalistic	2	50 (12/38), 19 (8/11)[d]	444 K + 130 K images

[a]The State Farm Distracted Driver dataset.
[b]The AUC Distracted Driver dataset.
[c]The Low Lighting Support dataset.
[d]For daytime and nighttime separately.

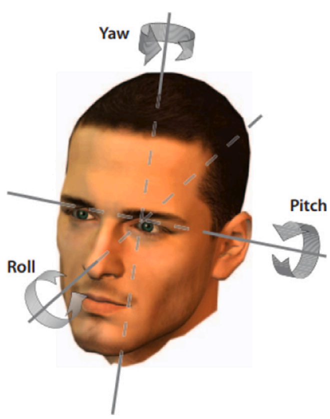


Fig. 5. Pitch, yaw and roll for head pose estimation [19].

features for driver inattention monitoring, in which they detect visual distraction based on head pose. To address the challenges such as large spatial head movements, occlusion by hand, and various lighting conditions due to single perspective, Martin et al. [71] propose a system for monitoring driver head dynamic from multi-perspective by three cameras. The head pose is estimation is based on facial features and their respective geometry. And the result of the current frame is used to decide which camera perspective should be selected in the next frame. To put the work one step further, Tawari et al. [72] propose a distributed camera framework named CoHMEt. Each camera perspective works independently, and for each camera perspective, the facial features are detected and tracked to build a geometric model for head pose estimation. These outputs along with the temporal dynamics are used during the perspective selection procedure to provide the final head pose estimation.

We can utilize position and orientation to model driver head in six degrees of freedom in reality, while the researchers tend to use one or more of the Euler angles (**yaw, pitch, and roll**, illustrated in Fig. 5) for driver head pose estimation. Murphy-Chutorian and Trivedi [73] present a system comprised of three interconnected modules, which are aimed separately at driver head detection, head pose estimation, as well as head position and orientation tracking. The initial pose estimation module uses Localized Gradient Orientation histograms with Support Vector Regressors to estimate driver head pose [74], while the tracking module is a model-based system integrated with appearance-based particle filter to track the pitch, yaw, and roll of driver head [75]. To estimate driver head yaw angle, Narayanan et al. [76] propose a generic geometric model which can be customized into 12 different models, whereas Diaz-Chito et al. [32] simply use only three representative facial keypoints, namely the center of the eyes and the nose tip. As for estimating driver head yaw and pitch angles, Diaz-Chito et al. [77] exploit Histogram of Oriented Gradients, a manifold embedding projection, and a continuous regression. And Alioua et al. [78] propose an appearance-based discrete head pose estimation. They construct a feature vector by fusing four head descriptors (SF, HOG, Haar, and SURF), and use two SVMs (Pitch-SVM and Yaw-SVM) to classify head poses.

The researches mentioned above estimate driver head pose estimation mostly from the images taken by ordinary digital cameras. We also notice that there are some studies estimating head pose from stereo images, depth images, or intensity images, and even virtual data.

Hoffken et al. [79] propose a novel system for driver head separation, pose estimation (yaw and pitch), and pose tracking using only 3D information from a stereo camera. Borghi et al. [80] propose a system using Recurrent Neural Network for driver head pose estimation based on depth images only. The system could be implemented on embedded boards with real time performance. In [81], Borghi et al. propose the POSEidon+ framework to estimate driver head and shoulder pose based on depth images only. Inspired by the PointNet++ framework, Hu et al. [82] propose a point-cloud based approach to estimate driver head pose. They use point-clouds to represent depth camera recording, and employ the set abstraction layers in PointNet++ to extract features, creating a 6D vector for head pose estimation.

Barros et al. [83] propose a fusion pipeline for extreme head pose estimation from intensity images. The pipeline consists of two independent schemes: tracking scheme using a geometric model with keypoints, and detection scheme using a synthetic head mesh with facial landmarks. These two head pose estimators are later fused into a dedicated Kalman Filter. Liu et al. [84] propose a lightweight framework called recurrent multi-task thin net (RM-ThinNet) to estimate driver pose, which consists of three parts: 9 body joints, 5 face land marks, and 3 head Euler angles. To boost the performance, they construct a virtual driving scene model, synthesize labeled virtual driver dataset, and employ the transfer metric learning method.

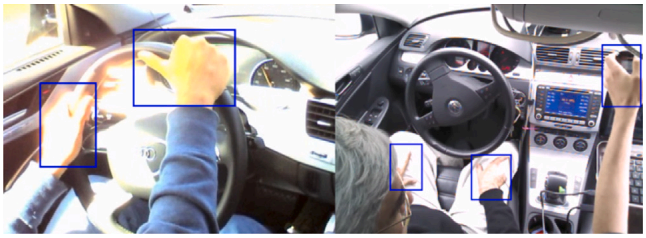


Fig. 6. Some examples of hand detection results [92].

4.2. Cues from hands

Similar to face detection, we have to deal with challenges such as illumination changes, occlusions, complex backgrounds, and varying viewpoints for hand detection. However, hands are deformable and smaller, which makes the detection a tougher task.

There are approaches taking advantage of different features for **generic hand detection**. Mittal et al. [85] present a two-stage approach for hand detection. Hand bounding boxes proposed by three complementary detectors independently are fed into a classifier to computer a confidence score. Zou et al. [86] present a mixture of multiscale deformable part models for hand detection against complex background. HOG feature pyramid is computed, allowing for taking advantage of the multi-scale features. The dataset is split into three groups here, and for each group there is a corresponding model, thus the mixture model trained via latent SVM is composed of three components. Each component includes a coarse root filter and three fine part filters. Deng et al. [87] present a system based on Faster R-CNN with a derotation layer after the ROI pooling for joint hand detection and rotation estimation. The system includes five parts: a shared network, a region proposal network, a rotation network, a derotation layer, and the detection network. Wang et al. [88] combine multi-feature based hand proposal generation with cascaded Convolutional Neural Network classification for hand detection.

We see that **contextual cues** are of critical importance for promoting driver hand detection. Region proposal and multi-scale features contribute greatly to the detection. Zhou et al. [89] present a hierarchical context-aware approach aggregating a general hand detector with the context detector for driver hand detection. Xia et al. [90] adopt Fast R-CNN framework for driver hand detection, using AdaBoost detector with Aggregated Channel Features (ACF) as hand region proposal. Yan et al. [91] incorporate multiscale features into Fast R-CNN for hand detection.

Significantly, Le et al. [92] present Multiple Scale Faster Region-based Convolutional Neural Network (MS-FRCNN) for hand detection in vehicles. And Fig. 6 show some examples of the hand detection results. They modify the multiscale deep feature extraction in both the Region Proposal Network and the detection network. In the further study, Le et al. [93] present Multiple scale Region-based Fully Convolutional Neural Networks (MS-RFCN) to detect hand both in vehicles and in the wild.

Besides, **hand tracking** depends on hand detection, and can improve hand detection performance in turn. Ohn-Bar and Trivedi [94] conduct driver hand detection in five regions of interest from color and depth images. With this information, they extract features and use a linear SVM to classify hands on wheel region particularly. Ohn-bar [95] et al. present two different frameworks for vision-based hand activity analysis, using motion cues and appearance cues separately. Ohn-bar and Trivedi [96] detect and track the driver hands, then use trajectory features and temporal models for activity classification, activity prediction, and abnormal event detection.

Rangesh et al. [97] present multiple hands tracking framework, which integrates hand detection with tracking using appearance and motion cues. A bipartite matching algorithm is used for data association, assigning tracks to the corresponding detections. To disambiguate hand tracks, an SVM classier with a linear kernel using HOG or CNN features is employed to classify the left hand versus the right hand. To address driver hand tracking problem, Rangesh et al. [98] present a combined tracking-detection framework with bipartite matching for data association. As for the challenge caused by self-occlusion, which is frequently seen during turn, a module encoding hand motion patterns is learned offline to handle this.

In addition to detection and tracking, there is some useful **semantic information** can be derived from driver hands. Some researchers commit to driver hand gestures recognition, hand grasp analysis, or how they interact with objects, especially the smart phones. while some are interested in hand parsing or segmentation.

Deo et al. [99] study the usage of Hidden Markov Models for in-vehicle hand gesture recognition. HOG and CNN features is served as shape descriptors, while dimensionality reduction and data augmentation are used to reduce overfitting. They also present a CNN-HMM hybrid framework, which outperforms the method using CNN just as feature extractor. Siddharth et al. [100] exploit YOLO to generate hand region proposals, then use a pixel-level mask to refine the proposals, and employ SVM to classify driver hand grasp state. Le et al. [101] present a driver behavior situational awareness system (DB-SAW) using Grammar-aware Driver Parsing (GDP) approach to segment parts of the driver, analyze seat belt usage and driver hands position. Weyers et al. [102] present a system using LSTM for driver hand activity recognition from images taken by a time-of-flight camera. The system consists of body key points localization, hand patches extraction, and activity recognition.

Rangesh and Trivedi [103] present a system named HandyNet to monitor a driver and assess his/her readiness from depth and RGB images. The system is based on Mask R-CNN for driver hand detection and segmentation. With this, they localize the hands and calculate the distance between driver and the steering wheel. Besides, the system is able to identify objects held by the driver. Yuen and Trivedi [104] present a system to locate in-vehicle occupant hands. They use a modified OpenPose to estimate joint and affinity heatmaps for driver and passenger's wrist and elbows. The part locations and association between matching parts are learnt simultaneously using Part Affinity Fields (PAFs).

Among various secondary tasks associated with driver hands while driving, there is a typical one called **cellphone usage**. Zhang et at. [105] apply Hidden Conditional Random Fields (HCRF) model with hand and mouth features to detect driver hand-held cellphone usage. Berri et al. [106] use SVM with various kernels to detect driver cellphone usage and illustrate that SVM with Polynomial kernel achieves best accuracy. Artan et al. [107] use a linear SVM classifier to detect driver cellphone usage from NIR images. Face regions are detected by the deformable part model (DPM) within the front windshield. The SVM classifier using several local aggregation descriptors is trained on full face images and half face images. Seshadri et al. [108] present a vision-based approach to detect driver cellphone usage on Strategic Highway Research Program (SHRP2) face view videos. the Supervised Descent Method (SDM) based facial landmark tracking algorithm is used to track facial landmarks and extract region of interest crops. Features are extracted from these crops, and represented by raw pixels or Histogram of Oriented Gradients (HOG). Different classifiers including Real Adaboost, Support Vector Machines (SVMs), and Random Forest (RF) are chosen to determine if the driver is using a cellphone.

Recently, deep learning methods are favored by the researchers to detect the use of mobile phone. Le et al. [109] present Multiple Scale Faster-RCNN approach to detect if a driver is using a cellphone, and to count the number of hands on the wheel. Regions of interest including hands, face, and steering wheel are extracted, then the geometric information is used for decision making. Torres et al. [110] use CNN to detect whether a driver is texting or talking on the phone. Celaya-Padilla et al. [111] fine-tune a pre-trained Inception-v3 to detect driver distraction in the form of texting while driving. The data is collected using a wide-angle camera mounted on the roof inside the car.

There is an interesting study by Watkins et al. [112], they provide a noel perspective to study the driver cellphone usage problem. Most of the researches focus on how the driving performance is influenced by texting. However, they look at it the other way, measuring how the texting dynamic is influenced by driving.

4.3. More cues

"More cues" are given two meanings here. On the one hand, there are some more cues we could make use of for the problem, for example, the foot of the driver. On the other hand, we could use more cues rather than a single cue to address the problem.

There are a few works involving driver foot activity analysis. Tran et al. design a Hidden Markov Model (HMM) model for driver foot behavior modeling and prediction, which use the output of optical flow based foot tracking in addition to embedded pedal sensor information [113,114]. Examining foot placement from naturalistic driving data, Wu et al. [115] use random forest tree to predict the pedal application, and multinomial logistic regression estimate foot placement types. In their following study, they analyze the various foot to pedal trajectories using functional principal components analysis (FPCA), so as to discover the unique patterns of foot movements that might be early indicators of pedal errors [116]. Rangesh and Trivedi [117] take driver foot activity classification as a case of image classification tasks where the output classes are tied to relative spatial locations of salient objects in the scene. They propose to augment the standard cross-entropy (classification) loss with a domain dependent Forced Spatial Attention (FSA) loss for the task and demonstrate the benefits of the loss function.

There are several studies making good use of multi-cues from the driver for driver inattention monitoring. Mbouna et al. [118] propose a system to monitor driver alertness, using visual features named eye index, pupil activity, and head pose. They construct a generic 3D head model and calculate three Euler angles for head pose estimation. The results show that combining head pose with eye information outperforms using eye information only. BTW, it is not uncommon to combine head pose with eye information, and we will see more examples in the coming Section 5.

Martin et al. [119] propose a vision-based analysis framework for in-vehicle activities recognition, in which two cameras are used to look at driver's head and hand separately. Feature including head pose, eye opening and hand locations are derived from head and hand cues. Similarly, head, eye, and hand cues are fused to classify driver activity into instrument cluster region, gear region, or wheel region activity [120]. Ohn-Bar et al. [121] utilize the information from the driver, the ego-vehicle, and the environment for on-road maneuver analysis, where the cues extracted from the driver include head, hand, and foot. Recently, face and hands cues are used to train the CNNs for distracted driver posture classification in [122], and the authors bring in skin cue for the training in their follow-up research [42].

5. Visual distraction detection

Previously we have discussed some work on driver visual attention, and head pose estimation in particular, in Section 4. Here, we will take one step further on this topic. Not only do we look inside the vehicle cabin and focus on the driver, but also look outside the car and pay attention to the driving context. More specifically, we will derive driver visual attention from gaze estimation.

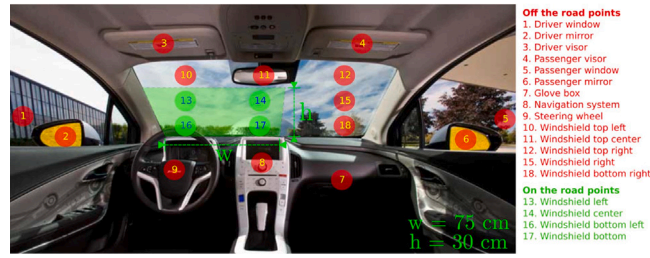


Fig. 7. The location and partition of 18 target points in [136].

5.1. From gaze direction to visual distraction

We roughly divide driver gaze estimation into accurate or coarse gaze direction estimation. In this subsection, we pay attention to associate driver visual distraction with gaze, regardless of accurate or coarse gaze direction. While in the next subsection, we will discuss driver gaze zone estimation, which is corresponding to coarse gaze estimation.

Smith et al. [123] reconstruct the 3D gaze direction with a single camera. Based on head rotation and eye blink information, they use Finite State Automata (FSM) to model driver visual attention. Alam and Hoque [124] use head and eye movements for gaze analysis to detect driver distraction. De Castro et al. [125] explore the usage of eye gaze angles as the only feature for driver visual distraction detection and prove its efficacy. In their following study [126], they use the combination of facial landmarks, eye gaze, and action units for driver visual distraction detection. Dua et al. [127] present a system called AutoRate, using generic features and specific facial features such as head pose, eye closure and eye gaze to rate driver attention level.

Jha and Busso [128] propose a probabilistic approach based on head position and orientation to estimate driver gaze coarsely. Gaussian process regression is used to create a probabilistic map of driver visual attention, from which confidence regions could be projected into the windshield to indicate the driver gaze directions. In their further research [129], they convert the continuous gaze angles into intervals, and use Convolutional Neural Networks to infer a generic probabilistic distribution of the driver visual attention. With this, they formulate driver gaze estimation as a dense classification problem rather than a regression problem.

5.2. Gaze zone estimation

Gaze zone/region estimation refers to coarse/rough gaze direction in some studies. Driver gaze is an important cue to recognize driver visual distraction. Making use of the cabin structure and taking the environment into account, we can correlate the driver's gaze direction with a certain region. It maps the continuous gaze angles into discrete gaze zones, which are more in line with our habits and provide more semantic information. Table 2 summarizes the related researches on driver gaze zone estimation. Next, we are going to introduce the researches on driver gaze zone estimation.

The number of gaze zones considered ranges from 2 to 18. There are up to 18 gaze zones considered in [130,131]. A simple way of gaze zone partition can be seen in [132,133], where the gaze direction is mapped into 9 regions in the form of a 3 by 3 square grid. However, it is short of semantic information. As a contrast, there are fewer but more meaningful gaze zones in [134,135], where the 5 gaze zones are explicitly correlated with the interior areas of vehicle cabin. The eyes off the road detection [136,137] involve 2 gaze zones only: on/off the road. What we have to say is that, Vicente et al. [136] mark 18 target locations (seen in Fig. 7) and present a paradigm to simplify the detection as a binary classification problem, transferring from the side in details to the other side in brief.

Table 2
Summary of researches on driver Gaze zone estimation.

Reference	Objective	Scenarios	#Cameras	#Zones	Cues	Methods
Wang et al. [130]	Head pose-free eye gaze prediction	Naturalistic	1	18	Facial landmarks, eye features	POSIT, Sparse encoding, Random forest
Lee et al. [131]	Gaze zone estimation	Naturalistic	1	18	Head orientation	SVM
Ji and Yang [132]	Driver vigilance monitoring	Simulated	2	9	Eyelid movement, face orientation, gaze	computer vision algorithms
Cheng and Xu [133]	Gaze zone estimation	Simulated	2	9	Face, eyes, mouth	3D triangle model, Neural networks
Lundgren et al. [134]	Gaze zone estimation, a probabilistic framework	Naturalistic	1	5	Head pose, gaze direction, eye closure	Bayesian filtering, Gaussian processes
Kim et al. [135]	Lightweight driver status recognition	Simulated	1	5	Face direction, eye closure, mouth opening	Multi-task mobilenets
Vicente et al. [136]	Gaze tracking and eyes off the road detection	Naturalistic	1	18 or 2	Facial landmarks, head pose	3D geometric reasoning
Zeng et al. [137]	Driver distraction detection and identity recognition	Simulated	1	2	Head motion, eye states	machine learning classifiers
Choi et al. [138]	Gaze zone estimation	Naturalistic	1	9	Face	AlexNet
Fu et al. [139]	Gaze zone estimation, automatic calibration	Naturalistic	1	12	Facial features, head orientation	self-learning, particle filtering
Vasli et al. [140]	Gaze zone estimation	Naturalistic	1	6	Head and eye cues	SVM
Vora et al. [141]	Generalizing gaze zone estimation	Naturalistic	1	7	Face	AlexNet, VGG-16
Vora et al. [142]	Generalized gaze zone estimation	Naturalistic	1	7	Face	AlexNet, VGG-16, ResNet-50, SqueezeNet
Hernandez et al. [143]	IVISs induced distraction monitoring	Simulated	2	10	Face pose, eye direction	3D model, POSIT
Jimenez et al. [144]	IVISs induced distraction monitoring	Simulated	2	11	Face pose, eye direction	Based on 3D face model
Tawari & Trivedi [145]	Gaze zone estimation	Naturalistic	2	8 or 7	Head pose	Random forest
Tawari et al. [146]	Gaze zone estimation	Naturalistic	2	6	Head pose, gaze angles	Random forest
Fridmand et al. [147]	Gaze zone estimation	Naturalistic	1	6	Head pose	Random forest
Fridman et al. [148]	Gaze zone estimation, head pose + eye pose	Naturalistic	1	6	Head pose, eye pose	Random forest

Fig. 8. Illustration of 6 driver gaze zones considered in [142].

It is likely that 6 to 10 gaze zones are more practical and suitable as a tradeoff. Choi et al. [138] propose a system using Convolutional Neural Networks to categorize driver gaze into one of the 9 gaze zones (including eye blink). A Haar feature face detector and a MOSSE face tracker are combined for face detection, then AlexNet is used for gaze zone estimation. Fu et al. [139] propose a system for head pose and eye gaze estimation with automatic calibration. Face detection, self-learning, and particle filtering are combined to categorize the driver gaze into one of the 12 gaze zones.

Vasli et al. [140] propose a system to classify driver gaze into 6 gaze zones, combining geometric and learning based methods. Geometric method takes advantage of the physical constraints of the car. The features from both methods are inputted to a SVM classifier. Vora et al. [141] use Convolutional Neural Networks to classify driver gaze into 6 gaze zones (illustrated in Fig. 8) or eye closed. Images are pre-processed to create different region crops of interest. The images crops are fed into the pre-trained AlexNet and VGG-16. To boost the generalization of the system, they [142] consider one more region crops, explore two more network architectures, and undertake ablation experiments on a large naturalistic driving dataset. Additionally, the system is tested on the Columbia Gaze dataset.

5.3. Head and eye cues for gaze estimation

Next we will discuss how the **head pose and eye cue** contribute to driver gaze estimation. There is a scheme to calculate gaze direction using eye cue in previous section [66]. And Wang et al. [130] map eye image feature into gaze zone, without estimating head pose. But the fact is, it is rare to estimate driver gaze relying on eye cue only. In contrast, head cue is frequently used for driver gaze direction or gaze zone estimation, alone or together with eye cue. Lundgren et al. [134] use different sets of head and eye information for driver gaze zone estimation. Jha and Busso [149] use head pose as a coarse estimation of the driver visual attention to explore the relationship between head pose and gaze. They propose linear regression models to predict driver gaze location based on the head position and orientation. And the ground truth estimations are derived from AprilTags attached to a headband. The results show the efficacy of using head pose to infer visual attention, and the limitation as well due to eye movements.

There are different research teams investigating the gain of adding eye cue to head pose for gaze estimation in a two-stage manner. In [150] driver gaze focalization is derived from face pose, while in [143,144], gaze zone estimation is computed from face pose and eye direction. Tawari and Trivedi [145] explore the efficacy of the driver head pose dynamics for coarse gaze direction estimation in a distributed camera system, addressing occlusions from large head movements. Later on, Tawari et al. [146] use both head and eye cues for driver gaze zone estimation. Fridmand et al. [147] propose a system to explore how well head pose estimation can do to predict driver gaze region without relying on eye movement. In the further study [148], Fridman et al. take eye pose into account in addition to head pose. This is to see the contribution of eye pose information, and the variation of improvement under different gaze strategies. The results reveal the effectiveness of adding eye pose for driver gaze estimation, especially in the "lizard" mode, i.e. when the eyes move only but the head keeps still.

In this subsection, we have discussed the contributions of head and eye cues for driver gaze estimation. In short, despite that head pose is more robust and reliable than eye state for visual distraction detection, the combination of them both is likely to work better. Head pose is employed to determine the gaze direction coarsely, whereas in some cases eye information is needed to tell the subtle differences.

6. Manual distraction detection

Similar to Section 5, this section makes a further effort to discuss the topic aforementioned in Section 4. Looking back upon Section 4, we know that hands off the wheel possibly points to the manual distraction of the driver. And we introduce driver cellphone usage as a typical distraction involving hands in particular. In this section, we will see more types of manual distraction related to driver hands, such as eating, drinking, smoking, or operating the IVISs. But above all, we are going to detect driver manual distraction from a global view instead of a local view.

The researches on driver posture recognition help to detect distracted driver, and identify the cause of distraction as well. Table 3 summarizes the related researches on driver manual distraction analysis. Many of the researches are evaluated on the State Farm Distracted Driver dataset (shorted as StateFarm in the table) [151–154], the AUC Distracted Driver dataset (shorted as AUC in the table) [42,122,155,156], or both of them [157,158], while the others are experimented on the custom datasets [159–161]. For a better understanding and comparison, next we are about to introduce the researches mainly according to the methods used.

6.1. Prevailing methods

The **deep learning** methods are extremely popular for driver manual distraction recognition whereas only few studies employ the traditional machine learning methods such as SVM [162], random forest [176], decision tree (DT) [160]. The results and discussions of [162] provide insights into such situation. The authors compare deep learning method with traditional machine learning method. Three pre-trained Convolutional Neural Network (CNN) models, namely AlexNet, VGG-16, and ResNet-152, are fine-tuned, and all of them outperform the SVM classifier with handcrafted features.

Among various deep learning methods, here **Convolutional Neural Networks (CNNs)** are favored by the researchers. Bailing Zhang led a research group making the early attempts to adopt CNN for driver activity recognition on the SEU dataset. Yan et al. [163] use Convolutional Neural Networks to recognize 6 states of the driver's eyes, mouth, and ear. They verify the proposed approach on a dataset involving normal driving, falling asleep, eating, and answering the phone. Yan et al. [164] present a system using CNN to recognize driver postures based on hand position information. There are 4 pre-defined driver postures, including normal driving, operating the shift gear, eating or smoking, and answering the phone. The proposed approach is further tested on another two datasets involving poor illumination and different road conditions [165]. Yan et al. [166] use Gaussian Mixture Model (GMM) to extract skin-like regions, which are fed into R*CNN as the secondary region so as to identify driver actions.

The **network architectures** including AlexNet, VGG, GoogLeNet, ResNet, and Inception are widely used. It is common to see one or more of them are adopted to recognize driver behaviors. Koesdwiady et al. [167] present an end-to-end system to recognize driver distraction based on pre-trained VGG-19. Masood et al. [152] use VGG-16 and VGG-19 for driver distraction recognition. The results show that the VGG-16 pretrained on ImageNet dataset yields better accuracy. Baheti et al. [155] use a modified VGG-16 with dropout, regularization, and bath normalization to identify driver distraction. Arefin et al. [156] use a modified AlexNet aggregating with HOG features to create a small size model for driver distraction recognition. Cengil and Cinar [168] use Laplacian of Gaussian filter to preprocess the images and GooLeNet to extract features for driver distraction recognition.

A few researchers are committed to exploring and comparing several network architectures. Tran et al. [159] develop an embedded computer system for driver distraction detection and alert. A dataset is collected using an assisted-driving tested, and four CNN models including VGG-16, AlexNet, GoogLeNet, and ResNet are implemented and compared for the detection system. VGG-16 runs faster with lower accuracy, ResNet has higher accuracy with slower speed, GoogLeNet achieves the best performance in terms of balancing the accuracy and speed. Ou et al. [161] present a dual mode system which can detect driver distraction by day and at night, using RGB images and Near Infrared (NIR) images as input separately. The results show that ResNet outperforms SqueezeNet on both modes in terms of accuracy, but it is at the cost of speed. Kapoor et al. [169] present a real-time resource-efficient system for driver distraction detection using Software Defined Cockpit. Four pre-trained CNN models including MobileNetV1, MobileNetV2, Inception-v3, and VGG-16 are fine-tuned on the State Farm Distracted dataset, then the one with relatively better performance, MobileNetV1, is chosen and retrained on a custom dataset to boost the performance.

6.2. Optimizing techniques

Various techniques have been exploring to improve the performance of driver manual distraction recognition. Given the size and variability of the dataset, **data augmentation** can be used to improve generalization and accuracy. Cronje and Engelbrecht [151] use a modified Darknet to recognize driver distraction. Class based data augmentation is applied during the training period to reduce overfitting. Ou and Karray [177] augment the collected data with Generative Adversarial Networks. They demonstrate that the augmented images of drivers in different driving scenarios improve the accuracy for driver distraction recognition significantly.

The **transfer learning** is frequently used to pre-train the deep learning models [154,170–172]. The size of the data and performance improvements account for the strategy. Oliveira and Farias [170] conduct comparative study using VGG-19, Inception-v3, ResNet-152 and DenseNet-161 with three transfer learning schemes for driver distraction detection. The experiments illustrate that DenseNet-161 with end-to-end transfer learning achieves the best accuracy. Varaich and Khalid [154] explore the usage of Inception-v3 and Xception with random initialization or transfer learning initialization to recognize driver distraction. The ResNet-50 pre-trained on ImageNet is fine-tuned on two datasets with different training strategies [171]. Taking advantage of deep learning and fuzzy logic, Ou et al. [172] present a driver inattention monitoring system composed of head pose estimation, distraction recognition, and fuzzy inference. As with [171], The ResNet-50 pre-trained on ImageNet is fine-tuned to sever as the distraction recognition module.

Some studies illustrate the effectiveness of **ensemble learning** [42,122,173]. Abouelnaga et al. [122] present a genetically-weighted ensemble of Convolutional Neural Networks (CNNs) for distracted driver posture classification. AlexNet and Inception-v3 are trained on raw images, face images, hands images, and "face+hands" images to investigate the effect of different elements. In their following study [42], the dataset is extended, one more visual element is added, and another two CNN architectures (ResNet and VGG-16) are ensembled. Chawan et al. [173] use a small CNN along with VGG-16, VGG-19, and Inception-v3 to recognize driver distraction. Compared with a single CNN model, the ensemble method by averaging the results of different CNN models decreases the log loss.

More broadly, different network architectures can be mixed and exploited in some more way [157,174]. Huang et al. [174] present a hybrid CNN framework combining ResNet50, Inception-v3, and Xception in parallel for driver distraction recognition. Munif Alotaibi and Bandar Alotaibi [157] integrate one ResNet block and two LSTM layers with the Inception module to detect driver distraction.

Table 3
Summary of researches on driver manual distraction analysis.

Reference	Objective	Dataset	#Classes	Cues	Methods
Hssayeni et al. [162]	Driver distraction detection, comparing SVM with CNNs	unspecified	7	Hands	SVM, CNNs (AlexNet, VGG-16, ResNet-152)
Yan et al. [163]	Driver state classification	SEU	4	Eyes, mouth, ear	CNN
Yan et al. [164]	Driving posture recognition	SEU	4	Hands	CNN
Yan et al. [165]	Driving posture recognition	SEU	4	Hands	CNN
Yan et al. [166]	Driver behavior recognition	SEU	6	Hands, skin	R*CNN
Koesdwiady et al. [167]	Driver distraction recognition	Custom	4	Hands	VGG-19
Masood et al. [152]	Driver distraction detection	StateFarm	10	Hands, head	VGG-16, VGG-19
Baheti et al. [155]	Driver distraction detection	AUC	10	Hands, head	VGG-16
Arefin et al. [156]	Driver distraction detection	AUC	10	Hands, head	AlexNet
Cengil and Cinar [168]	Driver distraction detection	StateFarm	10	Hands, head	GoogLeNet
Tran et al. [159]	Driver distraction recognition, embedded platform	Custom	10	Hand, body movements	VGG-16, AlexNet, GoogleNet, ResNet
Ou et al. [161]	Driver distraction detection, day & night	Custom	4	Hands, head	SqueezeNet, ResNet
Kapoor et al. [169]	Driver distraction detection, Android smartphone	StateFarm, custom	10, 4	Hands, head	MobileNetV1 & V2, InceptionV3, VGG-16
Cronje and Engelbrecht [151]	Driver distraction detection, data augmentation	StateFarm	10	Head, hands	Darknet
Oliveira and Farias [170]	Cellphone usage detection, transfer learning	Part of StateFarm	5	Hands	VGG-19, Inception-v3, ResNet-152, DenseNet-161
Varaich and Khalid [154]	Driver distraction recognition	StateFarm	10	Hands, head	Inception-v3, Xception
Ou et al. [171]	Driver distraction recognition, transfer learning	Custom	3	Hands, head	ResNet-50
Ou et al. [172]	Driver inattention monitoring, fuzzy logic theory	Custom	3	Hands, head	ResNet-50
Abouelnaga et al. [122]	Driver posture recognition, ensemble of CNNs	AUC	10	Face, hands	Ensemble of CNNs (AlexNet, Inception-v3)
Eraqi et al. [42]	Driver distraction identification, ensemble of CNNs	AUC	10	Face, hands, skin	Ensemble of CNNs (AlexNet, Inception-v3, ResNet-50, VGG-16)
Chawan et al. [173]	Driver distraction detection	StateFarm,	10	Hands, head	Small CNN, VGG-16, VGG-19, Inception-v3
Huang et al. [174]	Driver distraction detection, hybrid CNN framework	StateFarm,	10	Hands, head	Hybrid CNNs (ResNet50, Inception-v3, Xception)
Munif Alotaibi and Bandar Alotaibi [157]	Driver distraction classification, modified Inception	StateFarm, AUC	10	Hands, head	Inception + ResNet + LSTM
Valeriano et al. [153]	Driver distraction recognition, video-based method	StateFarm, Kinetics	10	Appearance, motion	3D CNN
Moslemi et al. [158]	Driver distraction monitoring, temporal information	StateFarm, AUC, Kinetics	10	Hands, temporal info	3D CNN
Yan et al. [175]	Driving behavior recognition	SEU	8	Hands, skin	Hierarchical classification
Weyers et al. [160]	Driver state monitoring, hierarchical classification	Custom	10	Hands, object	Decision tree + CNN
Majdi et al. [176]	Driver distraction detection	StateFarm	10	Hands	CNN + Random forest

Moreover, we can take **temporal information** [153,157,158,175] into consideration and make use of hierarchical classification [160,175, 176] for boosting the performance. Valeriano et al. [153] explore the contribution of temporal information for driver distraction recognition by comparing video-based methods with a frame-based method. Their best proposed approach use the I3D architecture to combine RGB with optical flow, which achieves the classification accuracy of 96.67 Taking advantage of the temporal information for driver distraction recognition, Moslemi et al. [158] use 3D Neural Networks and optical flow on video clips to boost the performance.

Driver postures are inherently linked with multi-cues, thus we can take advantage of local view **hierarchically** for classification. Yan et al. [175] present a hierarchical classification system for video-based driver behavior recognition. Weyers et al. [160] incorporate a hierarchical label structure into CNN classifier to detect driver absence and distraction. Majdi et al. [176] present a cascaded classifier named Drive-net composed of a CNN and a random decision forest.

In this section, we have mainly discussed using deep learning methods to recognize driver manual distraction over recent years, including some optimization techniques of deep learning concerning models, algorithms, and training. For the implementation of deep learning on hardware accelerators, readers are encouraged to refer to several surveys [178–181]. However, we still hold the view that it is worthwhile to keep an eye on the traditional machine learning methods. In the next section, we will discuss topics including this.

7. General discussion and future directions

In this section, we are going to discuss the developments and future trends of driver distraction analysis for safe driving.

Traditional Methods vs Deep Learning: There are a few works using traditional machine learning methods to detect driver distraction [182–184], among which SVM, RF and DT are frequently used. From the above sections, we also see that there have been many more researchers tend to utilizing deep learning to address this problem, especially for the past five years. Checking the up-to-date articles, the observations are mostly in keep with what we have discussed above. The Convolutional Neural Network (CNN) [48,185,186] is very popular for vision-based driver distraction detection, where the network architectures such as VGG-16 [187,188] and ResNet [189,190] are frequently employed. There are some other findings that are likely to shed light on the future directions.

Jain et al. [191] explore a CapsNet-based method to detect driver distraction, while Lu et al. [192] present deformable and dilated Faster R-CNN (DD-RCNN), using the cues from motion-specific objects to classify driver action. Targeting at both speed and accuracy, Baheti et al. [193] present mobileVGG to achieve a computationally efficient

CNN while maintaining good accuracy, whereas Gumaei et al. [194] take advantage of two deep learning models along with the power of edge and cloud computing. In addition to network architecture and computing resource, we can investigate hybrid approaches to make use of various architectures or features, as illustrated in [195,196]. Besides, we believe that attention mechanism will be promising for driver distraction detection and driver action recognition [197–199].

Despite that it is obvious that deep learning methods help to improve the accuracy significantly, we cannot ignore the consequent issues such as computation complexity and interpretability. Thus, it is worthwhile to explore the potential of the traditional machine learning methods such as SVM, RF, DT, Bayesian Networks (BN), K-Nearest Neighbor (KNN), and rule-based method. Exploiting the complementary features of traditional methods and deep learning effectively, the combination of them both will be promising for driver distraction monitoring.

RGB Cameras vs Depth Sensors: Sensors are critical for vision-based approaches for the detection, as the systems need "eyes" just like we human beings, so that the computer vision or machine vision can be established. The most frequently used sensor discussed so far is camera. It is obviously that it has lots of advantages, including small, convenient, effective, and cost-saving. Sometime only a single camera can serve, sometimes more are in need to obtain different perspectives. In terms of the camera/sensor, the topics can include the position of the camera, the number of cameras, and the type of camera. Recently, there are emerging works proving the efficacy of Kinect for driver distraction detection [200–203]. With the usage of Kinect, the researchers can get the depth data in addition to the images [204,205].

Images vs Videos: Image data is extensively used for computer vision tasks such as detection and recognition. However, it is short of the important temporal information for tracking or prediction. As for vision-based driver behavior analysis, we have mentioned the use of temporal information for hand tracking in Section 4, and for manual distraction detection in Section 6. Learning from action recognition [206], we can model and analyze driver behavior dynamically [46, 207–209] With temporal features from the video data.

Paying attention to the temporal features for driver distraction analysis, we shall be aware of the dynamics of distraction. Studies show that non-specific eyeglance away from the forward roadway lasting for a certain time dramatically increases the risk of driving, as pointed out in the Introduction. From this aspect, it is not just about detecting the presence of a mobile phone close in one frame, it is not even about detecting mobile phones in a sequence of frames, distraction detection is about estimating when attention has been diverted away from what is relevant for the driving task too often or for too long. That is, driver distraction detection algorithms also need to deal with time-series data, not just focus on a single time instant.

Cabin vs Surroundings: The surroundings define what you need to attend to, and consequently when you are distracted from what you need to attend. Thus, this is a very important future research topic that must be addressed. Only when we have context-sensitive distraction detection algorithms will we reach a solution that is acceptable for the driver, with high sensitivity as well as high specificity (and few false alarms).

Unlike driver fatigue detection mainly focusing on the drivers themselves, we also need to pay attention to the driving environment for driver distraction analysis. As for driver fatigue detection, what we are interested in may be "who" and "what". However, driver distraction analysis should also take "where" and "when" into consideration. That is, we have to address the problem of situation awareness. For example, to detect whether a driver is under visual distraction or not, we have to consider questions such as how would you classify a gaze towards the road in a curve, where the road is outside the forward region? Or a gaze to the left or right when driving past an intersection? Because it is dangerous to call a driver distracted when he/she is monitoring relevant traffic.

Early in 2009, there were researches taking driving environment into consideration for driver distraction monitoring. Doshi and Trivedi [210] fuse driver gaze patterns and environmental motion saliency maps investigate the relationships between driver visual search and driver intentions. Fletcher and Zelinsky [211] make attempts to correlate driver gaze with road events such as obstacle, sign, and pedestrian for driver distraction determination. In [212], Rezaei and Klette monitor driver head orientation and eye state, as well as vehicle position and weather state. And in [213], they combine head pose estimation with vehicle detection and distance estimation to look at the driver and the road simultaneously. Tawari et al. [214] present Driver Attention Guard and Lane Change Recommendation systems which look at driver attention and dynamic surrounding for safe merging and breaking. Yun et al. [207] look at both inside and outside of an ego vehicle to mitigate driver distraction by monitoring driver inattention, tracking and analyzing surround vehicles. Recently, Nambi et al. [215] develop HAMS, a system implemented on an Android smartphone, which can monitor both driver and driving using the front camera and back camera. The driver monitoring module detects driver drossiness, distraction such as talking on the phone, and gaze information.

In view of driving context, the big picture of driver distraction analysis can be drawn in a stepwise approach. Early researches provide a framework and a coarse-grained understanding to monitor the driver and surroundings simultaneously. Then we can get some fine-grained understanding, knowing more precisely about the drivers and the surroundings. With this, a holistic and comprehensive understanding of the driver distraction can be obtained. As driving is a complex dynamic process involving driver, vehicle, and the driving environment [24], the systematical design consideration of a driving monitoring and assistance systems (DMAS) in future are encouraged to take into account driver behavior dissemination [10] and Internet-of-Vehicles (IoV) [216], connectivity [24] and autonomous vehicular network (AVN) [217,218], fusing driver gaze information and vehicular ego state with the current road scenario [1], as well as integrated safety [28].

Multi-modal Data: We pay attention to vision-based approaches, which extract driver behavior features from images or videos. Similar to driver fatigue detection, we can take advantage of multi-modal data from reliable multi-sensor driver assistance systems [219] to boost the performance. With the aforementioned visual cues, non-visual features from driver physiological measures and vehicle driving parameters can provide additional information as a complementary perspective.

Ryu et al. [220] use Hybrid Bayesian Network to detect abnormal driver states including fatigue, distraction and workload from multi-modal data. As for driver distraction, they make use of image, voice, biometrics, and vehicle information. Koesdwiady et al. [221] use data from camera and pressure sensors to monitor driver inattention. Craye et al. [222] present a system for driver fatigue and distraction assessment, which is composed of three modules: the vision module, audio module, and other signals module. The system is able to process multi-modal signals including audio, color video, depth map, heart rate, and steering wheel and pedals positions. Streiffer et al. [223] present Darnet to detect distracted driving using both image data from a camera and Inertial Measurement Unit (IMU) data from a mobile device.

Including Cognitive Distraction Detection: In this review, we zoom in and focus on driver behavior distraction analysis. To address the driver distraction detection problem, we also need zoom out and enlarge our vision to depict a broad landscape, and considering cognitive distraction is small one of such steps. On the other hand, one type of distraction always accompanies with other types of distraction. For example, answering a hand-held cellphone or operating navigating system may involve manual distraction and cognitive distraction. The multi-modal data make it possible for us to monitor diverse driver distraction. For example, Li and Busso [224] use multi-modal features from the CAN-bus signal, microphone array, and two video cameras facing the road and the driver to predict driver visual and cognitive distraction.

Prediction: Not satisfied with detecting driver distraction only, we target at predicting the occurrence of such event so as to prevent potential traffic accidents and improve road safety for all. Ohn-Bar et al. [121] extract cues from the driver, the ego-vehicle, and the surround. They fuse the information to study the case of overtake maneuvers prediction. Martin and Trivedi [225] use gaze fixations and transition frequencies to represent driver gaze dynamics, which are employed to model driver behavior and predict driver maneuvers. The prediction of these maneuvers sheds light on the transition of driver visual attention, which can be further used for driver distraction analysis.

There are some other topics for vision-based driver distraction analysis, such as naturalistic driving data, system design and implementation, and privacy protection as well. Despite that the simulated data involves simple cases and controlled conditions, it still has theoretical meanings to some extent. On the other hand, the virtual data can be used to model and analyze driver behavior, especially for the extreme cases which may be dangerous for naturalistic driving. As to vision-based systems for driver distraction analysis, researchers can resort to embedded systems or smartphones. In such cases, the trade-off between accuracy and speed is almost inevitable. Along with the technical challenges, there can be some social issues like the privacy protection [29,108,111,223,226], as the data may contain personally identifiable information.

Associating the evolvement of driver distraction analysis with the levels of driving automation (SAE J3016 [6]), we have been in the Level 2 (partial automation) and the early stage of Level 3 (conditional automation), as driver assistance has been prevalent worldwide and collaborative driving has been emerging. We look forward to the day when we achieve high automation even full automation driving, where driver eyes off the road, hands off the wheel, or mind off driving are no longer risky. At the same time, we hope that more researchers can devote themselves to addressing the growing distracted driving problem and enhance road safety.

8. Conclusion

Driver inattention is a significant cause of traffic accidents all over the world. As one type of driver inattention, driver fatigue has been deeply studied. As another important type of driver inattention, driver distraction has contributed more to motor vehicle crashes. However, less research attention was paid to the increasingly prevalent driver distraction. The methods making use of visual features to detect driver distraction are highly praised for their non-intrusive, user-friendly nature. With the development of computer vision and artificial intelligence, the accuracy of vision-based approaches for driver distraction analysis has improved significantly.

In the paper, we presented a comprehensive survey of vision-based driver distraction analysis for safe driving. After an overview of driver distraction, we defined driver behavior distraction and introduced the available datasets and useful cues for the detecting tasks subsequently. Then two forms of driver behavior distraction (visual distraction and manual distraction) were discussed in detail. Lastly, we concluded the evolvement and future trends for driver distraction analysis. As far as we know, we are the first to propose driver behavior distraction and analyze it hierarchically. To reduce traffic accidents and improve road safety, driver distraction analysis has been an important part of Advanced Driver Assistance Systems (ADAS), and will be an essential module for collaborative driving.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China Grant No. 61932010, 61902118, 61972139, 62133014, China Postdoctoral Science Foundation No. 2019M662771.

References

[1] Y. Dong, Z. Hu, K. Uchimura, N. Murayama, Driver inattention monitoring system for intelligent vehicles: A review, IEEE Trans. Intell. Transp. Syst. 12 (2) (2011) 596–614, http://dx.doi.org/10.1109/TITS.2010.2092770.

[2] National Highway Traffic Safety Administration traffic safety facts, 2018, URL https://www.nhtsa.gov/risky-driving/distracted-driving/ (Accessed 24 October 2020).

[3] Global Status Report on Road Safety 2018, World Health Organization, 2018.

[4] S.G. Klauer, T.A. Dingus, V.L. Neale, J.D. Sudweeks, D.J. Ramsey, The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data, 2006.

[5] L. Blincoe, T.R. Miller, E. Zaloshnja, B.A. Lawrence, The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised), Technical Report, 2015.

[6] Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles (J3016), 2021, URL https://www.sae.org/standards/content/j3016_202104/.

[7] A. El Khatib, C. Ou, F. Karray, Driver inattention detection in the context of next-generation autonomous vehicles design: A survey, IEEE Trans. Intell. Transp. Syst. 21 (11) (2020) 4483–4496, http://dx.doi.org/10.1109/TITS.2019.2940874.

[8] A. Eriksson, N.A. Stanton, Takeover time in highly automated vehicles: Non-critical transitions to and from manual control, J. Human Factors Ergon. Soc. 59 (4) (2017) 689–705, http://dx.doi.org/10.1177/0018720816685832.

[9] N. Deo, M.M. Trivedi, Looking at the driver/rider in autonomous vehicles to predict take-over readiness, IEEE Trans. Intell. Veh. 5 (1) (2018) 41–52, http://dx.doi.org/10.1109/TIV.2019.2955364, arXiv:1811.06047.

[10] S. Kaplan, M.A. Guvensan, A.G. Yavuz, Y. Karalurt, Driver behavior analysis for safe driving: A survey, IEEE Trans. Intell. Transp. Syst. 16 (6) (2015) 3017–3032, http://dx.doi.org/10.1109/TITS.2015.2462084.

[11] Global Status Report on Road Safety 2015, 2015, URL https://www.who.int/violence_injury_prevention/road_safety_status/2015/en/ (Accessed 24 October 2020).

[12] M.A. Regan, C. Hallett, C.P. Gordon, Driver distraction and driver inattention: Definition, relationship and taxonomy, Accid. Anal. Prev. 43 (5) (2011) 1771–1781, http://dx.doi.org/10.1016/j.aap.2011.04.008.

[13] M.A. Regan, D.L. Strayer, Towards an understanding of driver inattention: taxonomy and theory, Ann. Adv. Automot. Med. 58 (2014) 5.

[14] J.D. Lee, K.L. Young, M.A. Regan, Defining driver distraction, Driver distraction: Theory, effects, and mitigation 13 (4) (2008) 31–40.

[15] J. Engström, C.A. Monk, et al., A conceptual framework and taxonomy for understanding and categorizing driver inattention.

[16] T.A. Ranney, W.R. Garrott, M.J. Goodman, NHTSA Driver Distraction Research: Past, Present, and Future, Technical Report, SAE Technical Paper, 2001.

[17] Center for Disease Control and Prevention, 2020, URL https://www.cdc.gov/transportationsafety/distracted_driving/index.html (Accessed 24 October 2020).

[18] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, M. Fathy, A review on driver face monitoring systems for fatigue and distraction detection, Int. J. Adv. Sci. Technol. 64 (2014) 73–100, http://dx.doi.org/10.14257/ijast.2014.64.07.

[19] E. Murphy-Chutorian, M. Trivedi, Head pose estimation in computer vision: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 31 (4) (2008) 607–626, http://dx.doi.org/10.1109/TPAMI.2008.106.

[20] K. Lipovac, M. Đerić, M. Tešić, Z. Andrić, B. Marić, Mobile phone use while driving-literary review, Transp. Res. F 47 (2017) 132–142, http://dx.doi.org/10.1016/j.trf.2017.04.015.

[21] O. Oviedo-Trespalacios, M.M. Haque, M. King, S. Washington, Understanding the impacts of mobile phone distraction on driving performance: A systematic review, Transp. Res. C 72 (2016) 360–380, http://dx.doi.org/10.1016/j.trc.2016.10.006.

[22] E. Ohn-Bar, M.M. Trivedi, Looking at humans in the age of self-driving and highly automated vehicles, IEEE Trans. Intell. Veh. 1 (1) (2016) 90–104, http://dx.doi.org/10.1109/TIV.2016.2571067.

[23] M.M. Trivedi, T. Gandhi, J. McCall, Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety, IEEE Trans. Intell. Transp. Syst. 8 (1) (2007) 108–120, http://dx.doi.org/10.1109/TITS.2006.889442.

[24] M.Q. Khan, S. Lee, A comprehensive survey of driving monitoring and assistance systems, Sensors 19 (11) (2019) 2574, http://dx.doi.org/10.3390/s19112574.

[25] R. Chhabra, S. Verma, C.R. Krishna, A survey on driver behavior detection techniques for intelligent transportation systems, in: 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, IEEE, 2017, pp. 36–41.

[26] S. Arun, K. Sundaraj, M. Murugappan, Driver inattention detection methods: A review, in: 2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT), no. October, IEEE, 2012, pp. 1–6, http://dx.doi.org/10.1109/STUDENT.2012.6408351.

[27] H.-B. Kang, Various approaches for driver and driving behavior monitoring: A review, in: 2013 IEEE International Conference on Computer Vision Workshops, IEEE, 2013, pp. 616–623, http://dx.doi.org/10.1109/ICCVW.2013.85.

[28] A. Koesdwiady, R. Soua, F. Karray, M.S. Kamel, Recent trends in driver safety monitoring systems: State of the art and challenges, IEEE Trans. Veh. Technol. 66 (6) (2017) 4550–4563, http://dx.doi.org/10.1109/TVT.2016.2631604.

[29] A. Fernández, R. Usamentiaga, J. Carús, R. Casado, Driver distraction using visual-based sensors and algorithms, Sensors 16 (11) (2016) 1805, http://dx.doi.org/10.3390/s16111805.

[30] Labeled Faces in the Wild, URL http://vis-www.cs.umass.edu/lfw/.

[31] YouTube Faces, URL http://www.cs.tau.ac.il/~wolf/ytfaces/.

[32] K. Diaz-Chito, A. Hernández-Sabaté, A.M. López, A reduced feature set for driver head pose estimation, Appl. Soft Comput. 45 (2016) 98–107, http://dx.doi.org/10.1016/j.asoc.2016.04.027.

[33] S. Martin, K. Yuen, M.M. Trivedi, Vision for intelligent vehicles & applications (VIVA): Face detection and head pose challenge, in: 2016 IEEE Intelligent Vehicles Symposium (IV), vol. 2016-Augus, no. Iv, IEEE, 2016, pp. 1010–1014, http://dx.doi.org/10.1109/IVS.2016.7535512.

[34] K. Yuen, S. Martin, M.M. Trivedi, On looking at faces in an automobile: Issues, algorithms and evaluation on naturalistic driving dataset, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 2777–2782, http://dx.doi.org/10.1109/ICPR.2016.7900056.

[35] A. Schwarz, M. Haurilet, M. Martinez, R. Stiefelhagen, DriveAHead – A large-scale driver head pose dataset, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), vol. 2017-July, IEEE, pp. 1165–1174, https://doi.org/10.1109/CVPRW.2017.155.

[36] G. Borghi, M. Venturelli, R. Vezzani, R. Cucchiara, Poseidon: Face-from-depth for driver pose estimation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 5494–5503, http://dx.doi.org/10.1109/CVPR.2017.583, arXiv:1611.10195.

[37] M. Roth, D.M. Gavrila, DD-Pose - a large-scale driver head pose benchmark, in: 2019 IEEE Intelligent Vehicles Symposium (IV), vol. 2019-June, no. Iv, IEEE, 2019, pp. 927–934, http://dx.doi.org/10.1109/IVS.2019.8814103.

[38] M. Selim, A. Firintepe, A. Pagani, D. Stricker, AutoPOSE: Large-scale automotive driver head pose and gaze dataset with deep head orientation baseline, in: Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 4, no. January, SCITEPRESS - Science and Technology Publications, 2020, pp. 599–606, http://dx.doi.org/10.5220/0009330105990606.

[39] R.F. Ribeiro, P.D. Costa, Driver gaze zone dataset with depth data, in: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), IEEE, 2019, pp. 1–5, http://dx.doi.org/10.1109/FG.2019.8756592.

[40] N. Das, E. Ohn-Bar, M.M. Trivedi, On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, IEEE, 2015, pp. 2953–2958, http://dx.doi.org/10.1109/ITSC.2015.473.

[41] G. Borghi, E. Frigieri, R. Vezzani, R. Cucchiara, Hands on the wheel: A dataset for driver hand detection and tracking, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 564–570, http://dx.doi.org/10.1109/FG.2018.00090.

[42] H.M. Eraqi, Y. Abouelnaga, M.H. Saad, M.N. Moustafa, Driver distraction identification with an ensemble of convolutional neural networks, J. Adv. Transp. 2019 (2019) 1–12, http://dx.doi.org/10.1155/2019/4125865, arXiv:1901.09097.

[43] State farm distracted driver detection, 2016, URL https://www.kaggle.com/c/state-farm-distracted-driver-detection.

[44] Academic purposes?, URL https://www.kaggle.com/c/state-farm-distracted-driver-detection/discussion/20043#114916.

[45] M.N.M. Yehya Abouelnaga, AUC Distracted Driver Dataset, URL https://heshameraqi.github.io/distraction_detection.

[46] T. Billah, S.M. Rahman, M.O. Ahmad, M. Swamy, Recognizing distractions for assistive driving by tracking body parts, IEEE Trans. Circuits Syst. Video Technol. 29 (4) (2019) 1048–1062, http://dx.doi.org/10.1109/TCSVT.2018.2818407.

[47] EBDD: EEE BUET distracted driving - Video database, URL https://mahbubur.buet.ac.bd/resources/ebdd_database.htm.

[48] D. Yang, X. Li, X. Dai, R. Zhang, L. Qi, W. Zhang, Z. Jiang, All in one network for driver attention monitoring, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2020, pp. 2258–2262, http://dx.doi.org/10.1109/ICASSP40776.2020.9053659.

[49] M.H. Saad, M.I. Khalil, H.M. Abbas, End-to-end driver distraction recognition using novel low lighting support dataset, in: Proceedings of ICCES 2020 - 2020 15th International Conference on Computer Engineering and Systems, 2020, pp. 1–6, https://doi.org/10.1109/ICCES51560.2020.9334619.

[50] O. Kopuklu, J. Zheng, H. Xu, G. Rigoll, Driver anomaly detection: A dataset and contrastive learning approach, in: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2021, pp. 91–100, http://dx.doi.org/10.1109/WACV48630.2021.00014, arXiv:2009.14660.

[51] M. Martin, A. Roitberg, M. Haurilet, M. Horne, S. ReiB, M. Voit, R. Stiefelhagen, Drive&Act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), vol. 2019-Octob, IEEE, 2019, pp. 2801–2810, http://dx.doi.org/10.1109/ICCV.2019.00289.

[52] J.D. Ortega, N. Kose, P. Cañas, M.-A. Chao, A. Unnervik, M. Nieto, O. Otaegui, L. Salgado, DMD: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12538 LNCS, 2020, pp. 387–405, http://dx.doi.org/10.1007/978-3-030-66823-5_23, arXiv:2008.12085.

[53] I. Jegham, A. Ben Khalifa, I. Alouani, M.A. Mahjoub, MDAD: A multimodal and multiview in-vehicle driver action dataset, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11678 LNCS, Springer International Publishing, 2019, pp. 518–529, http://dx.doi.org/10.1007/978-3-030-29888-3_42.

[54] I. Jegham, A. Ben Khalifa, I. Alouani, M.A. Mahjoub, A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3MDAD, Signal Process., Image Commun. 88 (July) (2020) 115960, http://dx.doi.org/10.1016/j.image.2020.115960.

[55] K. Yuen, S. Martin, M.M. Trivedi, Looking at faces in a vehicle: A deep CNN based approach and evaluation, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016, pp. 649–654, http://dx.doi.org/10.1109/ITSC.2016.7795622.

[56] K. Yuen, M.M. Trivedi, An occluded stacked hourglass approach to facial landmark localization and occlusion estimation, IEEE Trans. Intell. Veh. 2 (4) (2017) 321–331, http://dx.doi.org/10.1109/TIV.2017.2772889, arXiv:1802.02137.

[57] J. Nuevo, L. Bergasa, M. Sotelo, M. Ocana, Real-time robust face tracking for driver monitoring, in: 2006 IEEE Intelligent Transportation Systems Conference, IEEE, 2006, pp. 1346–1351, http://dx.doi.org/10.1109/ITSC.2006.1707410.

[58] J. Nuevo, L.M. Bergasa, P. Jiménez, RSMAT: Robust simultaneous modeling and tracking, Pattern Recognit. Lett. 31 (16) (2010) 2455–2463, http://dx.doi.org/10.1016/j.patrec.2010.07.016.

[59] Z. Hu, N. Uchida, Y. Wang, Y. Dong, Face orientation estimation for driver monitoring with a single depth camera, in: 2015 IEEE Intelligent Vehicles Symposium (IV), vol. 2015-Augus, no. Iv, IEEE, 2015, pp. 958–963, http://dx.doi.org/10.1109/IVS.2015.7225808.

[60] J. Jo, S.J. Lee, H.G. Jung, K.R. Park, J. Kim, Vision-based method for detecting driver drowsiness and distraction in driver monitoring system, Opt. Eng. 50 (12) (2011) 127202, http://dx.doi.org/10.1117/1.3657506.

[61] M.-H. Sigari, M. Fathy, M. Soryani, A driver face monitoring system for fatigue and distraction detection, Int. J. Veh. Technol. 2013 (1) (2013) 1–11, http://dx.doi.org/10.1155/2013/263983.

[62] M. Sabet, R.A. Zoroofi, K. Sadeghniiat-Haghighi, M. Sabbaghian, A new system for driver drowsiness and distraction detection, in: 20th Iranian Conference on Electrical Engineering (ICEE2012), IEEE, 2012, pp. 1247–1251, http://dx.doi.org/10.1109/IranianCEE.2012.6292547.

[63] C. Braunagel, E. Kasneci, W. Stolzmann, W. Rosenstiel, Driver-activity recognition in the context of conditionally autonomous driving, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, vol. 2015-Octob, IEEE, 2015, pp. 1652–1657, http://dx.doi.org/10.1109/ITSC.2015.268.

[64] M. Rezaei, R. Klette, 3D cascade of classifiers for open and closed eye detection in driver distraction monitoring, in: CAIP 2011, vol. 6855 LNCS, no. PART 2, 2011, pp. 171–179, http://dx.doi.org/10.1007/978-3-642-23678-5_19.

[65] M. Rezaei, R. Klette, Novel adaptive eye detection and tracking for challenging lighting conditions, in: ACCV 2012, vol. 7729 LNCS, no. PART 2, 2012, pp. 427–440, http://dx.doi.org/10.1007/978-3-642-37484-5_35.

[66] J. Wiśniewska, M. Rezaei, R. Klette, Robust eye gaze estimation, in: ICCVG 2014, vol. 8671, 2014, pp. 636–644, http://dx.doi.org/10.1007/978-3-319-11331-9_76.

[67] C. Ahlstrom, T. Victor, C. Wege, E. Steinmetz, Processing of eye/head-tracking data in large-scale naturalistic driving data sets, IEEE Trans. Intell. Transp. Syst. 13 (2) (2012) 553–564, http://dx.doi.org/10.1109/TITS.2011.2174786.

[68] C. Ahlstrom, K. Kircher, A. Kircher, A gaze-based driver distraction warning system and its effect on visual behavior, IEEE Trans. Intell. Transp. Syst. 14 (2) (2013) 965–973, http://dx.doi.org/10.1109/TITS.2013.2247759.

[69] C. Ahlstrom, T. Dukic, Comparison of eye tracking systems with one and three cameras, in: Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research - MB '10, ACM Press, New York, New York, USA, 2010, pp. 1–4, http://dx.doi.org/10.1145/1931344.1931347.

[70] L.M. Bergasa, J.M. Buenaposada, J. Nuevo, P. Jimenez, L. Baumela, Analysing Driver's attention level using computer vision, in: 2008 11th International IEEE Conference on Intelligent Transportation Systems, IEEE, 2008, pp. 1149–1154, http://dx.doi.org/10.1109/ITSC.2008.4732544.

[71] S. Martin, A. Tawari, M.M. Trivedi, Monitoring head dynamics for driver assistance systems: A multi-perspective approach, in: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), no. Itsc, IEEE, 2013, pp. 2286–2291, http://dx.doi.org/10.1109/ITSC.2013.6728568.

[72] A. Tawari, S. Martin, M.M. Trivedi, Continuous head movement estimator for driver assistance: Issues, algorithms, and on-road evaluations, IEEE Trans. Intell. Transp. Syst. 15 (2) (2014) 818–830, http://dx.doi.org/10.1109/TITS.2014.2300870.

[73] E. Murphy-Chutorian, M.M. Trivedi, Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness, IEEE Trans. Intell. Transp. Syst. 11 (2) (2010) 300–311, http://dx.doi.org/10.1109/TITS.2010.2044241.

[74] E. Murphy-Chutorian, A. Doshi, M.M. Trivedi, Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation, in: 2007 IEEE Intelligent Transportation Systems Conference, IEEE, 2007, pp. 709–714, http://dx.doi.org/10.1109/ITSC.2007.4357803.

[75] E. Murphy-Chutorian, M.M. Trivedi, Hyhope: Hybrid head orientation and position estimation for vision-based driver head tracking, in: 2008 IEEE Intelligent Vehicles Symposium, IEEE, 2008, pp. 512–517, http://dx.doi.org/10.1109/IVS.2008.4621320.

[76] A. Narayanan, R.M. Kaimal, K. Bijlani, Estimation of driver head yaw angle using a generic geometric model, IEEE Trans. Intell. Transp. Syst. 17 (12) (2016) 3446–3460, http://dx.doi.org/10.1109/TITS.2016.2551298.

[77] K. Diaz-Chito, J. Martinez Del Rincon, A. Hernandez-Sabate, D. Gil, Continuous head pose estimation using manifold subspace embedding and multivariate regression, IEEE Access 6 (2018) 18325–18334, http://dx.doi.org/10.1109/ACCESS.2018.2817252.

[78] N. Alioua, A. Amine, A. Rogozan, A. Bensrhair, M. Rziza, Driver head pose estimation using efficient descriptor fusion, EURASIP J. Image Video Process. 2016 (1) (2016) 2, http://dx.doi.org/10.1186/s13640-016-0103-z.

[79] M. Hoffken, E. Tarayan, U. Kresel, K. Dietmayer, Stereo vision-based driver head pose estimation, in: 2014 IEEE Intelligent Vehicles Symposium Proceedings, no. Iv, IEEE, 2014, pp. 253–260, http://dx.doi.org/10.1109/IVS.2014.6856578.

[80] G. Borghi, R. Gasparini, R. Vezzani, R. Cucchiara, Embedded recurrent network for head pose estimation in car, in: 2017 IEEE Intelligent Vehicles Symposium (IV), no. Iv, IEEE, 2017, pp. 1503–1508, http://dx.doi.org/10.1109/IVS.2017.7995922.

[81] G. Borghi, M. Fabbri, R. Vezzani, S. Calderara, R. Cucchiara, Face-from-depth for head pose estimation on depth images, IEEE Trans. Pattern Anal. Mach. Intell. 42 (3) (2018) 596–609, http://dx.doi.org/10.1109/TPAMI.2018.2885472, arXiv:1712.05277.

[82] T. Hu, S. Jha, C. Busso, Robust driver head pose estimation in naturalistic conditions from point-cloud data, in: 2020 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2020, pp. 1176–1182, http://dx.doi.org/10.1109/IV47402.2020.9304592.

[83] J.M.D. Barros, B. Mirbach, F. Garcia, K. Varanasi, D. Stricker, Fusion of keypoint tracking and facial landmark detection for real-time head pose estimation, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), vol. 2018-Janua, IEEE, 2018, pp. 2028–2037, http://dx.doi.org/10.1109/WACV.2018.00224.

[84] Y. Liu, P. Lasang, S. Pranata, S. Shen, W. Zhang, Driver pose estimation using recurrent lightweight network and virtual data augmented transfer learning, IEEE Trans. Intell. Transp. Syst. 20 (10) (2019) 3818–3831, http://dx.doi.org/10.1109/TITS.2019.2921325.

[85] A. Mittal, A. Zisserman, P.H. Torr, Hand detection using multiple proposals, in: BMVC 2011 - Proceedings of the British Machine Vision Conference 2011, 2011.

[86] C. Zou, Y. Liu, J. Wang, H. Si, Deformable part model based hand detection against complex backgrounds, in: Communications in Computer and Information Science, vol. 634, 2016, pp. 149–159, http://dx.doi.org/10.1007/978-981-10-2260-9_17.

[87] X. Deng, Y. Zhang, S. Yang, P. Tan, L. Chang, Y. Yuan, H. Wang, Joint hand detection and rotation estimation using CNN, IEEE Trans. Image Process. 27 (4) (2018) 1888–1900, http://dx.doi.org/10.1109/TIP.2017.2779600.

[88] Q. Wang, G. Zhang, S. Yu, 2D hand detection using multi-feature skin model supervised cascaded CNN, J. Signal Process. Syst. 91 (10) (2019) 1105–1113, http://dx.doi.org/10.1007/s11265-018-1406-3.

[89] T. Zhou, P.J. Pillai, V.G. Yalla, Hierarchical context-aware hand detection algorithm for naturalistic driving, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), no. Icd, IEEE, 2016, pp. 1291–1297, http://dx.doi.org/10.1109/ITSC.2016.7795723.

[90] Y. Xia, S. Yan, B. Zhang, Combination of ACF detector and multi-task CNN for hand detection, in: 2016 IEEE 13th International Conference on Signal Processing (ICSP), IEEE, 2016, pp. 601–606, http://dx.doi.org/10.1109/ICSP.2016.7877903.

[91] S. Yan, Y. Xia, J.S. Smith, W. Lu, B. Zhang, Multiscale convolutional neural networks for hand detection, Appl. Comput. Intell. Soft Comput. 2017 (2017) 1–13, http://dx.doi.org/10.1155/2017/9830641.

[92] T.H.N. Le, C. Zhu, Y. Zheng, K. Luu, M. Savvides, Robust hand detection in vehicles, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 573–578, http://dx.doi.org/10.1109/ICPR.2016.7899695.

[93] T.H.N. Le, K.G. Quach, C. Zhu, C.N. Duong, K. Luu, M. Savvides, Robust hand detection and classification in vehicles and in the wild, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), vol. 2017-July, IEEE, 2017, pp. 1203–1210, https://doi.org/10.1109/CVPRW.2017.159.

[94] E. Ohn-Bar, M. Trivedi, In-vehicle hand activity recognition using integration of regions, in: 2013 IEEE Intelligent Vehicles Symposium (IV), no. Iv, IEEE, 2013, pp. 1034–1039, http://dx.doi.org/10.1109/IVS.2013.6629602.

[95] E. Ohn-Bar, S. Martin, M.M. Trivedi, Driver hand activity analysis in naturalistic driving studies: challenges, algorithms, and experimental studies, J. Electron. Imaging 22 (4) (2013) 041119, http://dx.doi.org/10.1117/1.JEI.22.4.041119.

[96] E. Ohn-Bar, M.M. Trivedi, Beyond just keeping hands on the wheel: Towards visual interpretation of driver hand motion patterns, in: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), IEEE, 2014, pp. 1245–1250, http://dx.doi.org/10.1109/ITSC.2014.6957858.

[97] A. Rangesh, E. Ohn-Bar, M.M. Trivedi, Long-term multi-cue tracking of hands in vehicles, IEEE Trans. Intell. Transp. Syst. 17 (5) (2016) 1483–1492, http://dx.doi.org/10.1109/TITS.2015.2508722.

[98] A. Rangesh, E. Ohn-Bar, M.M. Trivedi, Hidden hands: Tracking hands with an occlusion aware tracker, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2016, pp. 1224–1231, https://doi.org/10.1109/CVPRW.2016.155.

[99] N. Deo, A. Rangesh, M. Trivedi, In-vehicle hand gesture recognition using hidden Markov models, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016, pp. 2179–2184, http://dx.doi.org/10.1109/ITSC.2016.7795908.

[100] Siddharth, A. Rangesh, E. Ohn-Bar, M.M. Trivedi, Driver hand localization and grasp analysis: A vision-based real-time approach, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016, pp. 2545–2550, http://dx.doi.org/10.1109/ITSC.2016.7795965.

[101] T.H.N. Le, C. Zhu, Y. Zheng, K. Luu, M. Savvides, DeepSafeDrive: A grammar-aware driver parsing approach to Driver Behavioral Situational Awareness (DB-SAW), Pattern Recognit. 66 (December 2016) (2017) 229–238, http://dx.doi.org/10.1016/j.patcog.2016.11.028.

[102] P. Weyers, D. Schiebener, A. Kummert, Action and object interaction recognition for driver activity classification, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 2019, pp. 4336–4341, http://dx.doi.org/10.1109/ITSC.2019.8917139.

[103] A. Rangesh, M.M. Trivedi, HandyNet: A one-stop solution to detect, segment, localize & analyze driver hands, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), vol. 2018-June, IEEE, 2018, pp. 1216–12167, arXiv:1804.07834, https://doi.org/10.1109/CVPRW.2018.00158.

[104] K. Yuen, M.M. Trivedi, Looking at hands in autonomous vehicles: A ConvNet approach using part affinity fields, IEEE Trans. Intell. Veh. (2019) 1, http://dx.doi.org/10.1109/TIV.2019.2955369, arXiv:1804.01176.

[105] X. Zhang, N. Zheng, F. Wang, Y. He, Visual recognition of driver hand-held cell phone use based on hidden CRF, in: Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety, IEEE, 2011, pp. 248–251, http://dx.doi.org/10.1109/ICVES.2011.5983823.

[106] R.A. Berri, A.G. Silva, R.S. Parpinelli, E. Girardi, R. Arthur, A pattern recognition system for detecting use of mobile phones while driving, in: VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications, vol. 2, SCITEPRESS, 2014, pp. 411–418.

[107] Y. Artan, O. Bulan, R.P. Loce, P. Paul, Driver cell phone usage detection from HOV/HOT NIR images, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2014, pp. 225–230, https://doi.org/10.1109/CVPRW.2014.42.

[108] K. Seshadri, F. Juefei-Xu, D.K. Pal, M. Savvides, C.P. Thor, Driver cell phone usage detection on Strategic Highway Research Program (SHRP2) face view videos, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), vol. 2015-Octob, IEEE, 2015, pp. 35–43, https://doi.org/10.1109/CVPRW.2015.7301397.

[109] T.H.N. Le, Y. Zheng, C. Zhu, K. Luu, M. Savvides, Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2016, pp. 46–53, https://doi.org/10.1109/CVPRW.2016.13.

[110] R. Torres, O. Ohashi, E. Carvalho, G. Pessin, A deep learning approach to detect distracted drivers using a mobile phone, in: ICANN 2017, vol. 10614 LNCS, 2017, pp. 72–79, http://dx.doi.org/10.1007/978-3-319-68612-7_9.

[111] J.M. Celaya-Padilla, C.E. Galván-Tejada, J.S.A. Lozano-Aguilar, L.A. Zanella-Calzada, H. Luna-García, J.I. Galván-Tejada, N.K. Gamboa-Rosales, A. Velez Rodriguez, H. Gamboa-Rosales, "Texting & Driving" detection using deep convolutional neural networks, Appl. Sci. 9 (15) (2019) 2962, http://dx.doi.org/10.3390/app9152962.

[112] M.L. Watkins, I.A. Amaya, P.E. Keller, M.A. Hughes, E.D. Beck, Autonomous detection of distracted driving by cell phone, in: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), IEEE, 2011, pp. 1960–1965, http://dx.doi.org/10.1109/ITSC.2011.6083026.

[113] C. Tran, A. Doshi, M.M. Trivedi, Pedal error prediction by driver foot gesture analysis: A vision-based inquiry, in: 2011 IEEE Intelligent Vehicles Symposium (IV), no. Iv, IEEE, 2011, pp. 577–582, http://dx.doi.org/10.1109/IVS.2011.5940548.

[114] C. Tran, A. Doshi, M.M. Trivedi, Modeling and prediction of driver behavior by foot gesture analysis, Comput. Vis. Image Underst. 116 (3) (2012) 435–445, http://dx.doi.org/10.1016/j.cviu.2011.09.008.

[115] Y. Wu, L.N. Boyle, D. McGehee, C.A. Roe, K. Ebe, J. Foley, Foot placement during error and pedal applications in naturalistic driving, Accident Anal. Prevention 99 (2017) 102–109, http://dx.doi.org/10.1016/j.aap.2016.10.019.

[116] Y. Wu, L.N. Boyle, D.V. McGehee, Evaluating variability in foot to pedal movements using functional principal components analysis, Accident Anal Prevention 118 (January) (2018) 146–153, http://dx.doi.org/10.1016/j.aap.2018.02.011.

[117] A. Rangesh, M. Trivedi, Forced spatial attention for driver foot activity classification, in: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), IEEE, 2019, pp. 2514–2521, arXiv:1907.11824, https://doi.org/10.1109/ICCVW.2019.00308.

[118] R. Oyini Mbouna, S.G. Kong, M.-G. Chun, Visual analysis of eye state and head pose for driver alertness monitoring, IEEE Trans. Intell. Transp. Syst. 14 (3) (2013) 1462–1469, http://dx.doi.org/10.1109/TITS.2013.2262098.

[119] S. Martin, E. Ohn-Bar, A. Tawari, M.M. Trivedi, Understanding head and hand activities and coordination in naturalistic driving videos, in: 2014 IEEE Intelligent Vehicles Symposium Proceedings, no. Iv, IEEE, 2014, pp. 884–889, http://dx.doi.org/10.1109/IVS.2014.6856610.

[120] E. Ohn-Bar, S. Martin, A. Tawari, M.M. Trivedi, Head, eye, and hand patterns for driver activity recognition, in: 2014 22nd International Conference on Pattern Recognition, IEEE, 2014, pp. 660–665, http://dx.doi.org/10.1109/ICPR.2014.124.

[121] E. Ohn-Bar, A. Tawari, S. Martin, M.M. Trivedi, Vision on wheels: Looking at driver, vehicle, and surround for on-road maneuver analysis, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2014, pp. 185–190, https://doi.org/10.1109/CVPRW.2014.33.

[122] Y. Abouelnaga, H.M. Eraqi, M.N. Moustafa, Real-time distracted driver posture classification, 2017, arXiv:1706.09498.

[123] P. Smith, M. Shah, N. da Vitoria Lobo, Determining driver visual attention with one camera, IEEE Trans. Intell. Transp. Syst. 4 (4) (2003) 205–218, http://dx.doi.org/10.1109/TITS.2003.821342.

[124] L. Alam, M.M. Hoque, Real-time distraction detection based on Driver's visual features, in: 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, 2019, pp. 1–6, http://dx.doi.org/10.1109/ECACE.2019.8679191.

[125] M.J.C. De Castro, J.R.E. Medina, J.P.G. Lopez, J.C. De Goma, M. Devaraj, A non-intrusive method for detecting visual distraction indicators of transport network vehicle service drivers using computer vision, in: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM), IEEE, 2018, pp. 1–5, http://dx.doi.org/10.1109/HNICEM.2018.8666306.

[126] M.J.C. De Castro, J.C. De Goma, M. Devaraj, J.P.G. Lopez, J.R.E. Medina, Distraction detection through facial attributes of transport network vehicle service drivers, in: Proceedings of the 2018 International Conference on Information Hiding and Image Processing - IHIP 2018, ACM Press, New York, New York, USA, 2018, pp. 112–118, http://dx.doi.org/10.1145/3292425.3293464.

[127] I. Dua, A.U. Nambi, C. Jawahar, V. Padmanabhan, AutoRate: How attentive is the driver? in: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), no. i, IEEE, 2019, pp. 1–8, http://dx.doi.org/10.1109/FG.2019.8756620.

[128] S. Jha, C. Busso, Probabilistic estimation of the driver's gaze from head orientation and position, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), vol. 2018-March, IEEE, 2017, pp. 1–6, http://dx.doi.org/10.1109/ITSC.2017.8317841.

[129] S. Jha, C. Busso, Probabilistic estimation of the gaze region of the driver using dense classification, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), vol. 2018-Novem, IEEE, 2018, pp. 697–702, http://dx.doi.org/10.1109/ITSC.2018.8569709.

[130] Y. Wang, T. Zhao, X. Ding, J. Bian, X. Fu, Head pose-free eye gaze prediction for driver attention study, in: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2017, pp. 42–46, http://dx.doi.org/10.1109/BIGCOMP.2017.7881713.

[131] S.J. Lee, J. Jo, H.G. Jung, K.R. Park, J. Kim, Real-time gaze estimator based on Driver's head orientation for forward collision warning system, IEEE Trans. Intell. Transp. Syst. 12 (1) (2011) 254–267, http://dx.doi.org/10.1109/TITS.2010.2091503.

[132] Q. Ji, X. Yang, Real-time eye, gaze, and face pose tracking for monitoring driver vigilance, Real-Time Imag. 8 (5) (2002) 357–377, http://dx.doi.org/10.1006/rtim.2002.0279.

[133] W.-C. Cheng, Y.-S. Xu, Real-time driver gaze direction detection using the 3D triangle model and neural networks, in: 2013 7th Asia Modelling Symposium, IEEE, 2013, pp. 41–45, http://dx.doi.org/10.1109/AMS.2013.11.

[134] M. Lundgren, L. Hammarstrand, T. McKelvey, Driver-gaze zone estimation using Bayesian filtering and Gaussian processes, IEEE Trans. Intell. Transp. Syst. 17 (10) (2016) 2739–2750, http://dx.doi.org/10.1109/TITS.2016.2526050.

[135] W. Kim, W.-S. Jung, H.K. Choi, Lightweight driver monitoring system based on multi-task mobilenets, Sensors 19 (14) (2019) 3200, http://dx.doi.org/10.3390/s19143200.

[136] F. Vicente, Z. Huang, X. Xiong, F. De la Torre, W. Zhang, D. Levi, Driver gaze tracking and eyes off the road detection system, IEEE Trans. Intell. Transp. Syst. 16 (4) (2015) 2014–2027, http://dx.doi.org/10.1109/TITS.2015.2396031.

[137] J. Zeng, Y. Sun, L. Jiang, Driver distraction detection and identity recognition in real-time, in: 2010 Second WRI Global Congress on Intelligent Systems, vol. 3, IEEE, 2010, pp. 43–46, http://dx.doi.org/10.1109/GCIS.2010.83.

[138] I.-H. Choi, S.K. Hong, Y.-G. Kim, Real-time categorization of driver's gaze zone using the deep learning techniques, in: 2016 International Conference on Big Data and Smart Computing (BigComp), IEEE, 2016, pp. 143–148, http://dx.doi.org/10.1109/BIGCOMP.2016.7425813.

[139] X. Fu, X. Guan, E. Peli, H. Liu, G. Luo, Automatic calibration method for Driver's head orientation in natural driving environment, IEEE Trans. Intell. Transp. Syst. 14 (1) (2013) 303–312, http://dx.doi.org/10.1109/TITS.2012.2217377.

[140] B. Vasli, S. Martin, M.M. Trivedi, On driver gaze estimation: Explorations and fusion of geometric and data driven approaches, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016, pp. 655–660, http://dx.doi.org/10.1109/ITSC.2016.7795623.

[141] S. Vora, A. Rangesh, M.M. Trivedi, On generalizing driver gaze zone estimation using convolutional neural networks, in: 2017 IEEE Intelligent Vehicles Symposium (IV), no. Iv, IEEE, 2017, pp. 849–854, http://dx.doi.org/10.1109/IVS.2017.7995822.

[142] S. Vora, A. Rangesh, M.M. Trivedi, Driver gaze zone estimation using convolutional neural networks: A general framework and ablative analysis, IEEE Trans. Intell. Veh. 3 (3) (2018) 254–265, http://dx.doi.org/10.1109/TIV.2018.2843120, arXiv:1802.02690.

[143] N. Hernandez, P. Jimenez, L. Bergasa, I. Parra, I. Garcia, M. Ocana, B. Delgado, M. Sevillano, Assessment of distractions inferred by in-vehicle information systems on a naturalistic simulator, in: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), IEEE, 2011, pp. 1279–1284, http://dx.doi.org/10.1109/ITSC.2011.6082886.

[144] P. Jimenez, L.M. Bergasa, J. Nuevo, N. Hernandez, I.G. Daza, Gaze fixation system for the evaluation of driver distractions induced by IVIS, IEEE Trans. Intell. Transp. Syst. 13 (3) (2012) 1167–1178, http://dx.doi.org/10.1109/TITS.2012.2187517.

[145] A. Tawari, M.M. Trivedi, Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos, in: 2014 IEEE Intelligent Vehicles Symposium Proceedings, vol. 12, no. 1, IEEE, 2014, pp. 344–349, http://dx.doi.org/10.1109/IVS.2014.6856607.

[146] A. Tawari, K.H. Chen, M.M. Trivedi, Where is the driver looking: Analysis of head, eye and iris for robust gaze zone estimation, in: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), IEEE, 2014, pp. 988–994, http://dx.doi.org/10.1109/ITSC.2014.6957817.

[147] L. Fridman, P. Langhans, J. Lee, B. Reimer, Driver gaze region estimation without use of eye movement, IEEE Intell. Syst. 31 (3) (2016) 49–56, http://dx.doi.org/10.1109/MIS.2016.47, arXiv:1507.04760.

[148] L. Fridman, J. Lee, B. Reimer, T. Victor, Owl and lizard: Patterns of head pose and eye pose in driver gaze classification, IET Comput. Vis. 10 (4) (2016) 308–314, arXiv:1508.04028.

[149] S. Jha, C. Busso, Analyzing the relationship between head pose and gaze to model driver visual attention, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016, pp. 2157–2162, http://dx.doi.org/10.1109/ITSC.2016.7795905.

[150] N. Hernandez, P. Jimenez, L. Bergasa, B. Delgado, M. Sevillano, Vision-based distraction analysis tested on a realistic driving simulator, in: 13th International IEEE Conference on Intelligent Transportation Systems, IEEE, 2010, pp. 895–902, http://dx.doi.org/10.1109/ITSC.2010.5625006.

[151] J. Cronje, A.P. Engelbrecht, Training convolutional neural networks with class based data augmentation for detecting distracted drivers, in: Proceedings of the 9th International Conference on Computer and Automation Engineering - ICCAE '17, vol. Part F1278, ACM Press, New York, New York, USA, 2017, pp. 126–130, http://dx.doi.org/10.1145/3057039.3057070.

[152] S. Masood, A. Rai, A. Aggarwal, M. Doja, M. Ahmad, Detecting distraction of drivers using convolutional neural network, Pattern Recognit. Lett. (2018) 1–7, http://dx.doi.org/10.1016/j.patrec.2017.12.023.

[153] L.C. Valeriano, P. Napoletano, R. Schettini, Recognition of driver distractions using deep learning, in: 2018 IEEE 8th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), vol. 2018-Septe, IEEE, 2018, pp. 1–6, http://dx.doi.org/10.1109/ICCE-Berlin.2018.8576183.

[154] Z.A. Varaich, S. Khalid, Recognizing actions of distracted drivers using inception v3 and xception convolutional neural networks, in: 2019 2nd International Conference on Advancements in Computational Sciences (ICACS), IEEE, 2019, pp. 1–8, http://dx.doi.org/10.23919/ICACS.2019.8689131.

[155] B. Baheti, S. Gajre, S. Talbar, Detection of distracted driver using convolutional neural network, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), vol. 2018-June, IEEE, 2018, pp. 1145–11456, https://doi.org/10.1109/CVPRW.2018.00150.

[156] M.R. Arefin, F. Makhmudkhujaev, O. Chae, J. Kim, Aggregating CNN and HOG features for real-time distracted driver detection, in: 2019 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2019, pp. 1–3, http://dx.doi.org/10.1109/ICCE.2019.8661970.

[157] M. Alotaibi, B. Alotaibi, Distracted driver classification using deep learning, Signal Image Video Process. 14 (3) (2020) 617–624, http://dx.doi.org/10.1007/s11760-019-01589-z.

[158] N. Moslemi, R. Azmi, M. Soryani, Driver distraction recognition using 3D convolutional neural networks, in: 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA), IEEE, 2019, pp. 145–151, http://dx.doi.org/10.1109/PRIA.2019.8786012.

[159] D. Tran, H. Manh Do, W. Sheng, H. Bai, G. Chowdhary, Real-time detection of distracted driving based on deep learning, IET-ITS (2018) http://dx.doi.org/10.1049/iet-its.2018.5172.

[160] P. Weyers, A. Barth, A. Kummert, Driver state monitoring with hierarchical classification, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), vol. 2018-Novem, IEEE, 2018, pp. 3239–3244, http://dx.doi.org/10.1109/ITSC.2018.8569467.

[161] C. Ou, Q. Zhao, F. Karray, A.E. Khatib, Design of an end-to-end dual mode driver distraction detection system, in: ICIAR 2019, vol. 2, no. 1831795, Springer International Publishing, 2019, pp. 199–207, http://dx.doi.org/10.1007/978-3-030-27272-2_17.

[162] M. Hssayeni, S. Saxena, R. Ptucha, A. Savakis, Distracted driver detection: Deep learning vs handcrafted features, Electron. Imaging 2017 (10) (2017) 20–26, http://dx.doi.org/10.2352/ISSN.2470-1173.2017.10.IMAWM-162.

[163] C. Yan, H. Jiang, B. Zhang, F. Coenen, Recognizing driver inattention by convolutional neural networks, in: 2015 8th International Congress on Image and Signal Processing (CISP), no. Cisp, IEEE, 2015, pp. 680–685, http://dx.doi.org/10.1109/CISP.2015.7407964.

[164] C. Yan, B. Zhang, F. Coenen, Driving posture recognition by convolutional neural networks, in: 2015 11th International Conference on Natural Computation (ICNC), vol. 2016-Janua, IEEE, 2015, pp. 680–685, http://dx.doi.org/10.1109/ICNC.2015.7378072.

[165] C. Yan, F. Coenen, B. Zhang, Driving posture recognition by convolutional neural networks, IET Comput. Vis. 10 (2) (2016) 103–114, http://dx.doi.org/10.1049/iet-cvi.2015.0175.

[166] S. Yan, Y. Teng, J.S. Smith, B. Zhang, Driver behavior recognition based on deep convolutional neural networks, in: 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), no. 1, IEEE, 2016, pp. 636–641, http://dx.doi.org/10.1109/FSKD.2016.7603248.

[167] A. Koesdwiady, S.M. Bedawi, C. Ou, F. Karray, End-to-end deep learning for driver distraction recognition, in: F. Karray, A. Campilho, F. Cheriet (Eds.), ICIAR 2017, in: Lecture Notes in Computer Science, vol. 10317, Springer International Publishing, Cham, 2017, pp. 11–18, http://dx.doi.org/10.1007/978-3-319-59876-5_2.

[168] E. Cengil, A. Cinar, Classification of human driving behaviour images using convolutional neural network architecture, in: H. Pérez García, L. Sánchez González, M. Castejón Limas, H. Quintián Pardo, E. Corchado Rodríguez (Eds.), HAIS 2019, in: Lecture Notes in Computer Science, vol. 11734, Springer International Publishing, Cham, 2019, pp. 264–274, http://dx.doi.org/10.1007/978-3-030-29859-3_23.

[169] K. Kapoor, R. Pamula, S.V. Murthy, Real-time driver distraction detection system using convolutional neural networks, in: P.K. Singh, B.K. Panigrahi, N.K. Suryadevara, S.K. Sharma, A.P. Singh (Eds.), ICETIT 2019, in: Lecture Notes in Electrical Engineering, vol. 605, (January) Springer International Publishing, Cham, 2020, pp. 280–291, http://dx.doi.org/10.1007/978-3-030-30577-2_24.

[170] F.R. da Silva Oliveira, F.C. Farias, Comparing transfer learning approaches applied to distracted driver detection, in: 2018 IEEE Latin American Conference on Computational Intelligence (la-CCI), no. i, IEEE, 2018, pp. 1–6, http://dx.doi.org/10.1109/LA-CCI.2018.8625214.

[171] C. Ou, C. Ouali, F. Karray, Transfer learning based strategy for improving driver distraction recognition, in: ICIAR 2018, vol. 1, no. June, Springer International Publishing, 2018, pp. 443–452, http://dx.doi.org/10.1007/978-3-319-93000-8_50.

[172] C. Ou, C. Ouali, S.M. Bedawi, F. Karray, Driver behavior monitoring using tools of deep learning and fuzzy inferencing, in: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), vol. 2018-July, IEEE, 2018, pp. 1–7, http://dx.doi.org/10.1109/FUZZ-IEEE.2018.8491511.

[173] P.M. Chawan, S. Satardekar, D. Shah, R. Badugu, A. Pawar, Distracted driver detection and classification, Int. J. Eng. Res. Appl. 8 (April) (2018) 60–64, http://dx.doi.org/10.9790/9622-0804036064.

[174] C. Huang, X. Wang, J. Cao, S. Wang, Y. Zhang, HCF: A hybrid CNN framework for behavior detection of distracted drivers, IEEE Access 8 (2020) 109335–109349, http://dx.doi.org/10.1109/ACCESS.2020.3001159.

[175] C. Yan, F. Coenen, Y. Yue, X. Yang, B. Zhang, Video-based classification of driving behavior using a hierarchical classification system with multiple features, Int. J. Pattern Recognit. Artif. Intell. 30 (05) (2016) 1650010, http://dx.doi.org/10.1142/S0218001416500105.

[176] M.S. Majdi, S. Ram, J.T. Gill, J.J. Rodriguez, Drive-Net: Convolutional network for driver distraction detection, in: 2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), vol. 2018-April, IEEE, 2018, pp. 1–4, http://dx.doi.org/10.1109/SSIAI.2018.8470309, arXiv:2006.12586.

[177] C. Ou, F. Karray, Enhancing driver distraction recognition using generative adversarial networks, IEEE Trans. Intell. Veh. 5 (3) (2020) 385–396, http://dx.doi.org/10.1109/TIV.2019.2960930.

[178] S. Mittal, S. Vaishay, A survey of techniques for optimizing deep learning on GPUs, J. Syst. Archit. 99 (2019) 101635.

[179] A.G. Blaiech, K.B. Khalifa, C. Valderrama, M.A. Fernandes, M.H. Bedoui, A survey and taxonomy of FPGA-based deep learning accelerators, J. Syst. Archit. 98 (2019) 331–345.

[180] S. Mittal, A survey on optimized implementation of deep learning models on the NVIDIA Jetson platform, J. Syst. Archit. 97 (2019) 428–442.

[181] S. Mittal, et al., A survey of accelerator architectures for 3D convolution neural networks, J. Syst. Archit. (2021) 102041.

[182] M. Kutila, M. Jokela, G. Markkula, M.R. Rue, Driver distraction detection with a camera vision system, in: 2007 IEEE International Conference on Image Processing, vol. 6, no. 3 1, IEEE, 2007, http://dx.doi.org/10.1109/ICIP.2007.4379556, VI – 201–VI – 204.

[183] C. Zhao, B. Zhang, J. He, J. Lian, Recognition of driving postures by contourlet transform and random forests, IET Intell. Transport Syst. 6 (2) (2012) 161, http://dx.doi.org/10.1049/iet-its.2011.0116.

[184] C. Yan, B. Zhang, F. Coenen, Driving posture recognition by a hierarchal classification system with multiple features, in: 2014 7th International Congress on Image and Signal Processing, IEEE, 2014, pp. 83–88, http://dx.doi.org/10.1109/CISP.2014.7003754.

[185] A. Behera, Z. Wharton, A. Keidel, B. Debnath, Deep CNN, body pose and body-object interaction features for Drivers' activity monitoring, IEEE Trans. Intell. Transp. Syst. (2020) 1–8, http://dx.doi.org/10.1109/TITS.2020.3027240.

[186] M. Leekha, M. Goswami, R.R. Shah, Y. Yin, R. Zimmermann, Are you paying attention? Detecting distracted driving in real-time, in: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), IEEE, 2019, pp. 171–180, http://dx.doi.org/10.1109/BigMM.2019.00-28.

[187] K. A. AlShalfan, M. Zakariah, Detecting driver distraction using deep-learning approach, Comput. Mater. Continua 68 (1) (2021) 689–704, http://dx.doi.org/10.32604/cmc.2021.015989.

[188] A. Narayanan, V. Aiswaryaa, A.T. Anand, N. Kadiresan, Real-time detection of distracted drivers using a deep neural network and multi-threading, in: Advances in Intelligent Systems and Computing, vol. 1133, Springer Singapore, 2021, pp. 89–100, http://dx.doi.org/10.1007/978-981-15-3514-7_8.

[189] M. Tang, F. Wu, L.-L. Zhao, Q.-p. Liang, J.-w. Lin, Y.-b. Zhao, Detection of distracted driving based on MultiGranularity and middle-level features, in: 2020 Chinese Automation Congress (CAC), no. 2018, IEEE, 2020, pp. 2717–2722, http://dx.doi.org/10.1109/CAC51589.2020.9327078.

[190] T. Abbas, S.F. Ali, A.Z. Khan, I. Kareem, Optnet-50: An optimized residual neural network architecture of deep learning for Driver's distraction, in: 2020 IEEE 23rd International Multitopic Conference (INMIC), IEEE, 2020, pp. 1–5, http://dx.doi.org/10.1109/INMIC50486.2020.9318087.

[191] D.K. Jain, R. Jain, X. Lan, Y. Upadhyay, A. Thareja, Driver distraction detection using capsule network, Neural Comput. Appl. 33 (11) (2021) 6183–6196, http://dx.doi.org/10.1007/s00521-020-05390-9.

[192] M. Lu, Y. Hu, X. Lu, Driver action recognition using deformable and dilated faster R-CNN with optimized region proposals, Appl. Intell. 50 (4) (2020) 1100–1111, http://dx.doi.org/10.1007/s10489-019-01603-4.

[193] B. Baheti, S. Talbar, S. Gajre, Towards computationally efficient and realtime distracted driver detection with mobilevgg network, IEEE Trans. Intell. Veh. 5 (4) (2020) 565–574, http://dx.doi.org/10.1109/TIV.2020.2995555.

[194] A. Gumaei, M. Al-Rakhami, M.M. Hassan, A. Alamri, M. Alhussein, M.A. Razzaque, G. Fortino, A deep learning-based driver distraction identification framework over edge cloud, Neural Comput. Appl. 1 (2020) http://dx.doi.org/10.1007/s00521-020-05328-1.

[195] J. Mafeni Mase, P. Chapman, G.P. Figueredo, M. Torres Torres, A hybrid deep learning approach for driver distraction detection, in: 2020 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, 2020, pp. 1–6, http://dx.doi.org/10.1109/ICTC49870.2020.9289588.

[196] J. Mafeni Mase, P. Chapman, G.P. Figueredo, M. Torres Torres, Benchmarking deep learning models for driver distraction detection, in: G. Nicosia, V. Ojha, E. La Malfa, G. Jansen, V. Sciacca, P. Pardalos, G. Giuffrida, R. Umeton (Eds.), in: Lecture Notes in Computer Science, vol. 12566, Springer International Publishing, Cham, 2020, pp. 103–117, http://dx.doi.org/10.1007/978-3-030-64580-9_9.

[197] L. Ye, C. Chen, M. Wu, S. Nwobodo, A.A. Antwi, C.N. Muponda, K.D. Ernest, R.S. Vedaste, Using CNN and channel attention mechanism to identify Driver's distracted behavior, in: Z. Pan, A.D. Cheok, W. Müller, M. Zhang (Eds.), in: Lecture Notes in Computer Science, vol. 11782, Springer Berlin Heidelberg, Berlin, Heidelberg, 2020, pp. 175–183, http://dx.doi.org/10.1007/978-3-662-61510-2_17.

[198] I. Jegham, A.B. Khalifa, I. Alouani, M.A. Mahjoub, Soft spatial attention-based multimodal driver action recognition using deep learning, IEEE Sens. J. 21 (2) (2021) 1918–1925, http://dx.doi.org/10.1109/JSEN.2020.3019258.

[199] Z. Wharton, A. Behera, Y. Liu, N. Bessis, Coarse temporal attention network (CTA-Net) for Driver's activity recognition, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1279–1289.

[200] M. Martin, S. Stuehmer, M. Voit, R. Stiefelhagen, Real time driver body pose estimation for novel assistance systems, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), vol. 2018-March, IEEE, 2017, pp. 1–7, http://dx.doi.org/10.1109/ITSC.2017.8317722.

[201] M. Martin, J. Popp, M. Anneken, M. Voit, R. Stiefelhagen, Body pose and context information for driver secondary task detection, in: 2018 IEEE Intelligent Vehicles Symposium (IV), vol. 2018-June, no. Iv, IEEE, 2018, pp. 2015–2021, http://dx.doi.org/10.1109/IVS.2018.8500523.

[202] Y. Xing, J. Tang, H. Liu, C. Lv, D. Cao, E. Velenis, F.-Y. Wang, End-to-end driving activities and secondary tasks recognition using deep convolutional neural network and transfer learning, in: 2018 IEEE Intelligent Vehicles Symposium (IV), vol. 2018-June, no. Iv, IEEE, 2018, pp. 1626–1631, http://dx.doi.org/10.1109/IVS.2018.8500548.

[203] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, F.Y. Wang, Driver activity recognition for intelligent vehicles: A deep learning approach, IEEE Trans. Veh. Technol. 68 (6) (2019) 5379–5390, http://dx.doi.org/10.1109/TVT.2019.2908425.

[204] T.H. Dinh, M.T. Pham, M.D. Phung, D.M. Nguyen, V.M. Hoang, Q.V. Tran, Image segmentation based on histogram of depth and an application in driver distraction detection, in: 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), vol. 2014, no. December, IEEE, 2014, pp. 969–974, http://dx.doi.org/10.1109/ICARCV.2014.7064437.

[205] C. Craye, F. Karray, Driver distraction detection and recognition using RGB-D sensor, 2015, arXiv:1502.00250.

[206] M.-C. Wu, C.-T. Chiu, Multi-teacher knowledge distillation for compressed video action recognition based on deep learning, J. Syst. Archit. 103 (2020) 101695.

[207] Y. Yun, I.Y. Gu, M. Bolbat, Z.H. Khan, Video-based detection and analysis of driver distraction and inattention, in: 2014 International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2014, pp. 190–195, http://dx.doi.org/10.1109/SPIN.2014.6776946.

[208] P. Li, M. Lu, Z. Zhang, D. Shan, Y. Yang, A novel spatial-temporal graph for skeleton-based driver action recognition, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 2019, pp. 3243–3248, http://dx.doi.org/10.1109/ITSC.2019.8916929.

[209] N. Kose, O. Kopuklu, A. Unnervik, G. Rigoll, Real-time driver state monitoring using a CNN based spatio-temporal approach *, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 2019, pp. 3236–3242, http://dx.doi.org/10.1109/ITSC.2019.8917460, arXiv:1907.08009.

[210] A. Doshi, M. Trivedi, Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions, in: 2009 IEEE Intelligent Vehicles Symposium, IEEE, 2009, pp. 887–892, http://dx.doi.org/10.1109/IVS.2009.5164397.

[211] L. Fletcher, A. Zelinsky, Driver inattention detection based on eye gaze-road event correlation, Int. J. Robot. Res. 28 (6) (2009) 774–801, http://dx.doi.org/10.1177/0278364908099459.

[212] M. Rezaei, R. Klette, Simultaneous analysis of driver behaviour and road condition for driver distraction detection, Int. J. Image Data Fusion 2 (3) (2011) 217–236, http://dx.doi.org/10.1080/19479832.2011.590458.

[213] M. Rezaei, R. Klette, Look at the driver, look at the road: No distraction! no accident!, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2014, http://dx.doi.org/10.1109/CVPR.2014.24.

[214] A. Tawari, S. Sivaraman, M.M. Trivedi, T. Shannon, M. Tippelhofer, Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking, in: 2014 IEEE Intelligent Vehicles Symposium Proceedings, no. Iv, IEEE, 2014, pp. 115–120, http://dx.doi.org/10.1109/IVS.2014.6856600.

[215] A.U. Nambi, S. Bannur, I. Mehta, H. Kalra, A. Virmani, V.N. Padmanabhan, R. Bhandari, B. Raman, HAMS: Driver and driving monitoring using a smartphone, in: Proceedings of the 24th Annual International Conference on Mobile Computing and Networking - MobiCom '18, ACM Press, New York, New York, USA, 2018, pp. 840–842, http://dx.doi.org/10.1145/3241539.3267723.

[216] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, J. Ren, SDN/NFV-Empowered future IoV with enhanced communication, computing, and caching, Proc. IEEE 108 (2) (2020) 274–291, http://dx.doi.org/10.1109/JPROC.2019.2951169.

[217] Q. Ye, W. Shi, K. Qu, H. He, W. Zhuang, X. Shen, Joint RAN slicing and computation offloading for autonomous vehicular networks: A learning-assisted hierarchical approach, IEEE Open J. Veh. Technol. 2 (2021) 272–288.

[218] Q. Ye, W. Shi, K. Qu, H. He, W. Zhuang, X.S. Shen, Learning-based computing task offloading for autonomous driving: A load balancing perspective, in: ICC 2021 - IEEE International Conference on Communications, 2021, pp. 1–6, http://dx.doi.org/10.1109/ICC42927.2021.9500293.

[219] F. Bock, S. Siegl, P. Bazan, P. Buchholz, R. German, Reliability and test effort analysis of multi-sensor driver assistance systems, J. Syst. Archit. 85 (2018) 1–13.

[220] D.W. Ryu, H.B. Jeong, S.H. Lee, W.-S. Lee, J.H. Yang, Development of driver-state estimation algorithm based on hybrid Bayesian network, in: 2015 IEEE Intelligent Vehicles Symposium (IV), vol. 2015-Augus, no. Iv, IEEE, 2015, pp. 1282–1286, http://dx.doi.org/10.1109/IVS.2015.7225873.

[221] A. Koesdwiady, R. Abdelmoula, F. Karray, M. Kamel, Driver inattention detection system: A PSO-based multiview classification approach, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, vol. 2015-Octob, IEEE, 2015, pp. 1624–1629, http://dx.doi.org/10.1109/ITSC.2015.264.

[222] C. Craye, A. Rashwan, M.S. Kamel, F. Karray, A multi-modal driver fatigue and distraction assessment system, Int. J. Intell. Transp. Syst. Res. 14 (3) (2016) 173–194, http://dx.doi.org/10.1007/s13177-015-0112-9.

[223] C. Streiffer, R. Raghavendra, T. Benson, M. Srivatsa, DarNet: A deep learning solution for distracted driving detection, in: Proceedings of the 18th ACM/IFIP/USENIX Middleware Conference on Industrial Track - Middleware '17, ACM Press, New York, New York, USA, 2017, pp. 22–28, http://dx.doi.org/10.1145/3154448.3154452.

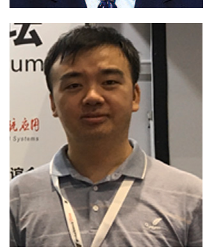
[224] N. Li, C. Busso, Predicting perceived visual and cognitive distractions of drivers with multimodal features, IEEE Trans. Intell. Transp. Syst. 16 (1) (2015) 51–65, http://dx.doi.org/10.1109/TITS.2014.2324414.

[225] S. Martin, M.M. Trivedi, Gaze fixations and dynamics for behavior modeling and prediction of on-road driving maneuvers, in: 2017 IEEE Intelligent Vehicles Symposium (IV), no. Iv, IEEE, 2017, pp. 1541–1545, http://dx.doi.org/10.1109/IVS.2017.7995928.

[226] J. Liu, R. Zhang, G. Han, N. Sun, S. Kwong, Video action recognition with visual privacy protection based on compressed sensing, J. Syst. Archit. 113 (2021) 101882.

Wanli Li received the B.Eng. degree in computer science and technology from Hunan University, China, in 2014. He is currently pursuing the Ph.D. degree with the Key Laboratory for Embedded and Network Computing of Hunan Province, Hunan University, China. His current research interests include machine learning, computer vision, intelligent transportation systems, advanced driver assistance systems, and driver behavior analysis.

Jing Huang received his Ph.D. degree in Computer Science and Technology from Hunan University, Changsha, China, in 2018. He is working as a Postdoctoral Researcher at Hunan University. His research interests include parallel computing, high-performance computing, distributed computing, energy-efficient computing, heterogeneous computing, cloud computing, machine learning.



Guoqi Xie is currently a Professor with Hunan University. He received the Ph.D. degree in computer science and engineering from Hunan University, China, in 2014. He was a postdoctoral research fellow with Nagoya University, Japan. His current research interests include embedded computing systems, real-time systems, cyber–physical systems (CPS), automotive electronics. He received the 2018 IEEE TCSC Early Career Researcher Award. He is currently serving on the editorial boards of the Journal of Systems Architecture, the Microprocessors and Microsystems, and the Journal of Circuits, Systems and Computers. He is an IEEE senior member and ACM senior member.

Fakhri Karray received the Ph.D. degree from the University of Illinois, Urbana-Champaign, Champaign, IL, USA. He is currently the University Research Chair Professor in electrical and computer engineering with the University of Waterloo, Waterloo, ON, Canada, and Co-Directs the University's Institute for Artificial Intelligence. His research interests are in the fields of intelligent systems design, big data analytic, soft computing, sensor fusion, and context aware machines with applications to intelligent transportation systems, Internet of Things, cognitive robotics. He has authored in these areas and has disseminated his work in journals, conference proceedings, and textbooks. He has also served as the Associate Editor/Guest Editor for various journals, including the IEEE Transactions on Cybernetics, the IEEE Transactions on Neural Networks and Learning, the IEEE Transactions on Mechatronics, and the IEEE Computational Intelligence Magazine. He is a Fellow of the Canadian Academy of Engineering and a Fellow of the Engineering Institute of Canada.

Renfa Li is a professor of computer science and electronic engineering, with the Hunan University, China. He is the director of the Key Laboratory for Embedded and Network Computing of Hunan Province, China. He is also an expert committee member of National Supercomputing Center in Changsha, China. His major research include computer architectures, embedded computing systems, cyber–physical systems, and Internet of things. He is a member of the council of CCF, a senior member of the IEEE, and ACM.