Identification and Analysis of Driver Postures for In-Vehicle Driving Activities and Secondary Tasks Recognition

Yang Xing, Chen Lv, *Member, IEEE,* Zhaozhong Zhang, Huaji Wang, Xiaoxiang Na, Dongpu Cao, *Member, IEEE,* Efstathios Velenis, Fei-Yue Wang, *Fellow, IEEE*

Abstract—Driver decisions and behaviours regarding the surrounding traffic are critical to traffic safety. It is important for an intelligent vehicle to understand driver behaviour and assist in driving tasks according to their status. In this study, the consumer range camera Kinect is used to monitor drivers and identify driving tasks in a real vehicle. Specifically, seven common tasks performed by multiple drivers during driving are identified in this study. The tasks include normal driving, left, right, and rear mirror-checking, mobile phone answering, texting using a mobile phone with one or both hands, and the setup of in-vehicle video devices. The first four tasks are considered safe driving tasks while the other three tasks are regarded as dangerous and distracting tasks. The driver behaviour signals collected from the Kinect consist of a colour and depth image of the driver inside the vehicle cabin. Additionally, three-dimensional head rotation angles and the upper body (hand and arm at both sides) joint positions are recorded. Then, the importance of these features to behaviour recognition is evaluated using Random Forests (RF) and Maximal Information Coefficient (MIC) methods. Next, a Feedforward Neural Network (FFNN) is used to identify the seven tasks. Finally, the model performance for task recognition is evaluated with different features (body only, head only, and combined). The final detection result for the seven driving tasks among five participants achieved an average of greater than 80% accuracy, and the FFNN tasks detector is proved to be an efficient model that can be implemented for real-time driver distraction and dangerous behaviour recognition.

Index Terms—Driver Behaviour, driver distraction, Kinect, Random Forest, Feedforward Neural Network.

I. INTRODUCTION

A. Motivations

D RIVER behaviour is the most important factor for on-road driving safety [1-6]. Since humans are the major users of

Y. Xing is with Advanced Vehicle Engineering Centre, Cranfield University, Bedford, MK43 0AL, UK, and also with Qingdao Huituo Intelligent Machine Company, Qingdao, China (email: y.xing@cranfield.ac.uk)

C. Lv, Z. Zhang, H. Wang, D. Cao and E. Velenis are with Advanced Vehicle Engineering Centre, Cranfield University, Bedford, MK43 0AL, UK (e-mail:{c.lyu,zhaozhong.zhang,huaji.wang,d.cao,e.velenis}@cranfield.ac.uk). (Corresponding author is C. Lv.)

X. Na is with the Department of Engineering, University of Cambridge, CB2 1PZ, United Kingdom (e-mail: xnhn2@eng.cam.ac.uk)

F.-Y. Wang is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P.R. China (e-mail: feiyue@ieee.org).

roads, their driving behaviours influence traffic safety and efficiency. More than 90% of traffic accidents for light-vehicles in the US were reported to be caused by driver errors such as misbehaviour and inadvertent errors, which is similar to other countries worldwide. It was also mentioned in [7-11] that traffic accidents could be reduced by 10% to 20% by correctly recognizing driver behaviours. Therefore, it is critical to have a clear perspective of driver behaviour and the tasks being performed.

Human drivers have been extensively studied since the 1970s. The study of human drivers is a massive project with many aspects. Most of the existing research lies in the scopes of driver behaviours, driver attention and intention, driver drowsiness and fatigue, driver cognitive and neural muscles, etc. All of these studies have a common objective, which is to gain a better understanding of driver status from either a psychological or physiological aspect so as to assist in driving tasks and increase driving safety [12-14].

Understanding human drivers is necessary both for conventional vehicles and for automated vehicles. In the US and China, accidents have occurred when a Tesla driver trusted or solely relied on the autopilot system while driving. For lower level automated vehicles, especially for level two and level three automated vehicles (based on the automation definition in SAE standard J3016), human drivers need to sit in the driver seat and are responsible for the safety issues. In these vehicles, the driver is allowed to perform secondary tasks for entertainment; however, due to the partially automated limitation, the driver has to take control in emergencies. Therefore, the monitoring of human drivers and determining whether they can return to the driving task is more important than in conventional vehicles.

In this study, a driver monitoring system is designed to detect driving and secondary tasks in real time. Specifically, the recognition model is designed to identify seven tasks performed by different drivers. There are four tasks considered as normal driving tasks: normal driving (front looking), right mirror checking, left mirror checking, and rear mirror checking. Meanwhile, according to [13], the three most common secondary tasks in automated vehicles are selected, which are using a video device, answering a mobile phone, and texting

using a mobile phone. To identify the driver postures, multimodal data is collected using a Kinect consumer RGB-D camera including the head rotation and body joint positions. The main objective of this study is to design a real-time driver behaviour model that does not require any history information for the recognition of normal driving and secondary driving tasks. Additionally, the importance of driver posture features to the identification of driving tasks is evaluated.

B. Related Works

In this study, the research scope is narrowed to the range of driving task recognition towards a normal driving and secondary task monitoring system for lower level and middle level automated vehicles. According to previous studies, driver behaviour can be classified into intended and non-intended behaviours [15-45]. The intended behaviour of the driver is the extension of the driver's mental thought, which can be used to infer the mental state and intent of the driver. In contrast, non-intended behaviours are usually caused by distractions due to outside and inside disturbances. Driver behaviour has been widely studied in previous literature. General driver behaviours include the study of driver head pose [15-16], eye gaze dynamics [17-18], hand motions and gestures [19], body movement [20-21], and foot dynamics [22]. This behaviour information has been successfully used to estimate driver fatigue, driver distraction, driver attention, etc. In this study, driver head and upper body information detected using a Kinect will be evaluated for normal driving and distraction identification.

When drivers are performing secondary tasks while driving, they are regarded as being distracted and many studies use the duration of eye-off-road to detect whether a driver is distracted by the secondary tasks. Therefore, the most common features for driver distraction detection are head pose and eye gaze information. Along with the driver behaviour, information of the vehicle such as vehicle speed, heading, and acceleration are important features for evaluating the level of driver distraction. In [23], an integration method combining the driver's hand, head, and eye for driver activity recognition was proposed. Rezaei and Klette introduced an intelligent driver assistance system to prevent rear-end crashes based on driver monitoring and front vehicle detection [24]. The head pose was estimated based on the proposed face appearance model and 3D head model mapping. In [25], a driver drowsiness alert system was proposed according to the driver head and eye dynamics. The driver head pose was estimated based on an Euler angle comparison between a single head region image and a 3D head model with known rotations. In [28], the authors analysed the relationship between head pose and eye gaze. A strong correlation was found between the head and gaze direction. The study showed that during natural driving, the participants tend to have less head rotation but more gaze searching to maintain safe driving.

In [45], a comprehensive in-vehicle perception system for driver surveillance and assistance was proposed. Multi-modal sensors were fused to integrate the major driver physical cues and traffic situations. In [29], driver acceleration profiles for a car following scenario on a highway were generated using recurrent neural networks. Specifically, a long short-term memory (LSTM) recurrent network was adapted since it can automatically learn the spatial and temporal features of the naturalistic driving data. In [30], a LSTM based recurrent neural network was proposed to detect driver distraction behaviours based on the simulated CAN bus signals. Thirty participants performed eight typical secondary tasks independently and the distraction levels were classified into binary, three levels, and six levels. In [31], the authors claimed that applying eye tracking is much more difficult in real vehicles than in the simulator. Therefore, gaze estimation was not adopted and only driving information through the CAN bus was used for driver visual searching distraction detection. In [32], a driving behaviour model for teenage drivers was studied. Different machine learning methods were evaluated based on the driving data, which was collected with a driving simulator. The authors reported that instead of predicting driving behaviour (steer, throttle, and brake) directly, more accurate results can be achieved using context-based prediction and indirect prediction methods.

Despite the driver and driving behaviours, other studies have used physiological sensors to identify driver distraction and other abnormal statuses. According to the study in [35], driver monitoring systems for drowsiness and distraction detection can be classified into visual-based and non-visual-based methods. Visual-based methods monitor driver head pose, eye movement and blinking, yawning, and facial expression. In contrast, non-visual based systems detect driver status with physiological sensors such as EEG, ECG, and EOG, along with the vehicle CAN bus signals. However, the effects of hand, arm, and body on the recognition of driver status were not discussed. Similarly, a stress detection system for drivers was studied in [34]. A specific type of continuous recurrent neural network named cellular neural network was used for the binary classification task.

The Kinect sensor, a low-cost range camera, has been successfully applied to human and driver behaviour detection since it was first made available by Microsoft in 2012. Kinect was first designed for indoor motion sensing and provides a colour image, depth image, and infrared image. In [36], the general architecture for human activity recognition was proposed using Kinect. The human activities were viewed as the spatiotemporal evolution of body postures. The estimated postures are classified using support vector machines, and finally, the HMM was used to model the activities as a time sequence of the different estimated postures. In [37], a Kinect-based wearable face recognition system for people with low-vision or blindness was proposed. The colour and depth images were simultaneously captured to identify the face and generate the 3D location for the user. In [27], a seven-point skeleton based driver upper body tracking system using Kinect depth images was applied. The proposed system is efficient for detecting driver merging and turning behaviours according to the detected body pose and arm motion. The system can also be used to analyse and compare the driving manoeuver styles of different drivers.

In this study, Kinect is adopted as the driver monitoring sensor to identify normal driving and secondary tasks. Similar research can be found in [26], where driver mirror-checking behaviour during normal driving and performing secondary tasks were analysed. The authors reported that mirror-checking

behaviour is one of the most important driving perception processes and reflects the attention level of the driver. In addition, mirror-checking behaviours are highly detectable manoeuvres and can achieve 95% detection accuracy using machine learning. However, that work only studied the binary classification scenarios without reporting the recognition accuracy for each task. Additionally, that study did not analyse the impact of body postures to the recognition of complex driving behaviours.

C. Contributions

In this study, driver head and body posture information is used for driving and non-driving related task recognition. The contributions of this study are threefold.

First, a driver posture detection method using a Kinect, a consumer range camera, inside a vehicle is introduced. The data characteristics of Kinect are analysed and the data processing technique for in-vehicle application is proposed. In addition, the head rotation signals from the Kinect are calibrated with a precise orientation sensor.

Second, the importance of head and body features to task prediction is estimated using an integrated algorithm. The feature importance estimation given by RF and MIC are compared and integrated. Then, the most important posture features for task recognition are determined. Unlike previous studies that use time sequence data for driver behaviour recognition [29, 30], this study focuses on identifying behaviour in a more natural way, only based on the instance samples. The objective is to design a human-like task detector that can identify driver behaviours according to a single image. Therefore, a FFNN model is evaluated and compared with multiple machine learning methods.

Finally, quantity analyses of the impact of the driver's head and body features to driver task recognition are performed, and the predicted important posture features are evaluated separately. Since the existing literature seldom considers the driver's body features, this study quantitatively proved that head and body features are required for driver behaviour recognition.

D. Paper Organization

The remainder of the paper is organized as follows. Section II introduces the experiment setup, data collection and processing methodologies. Section III proposes the method for feature importance prediction based on the integrated method and FFNN for task recognition is discussed. Then, the task recognition results and driver posture feature evaluation are performed in Section IV. Section V presents the results discussion and future work. Finally, the study is concluded in Section VI.

II. EXPERIMENT DESIGN AND DATA ANALYSIS

A. System Architecture

The procedure taken to construct a behaviour recognition model is described in this section. The driver monitoring system architecture is shown in Fig. 1. The general structure of this study consists of three parts. First, the driver head and body data are collected and time stamped. Then, the signals are smoothed and noise is filtered. Second, feature importance prediction is

proposed using a combination of Random Forests and MIC, the feature importance given by the two algorithms show strong consistency. The 'model selection' block in Fig. 1 processes the feature evaluation based on the feature importance provided by the 'feature importance estimation' block. Meanwhile, the influence of depth, head, and body features to the driver status detection will be studied. Then, real-time driver behaviour identification will be conducted using a FFNN model with *leave-one-out (LOO)* cross-validation. Finally, the performance with different features is analysed and a behaviour classification performance comparison between different algorithms will be proposed.
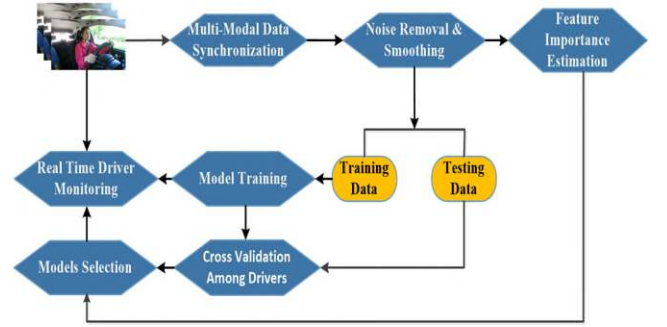

Fig. 1. Proposed driver task recognition architecture.

B. Experiment Setup and Data Collection

In this part, the experiment setup and data collection methods are introduced. Driver behaviour data is collected using the low-cost range camera Kinect, which was developed by Microsoft. In this study, the second version of Kinect (V2) was adopted. Kinect is a consumer camera that supports colour images, depth images, audio, and infrared information. It was first designed for indoor human interaction with computers and has been successfully applied in vehicles for driver monitoring [36-37].

Kinect supports tracking the head and the body skeletons of as many as six individuals. In this work, the head and upper body joint detection functions are integrated for collecting driver head and body signals. The head detection provided by the Kinect requires tracked body information. Therefore, to use the Kinect inside a vehicle, it must be mounted above the dashboard to have full vision of the driver's body. Considering the mounting requirements in [38], the Kinect is mounted in the middle of the dashboard, facing the driver, which does not interfere with the driver's field of view and allows for monitoring of the driver's entire upper body. Fig. 2 illustrates the detected head centre and upper body joints using Kinect and an example depth image.

Fig. 2. The left image shows the head and body joints for the seated driver detected using the Kinect sensor. The right image shows the depth image given by Kinect.

In this study, the head and body signals and colour and depth images are collected and synchronized with a time stamp. The sampling rate is eight frames per second. The data is sampled with an Intel ® Core i7 2.5 GHz computer and the code is written in C++ based on the Windows Kinect SDK and OpenCV. The size of the colour image captured using a Kinect is 1920 x 1080. However, to increase computational efficiency, the stored colour image was compressed to 640 x 360. According to [26], short-term driver mirror checking actions last from 0.5 s to 1 s. Therefore, the sampling frequency is fast enough to capture normal driver actions and behaviour. The three-dimensional head rotation vector contains yaw, pitch, and roll angles. The upper body joints are recorded using X and Y coordinates in the colour image and the corresponding depth value in the depth map. The 42 signals collected are shown in Table 1.

TABLE 1
MULTIMODAL FEATURES GIVEN BY KINECT

Multimodal Features from Kinect (42 Features)		
Head (12 Features)	Head Pitch Angle (Pitch) Head Roll Angle (Roll) Right Eye (X, Y, Z)	Head Yaw Angle (Yaw) Left Eye (X, Y, Z) Nose (X, Y, Z)
Body (30 Features)	Left Hand (X, Y, Z) Left Wrist (X, Y, Z) Left Elbow (X, Y, Z) Left Shoulder (X, Y, Z) Left Hand Tip(X, Y, Z)	Right Hand (X, Y, Z) Right Wrist (X, Y, Z) Right Elbow (X, Y, Z) Right Shoulder (X, Y, Z) Right Hand Tip(X, Y, Z)

C. Data Processing

The Kinect data processing methodologies used in this study are described in this section. The two data processing steps are head rotation calibration with an orientation sensor and noise removal and smoothing based on a combination of a median filter and an exponential filter.

1) Kinect Head Rotation Data Calibration

In [39], Kinect head rotation data were evaluated and compared with a high-precision head rotation detection device. The author concluded that the average errors in absolute yaw, pitch, and roll angles were $2.0 \pm 1.2°, 7.3 \pm 3.2°$, and $2.6 \pm 0.7°$, respectively. However, the experiment and data calibration were proposed for indoor environments in standard conditions. However, in this work, the Kinect V2 was implemented inside a vehicle, which is a more challenging environment. During the experiment, the Kinect detection signals inside the vehicle have more noise and are less stable than the signals collected inside the room. Therefore, the first step was to calibrate the Kinect head rotation data with a high-precision head rotation sensor. Since driver head rotation is a very important signal for determining the driver's attention and

distraction status, only the head rotation signals were evaluated in this study and the detected body positions provided by the Kinect were not calibrated.

To calibrate the estimated head rotation results of the Kinect, a head-mounted head tracker was used and three-degree rotation data from the head tracker was used as the ground truth. The head tracker is based on an Arduino microcontroller board and an intelligent nine-axis absolute orientation sensor (BNO055) designed by BOSCH. The sampling frequency of the orientation sensor is up to 100 HZ. The rotation sensor and Arduino data-recording sensor are fixed on a head-mounted harness belt strap, as shown in Fig. 3. And seven driver behaviours studied in this research are shown in Fig. 4.

Fig. 3. The left image illustrates the Kinect setup inside the vehicle and right image shows a participant seated in the passenger seat wearing the head tracker.

The Kinect sensor is mounted in the middle of the front dashboard. The optical axis of the Kinect camera is not perpendicular to the yaw axis of driver's head, which will influence the detected yaw angles. The rotation angle of the Kinect sensor in world coordinates is reflected by a constant bias of the detected yaw angle, as shown in Fig. 5. The blue line is the original yaw angle. The yellow line is the shifted yaw angle, which shifts the original signal by a constant offset ($30°$). The red line shows the ground truth results of the head tracker. The calibrated Kinect signal and ground truth have similar variations, which means that the head rotation angle detected by Kinect is reliable and can be used for further analysis.

The data recording frequency for the head tracker is 30 Hz, which is approximately three times greater than Kinect, therefore, the head tracker yaw angle shown in Fig. 5 is the smoothed version of the original signal. Finally, the mean error and standard deviation between the calibrated Kinect signal and the head tracker for yaw, pitch, and roll angles are $1.93 \pm 11.55°, 1.47 \pm 5.98°$, and $1.44 \pm 6.98°$, respectively.

2) Noise Removal and Data Smoothing

The temporal spikes due to noise can cause more serious problems. The body and head detection results using the Kinect can be influenced by lighting conditions or the location and distance to the driver and human gesture or body pose can influence joint detection, especially inside the vehicle. Due to

Fig. 4. Seven driver behaviours studied in this research. The first row shows normal driving, right-mirror checking, rear-mirror checking, left-mirror checking, and using video devices. The second row shows mobile phone texting and answering a mobile phone performed by multiple drivers.
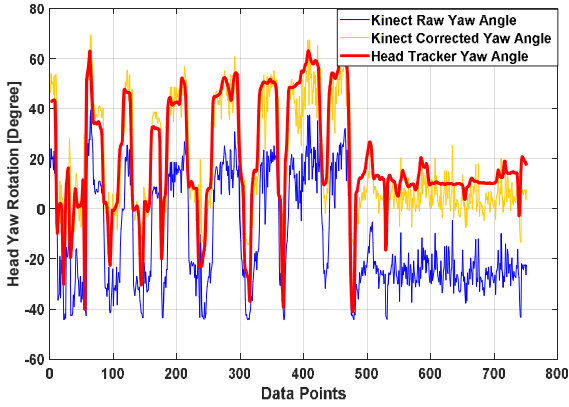

Fig. 5. Head yaw angle detection results given by Kinect and the orientation-based head tracker.
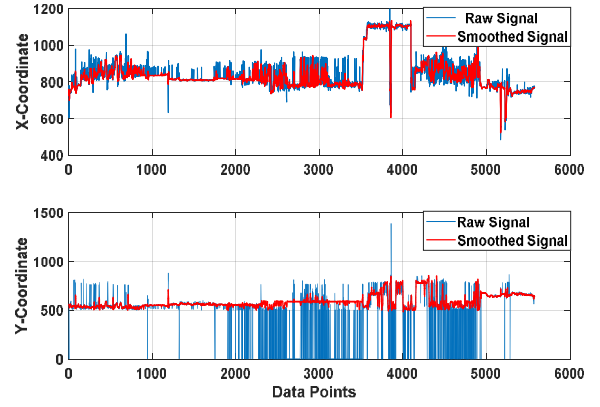

Fig. 6. Right wrist signal smoothing results. The upper image is the x-coordinates of the image. The bottom image shows the y-coordinates of the image.

the less precise detection results using the Kinect, an integrated signal process scheme combining two different filtering techniques is adopted in this study.

Specifically, an abnormal data removal and exponential smoothing filter are applied to the raw signals to smooth and track the detection results.

$$\hat{x}_n = \begin{cases} x_n, & x_n \neq 0 \\ mean(X_{pre}), & x_n = 0 \end{cases} \quad (1)$$

where $\hat{x}_n$ is the filtered data value, x_n is the raw data, and X_{pre} represents all the non-zero data before step n. The exponential smoothing filter is defined as (2):

$$\begin{cases} s_0 = x_0 \\ s_t = \alpha \sum_{i=0}^{W}(1-\alpha)^i x_{W-i} \end{cases} \quad (2)$$

where s_t is the smoothed version of raw signal x_t, W is the sliding window size that depends on the number of previous inputs used for smoothing, α is called the dampening factor, which controls the weight of previous inputs and $0 \leq \alpha \leq 1$.

As shown in Fig. 4, the driver's right arm is partially blocked by the steering wheel, which causes the Kinect to detect inaccurate body joints. During the data recording process, some data points will be lost or unreasonable due to the driver pose, lighting conditions, or the Kinect algorithms. First, these data points are recorded as zeroes to indicate abnormal detection status. Then, the data are fed into the hierarchical filter module to smooth the original signals. To track the signals, an abnormal data removal algorithm is applied. The zero data points are replaced by the mean value of the non-zero data. Then, the exponential smoothing filter is applied to further smooth the noisy signal. Fig. 6 shows the smoothing result of right wrist signal.

III. EVALUATION AND IDENTIFICATION ALGORITHMS DESIGN

In this section, driver feature evaluation is proposed to study the relationship between driver features and driver behaviour estimation. The most relevant features for driver behaviour recognition are detected. Then, a feedforward neural network is adopted as the driver behaviour classifier to identify the driver actions based on the selected feature vectors.

A. Feature Importance Evaluation Using RF and MIC

For some machine learning tasks, feature vector dimensions can be very high (hundreds or thousands, or even larger). Although machine learning methods are particularly suitable for modelling large datasets, they are always viewed as a black-box where it is difficult to analyse the intrinsic structure. Therefore, it is important to understand how the input features

influence or are associated with the model output. In this study, to understand how the driver signals influence behaviour detection, the relationship between body signals and driver behaviour is analysed. Such feature evaluation and selection enable subjective understanding of the relationship between driver body signals and behaviour.

Feature selection is a major research area of feature engineering. By selecting a subset of feature vectors, machine learning models can be trained more efficiently and better results can be obtained. In this section, to understand how driver features influence the corresponding behaviour detection and which features are important for the behaviour recognition task, two distinct feature selection methods are applied and compared. First, a random forest was used to estimate the driver feature importance with an out-of-bag (OOB) dataset. Second, maximal information coefficient (MIC) is used as another indicator for the association between features and the behaviour class. The final conclusion of feature importance will be summarized according to the results given by these two distinct algorithms.

1) RF for Feature Importance Estimation

Random forests, introduced by Breiman in 2001, were built on classification and regression trees [40]. It has proven to be a powerful machine-learning tool for many applications: In [41], the author evaluated the RF classification performance on 121 public datasets and the RF algorithm achieved the best classification result among 179 algorithms. RF is an ensemble learning machine that integrates multiple decision trees. One decision tree is constructed with one root node and multiple middle leaf nodes. The prediction ability for a single tree is limited, and given a large dataset, overfitting is common for a single decision tree. According to the drawbacks of a single decision tree, RF combines multiple decision trees and uses average or voting schemes to calculate the final results.
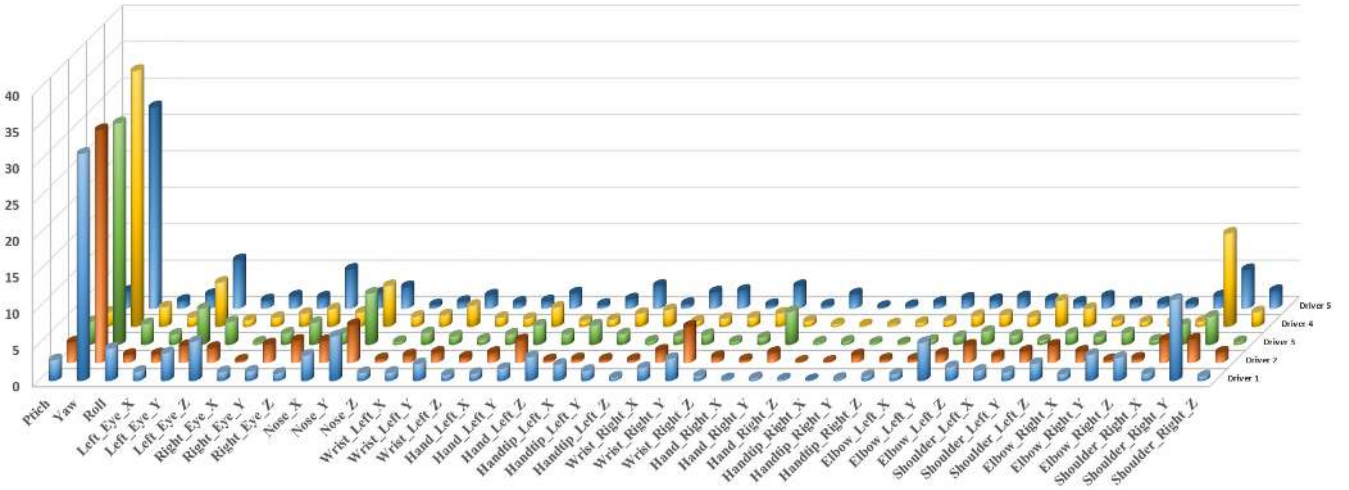

Fig. 7. Feature importance prediction using random forests based on the permutation method. 42 head and body features of five participants are evaluated.

To increase the diversity of each tree in the forest, RF is trained using a bootstrap aggregating (Bagging) technique. Specifically, the number of trees B in the RF is selected. Then, according to this number, B separate training datasets are chosen from the original dataset. Since Bagging is a random sampling technique with replacement, approximately one-third of the data is not used for training each subtree. The remaining dataset for each tree is the OOB dataset. Normally cross-validation is not necessary for training RF since the OOB can be used to evaluate the model performance by evaluating the OOB errors [40]. Moreover, the OOB dataset can be used to evaluate the feature importance for model accuracy. To obtain the feature importance, for each variable X_i, the variable is randomly permuted. The feature importance is calculated as follows:

$$I(X_i) = \frac{1}{B}\sum_t^B \widetilde{OOBerr}_{t^i} - OOBerr_t \tag{3}$$

where X_i is the permuted i^{th} feature in the feature vector X, B is the number of trees in the random forest, $\widetilde{OOBerr}_{t^i}$ is the model prediction error of the perturbed OOB sample with the permuted feature X_i for tree t, and $OOBerr_t$ is the untouched OOB data sample with permuted variable.

The concept of permutation feature importance is that a large importance value indicates the feature is influential in the prediction and permuting the feature value will influence the model prediction. In contrast, a small influential feature will have no or less impact on the model prediction. The predicted feature importance for the 42 driver signals using RF are illustrated in Fig. 7. From the importance estimation results, the driver yaw angles are extremely important for action classification for all five drivers. To verify the prediction results given by RF, the next section proposes another feature evaluation technique called the maximal information coefficient, which uses a completely different method to estimate feature importance.

2) MIC for Feature Importance Estimation

The MIC is designed to efficiently solve the mutual information estimation problem for continuous variables and continuous distributions. The MIC provides an equitable measurement for the linear or nonlinear strength association between two variables. The MIC introduced a maximal mutual information searching technique by varying the grid that drawn on a scatterplot of two variables [42]. Mutual information usually can be used to evaluate the mutual dependence between different variables and assess the amount of information the two

variables share, or more generally, the correlation between the joint distribution of the two variables and the product of the independent distribution of the two variables [43]. The mutual information for two discrete vectors is defined as:

$$MI_D(X,Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log(\frac{p(x,y)}{p(x)p(y)}) \qquad (4)$$

where MI_D is the mutual information of two discrete vectors, $p(x,y)$ is the joint probabilistic distribution of x and y. $p(x)$ and $p(y)$ are the marginal probability distribution functions of x and y, respectively. For continuous variables, the mutual information format is slightly changed to:

$$MI_C(X,Y) = \iint p(x,y) \log(\frac{p(x,y)}{p(x)p(y)}) dxdy$$

$$(5)$$

where MI_C is the mutual information for two continuous vectors, and $p(x,y)$, $p(x)$, and $p(y)$ represent the corresponding probabilistic density functions.

As shown in (2), calculating the mutual information of continuous variables is difficult. Therefore, the maximal information coefficient technique, which concentrates on the optimal binning method, is applied to assess the mutual information of the continuous case. Meanwhile, MIC enables the mutual information score to be normalized into the range [0,1], which makes assessing the dependency and co-relationship between two variables more convenient. In this study, in addition to the first three continuous head rotation angles, the remaining features are discrete image coordinates and depth values. Therefore, the MIC can be efficiently used for feature association predication.

3) Comparison of the Feature Importance Prediction

To evaluate the prediction results of feature importance using the two algorithms, the ten most important features for each subject are extracted and compared. Specifically, for each driver, the ten most important features are selected. Then, five selected feature vectors are fused into 42 bins and the count in each bin represents the number of occurrences for each feature of the five subjects. Therefore, the highest value, 5, indicates the feature is the one of the ten most important features for all five drivers. The statistical results are shown in Fig. 8.
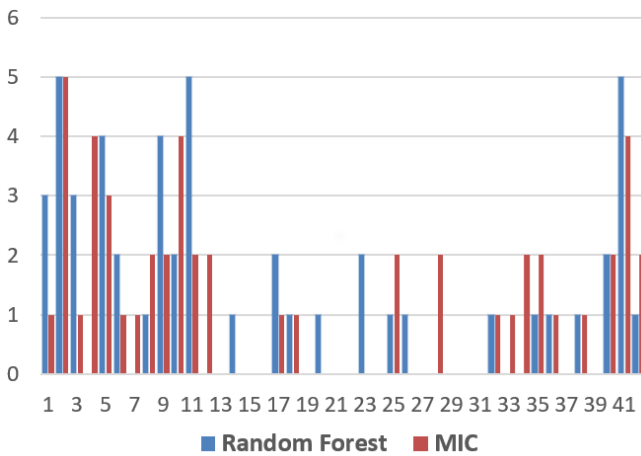

Fig. 8. Feature importance prediction results using random forest OOB permutation (blue bars) and maximal information coefficient methods (red bars).

As shown in Fig. 8, although the prediction results of the two algorithms are not identical, there is some consistency in the results of the two algorithms. For example, the driver yaw and y-coordinate of the right shoulder features (No. 2 and No. 41) are both significant. According to Fig. 8, the 12 most important features (marked as the ten most important features by at least two drivers) are listed in Table 2.

TABLE 2
FEATURE IMPORTANCE ESTIMATION RESULT

	Importance Order	Features
RF	1-5	Yaw, Left_eye_Y, Red_eye_Y, Nose_Y, Right_shoulder_Y
	6-12	Pitch, Roll, Left_eye_Z, Nose_X, Left_hand_Y, Right_wrist_Y, Right_shoulder_X
MIC	1-5	Yaw, Left_eye_X, Left_eye_Y, Nose_X, Right_shoulder_Y
	6-12	Right_eye_Y, Right_eye_Z, Nose_Y, Nose_Z, Right_hand_X, Left_shoulder_X, Right_shoulder_X

According to Table 2, the importance predictions given by RF and MIC are similar. The most important features are the head rotation angles (yaw, pitch, and roll), eye and nose position, shoulder position, and hand position. The remaining features such as the wrist, hand tip, and elbow positions are less likely to influence the behaviour detection result. A quantitative analysis of the feature impact on behaviour recognition based on a feedforward neural network is proposed in the next section.

B. Feedforward Neural Network for Driver Behaviour Classification

In this section, an Artificial Neural Network (ANN) is used for driver behaviour pattern recognition. Specifically, a one-way feedforward neural network (FFNN) is adopted. The FFNN passes the input vectors to the output layer-by-layer without any feedback connections. The FFNN is a powerful tool for solving complex nonlinear mapping problems. By learning the neuron parameters and the connection width, the FFNN model is able to construct a nonlinear mapping between the input and output. The FFNN can be approximately represented as follows:

$$y = f(X, \boldsymbol{\theta}) + \epsilon \qquad (6)$$

where y is the output of FFNN, $f()$ is the learned model mapping function with model parameter $\boldsymbol{\theta}$, X is the input data vectors, and ϵ is the bias between the actual output and the target.

For the FFNN, parameter $\boldsymbol{\theta}$ represents the set of activation function parameters and the width set between neurons. In this study, a two-layer FFNN with one hidden layer is used to train the driver behaviour recognition model. The sigmoid transfer function in the hidden layer is chosen. The sigmoid activation function for a single neuron is represented as:

$$f = \frac{1}{1+e^{-X}} \qquad (7)$$

where f is the neuron output and X is the neuron input, which has the following form:

$$X = \sum_{i=1}^{N} \omega_i x_i + b \qquad (8)$$

where ω_i is the weight of the ith input, and normally each neuron has a bias parameter b. An important reason for using

the sigmoid activation function is the computation efficiency during model training. The backpropagation requires the derivative of the neuron transfer function to be calculated. While the sigmoid function has the convenient derivative form [44]:

$$f' = f \cdot (1 - f)$$
(9)

Although the sigmoid function will cause the loss of a gradient problem in most scenarios, it is not a serious problem in this shallow network case. In this case, the supervised FFNN is trained with driver head and body signals as the input and an output of the corresponding behaviour among the seven actions. Unlike some existing research that uses time-series models, the FFNN used in this study does not consider the previous step status of the driver. The reason for this is that humans can normally distinguish the current driving behaviour using one image and do not require video sequences. Unlike the inner mental states of the driver, which is a long-term process and depends on previous states, the outer behaviours can be considered a transient state and are not highly dependent on prior information. Therefore, the FFNN is applied to detect the driving tasks frame-by-frame based on the collected driver body information.

Since time information is not considered in the model construction procedure, the training and testing dataset are reordered randomly. For model training, cross-validation is used. Specifically, the leave-one-out (Loo) method is adopted. For the five-driver dataset, data of four drivers are used for model training and validation, the data of the remaining driver is used to test the classification performance. The general classification accuracy is the average of the five classification results. Another hyper-parameter for FFNN is the number of neurons in the hidden layer. To evaluate the influence of neuron quantity on classification performance, different neuron numbers and cross-validation are studied. A boxplot of the classification results is shown in Fig. 9. The neuron numbers vary from 10 to 100 with an interval of 10. The red line represents the mean accuracy of the five drivers with different neurons. As shown in Fig. 9, variation in the number of neurons does not significantly influence performance. The most accurate detection occurs at the 100-neuron cases, with an accuracy of approximately 81.2%. In the next section, more detailed statistical results using FFNN with 60 neurons are proposed, and the results are compared with multiple machine learning methods.
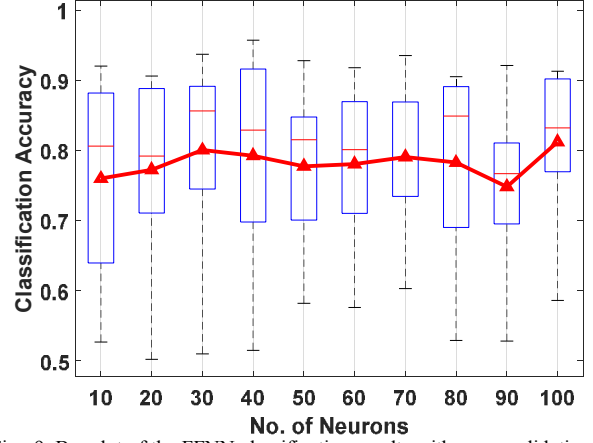
Fig. 9. Boxplot of the FFNN classification results with cross-validation of the neuron numbers in the hidden layer. The red line represents the average classification results of the five drivers with different neuron numbers.

IV. EXPERIMENT RESULTS AND ANALYSIS

In this section, the task recognition results are discussed. Specifically, task classification with FFNN is compared with other machine learning methods. In addition, the impact of the head, body and depth information on the classification results will be evaluated separately in part two.

A. Behaviour Recognition Results

In this section, the identification accuracy for driving and non-driving tasks is analysed. As mentioned in the previous section, the classification model is trained using the LOO-cross validation method. The prediction results for the five drivers are illustrated in Table 3. The first four mirror checking tasks are divided into driving-related tasks, while the remaining three tasks are divided into non-driving and distraction tasks.

TABLE 3
CLASSIFICATION RESULTS USING FFNN WITH ENTIRE FEATURES

	Driving Tasks				Non-Driving Tasks			
	T1	T2	T3	T4	T5	T6	T7	Ave
D1	0.905	0.856	0.843	0.925	0.686	0.896	1.00	0.883
D2	0.380	0.557	0.498	0.985	0.877	0.617	0.684	0.630
D3	0.985	0.976	0.690	0.998	0.988	1.00	0.662	0.898
D4	0.720	0.973	0.994	0.999	1.00	1.00	0.858	0.927
D5	0.583	0.977	0.801	1.00	0.798	0.969	0.991	0.871
Mean	0.715	0.838	0.747	0.981	0.867	0.884	0.838	0.824

The seven driver tasks are ordered as {Normal Driving, Right Mirror-Checking, Rear Mirror-Checking, Left Mirror-Checking, Using Video Device, Texting, and Answering Mobile Phone}.

As shown in the far right column of Table 3, the average classification result (Ave) for each driver is defined as the average of the seven tasks. The mean values shown in the bottom row represent the average classification accuracy for each task of the five drivers. Detection results equal to 1.00 shown in Table 3 indicate an accuracy of 100%. The FFNN classification model is trained with 60 neurons using the entire

feature vector (42 features). The classification results for driver 2 are much lower than the other four drivers, with an average of only 0.630. This is due to imprecise detection of the driver skeleton during data collection. To have a clear perspective of detection performance, the confusion matrix for driver 2 is shown in Fig. 10.

Confusion Matrix

Fig. 10. Confusion matrix of driving tasks classification results for driver 2. The bottom row illustrates the classification accuracy with respect to the ground truth label and the far right column illustrates the classification results with respect to the prediction value.

In the confusion matrix, the green diagonal shows the number of correct detection cases for that class. The bottom row shows the classification accuracy with respect to the target value, and the far right column shows the classification accuracy with respect to the predicted labels. As shown in Fig. 10, the normal driving behaviour for driver 2 only achieved 38% detection accuracy and 289 cases are classified into the phone answering task. This is mainly due to the similar postures between normal driving and phone answering behaviour. Once hand detection is inaccurate, it is very difficult to classify these two tasks only according to head pose. Detailed discussion will be proposed later. In addition, the low detection accuracy means the trained model using the other four drivers is not sufficient to precisely recognize all the behaviours for driver 2 due to the diversity of the drivers. However, once driver 2 is included in the training data, the model will obtain better detection results for the other four drivers. The most accurate detection occurs for driver 4, the relative results are shown in Fig. 11.

As shown in Fig. 11, the classification results for the seven tasks for driver 4 are much better than for driver 2. False detection between different classes decreased significantly. Similar results are achieved for the remaining three drivers. In conclusion, although very accurate results were not achieved for driver 2 compared with the other drivers, the general classification accuracy for the seven tasks was 82.4% (the mean value of the average column), which indicates efficient classification results.

In Table 4, the classification results of FFNN are compared with four other machine learning methods, which are random

forest (RF), support vector machine (SVM), naïve Bayes (NB), and K-nearest neighbour (KNN, K equals 5 in this case).

Confusion Matrix

Fig. 11. Confusion matrix of driving task classification results for driver 4.

The accuracy in Table 4 is defined as the average detection result for the five drivers, i.e., the average of the Ave column in Table 3. Meanwhile, to evaluate the driver distraction detection performance, the seven classification tasks are merged into a binary classification. Here, the negative group is defined as the combination of the first four normal driving tasks, and the true distraction group consists of the remaining three distracted driving tasks.

TABLE 4
CLASSIFICATION RESULTS USING DIFFERENT MACHINE LEARNING METHODS

	Accuracy	TPR	FPR	Training Cost [s]	Testing Cost [s]
FFNN	0.824	0.939	0.088	4.92	0.05
RF	0.736	0.900	0.144	33.55	0.41
SVM	0.747	0.913	0.177	2.85	0.03
NB	0.767	0.922	0.171	0.188	0.02
KNN	0.623	0.771	0.090	0.049	1.08

The TPR and FPR in Table 4 represent the true positive rate (sensitivity) and false positive rate, respectively. TPR and FPR are calculated as:

$$TPR = \frac{TP}{P} \tag{10}$$

and

$$FPR = \frac{FP}{N} \tag{11}$$

where TP is the number of correctly detected distracted cases. P is the total number of distracted cases, which is the total quantity of the three distracted cases. FP is the number of false detections. In this case, it represents the number of normal driving tasks that are classified in the abnormal driving group. Finally, N is the total amount of normal driving cases.

According to Table 4, FFNN binary classification outperforms the other four models, indicating that FFNN is a

powerful model suitable for driver behaviour modelling. Note that there are no optimization algorithms used in the other four models. These models are used with their default setup in MATLAB. The RF is constructed with 100 decision trees, and SVR uses a radial-based kernel. Better results may be obtained with parameter tuning and optimization, however, this is beyond the scope of this study. The binary classification model is able to distinguish normal driving behaviour and distracted behaviour. From the perspective of safety, although it may annoy the driver, it is safe to classify normal driving behaviour into distracted behaviour and warn the driver. On the other hand, if the model classifies distracted behaviour into the normal driving group, it is more dangerous than the previous case and this misclassification should be avoided. In the real world, in terms of non-driving tasks, the time constants are always much longer than normal driving tasks, texting or answering a phone can last for a few minutes. However, the mirror-checking actions usually last for one to two seconds. These time-properties of the different tasks can be adopted to predict the correct states in the future.

B. Feature Evaluation for Behaviour Classification Performance

In this section, the impact of the driver's head and body features on driving task classification will be analysed. The feature evaluation is divided into three parts. First, the depth information of the detected joints and facial landmarks (eyes and nose) are evaluated. Then, task classification using only head signals or only body signals is proposed. The classification results for these three parts are illustrated in Table 5.

TABLE 5
TASKS CLASSIFICATION BASED ON DIFFERENT FEATURES

		Driving Tasks				Non-Driving Tasks			
		T1	T2	T3	T4	T5	T6	T7	Ave
2D Only (29)	D1	93.8	48.1	60.5	88.0	100	91.7	70.4	79.4
	D2	66.8	79.6	41.4	88.2	62.7	84.6	37.1	62.7
	D3	58.7	64.4	100	100	97.8	96.0	48.9	77.6
	D4	69.3	73.1	96.9	99.7	99.1	96.7	91.0	87.6
	D5	68.5	54.3	39.3	100	32.7	95.5	99.5	71.1
	Mean	71.4	63.9	67.6	95.2	78.5	92.9	69.4	75.7
Head Only (3)	D1	78.9	92.3	99.7	96.8	5.4	24.9	0.7	61.8
	D2	0.0	85.4	15.7	66.7	93.0	90.8	26.0	44.6
	D3	39.3	4.2	16.6	100	98.5	96.2	40.5	52.7
	D4	94.6	99.8	77.2	100	33.2	99.3	24.8	74.8
	D5	52.8	91.1	33.5	99.6	97.4	99.6	50.0	76.6
	Mean	53.1	74.5	48.5	92.6	65.5	82.1	28.4	62.1
Body Only (30)	D1	76.2	0.0	2.0	0.0	90.2	90.0	96.0	47.0
	D2	12.0	0.0	26.6	0.0	40.3	1.2	32.9	16.3
	D3	1.2	100	0.0	7.4	96.9	94.0	53.5	48.2
	D4	55.4	45.0	0.0	97.0	97.9	95.1	62.1	58.9
	D5	44.0	2.1	0.0	97.3	35.3	75.0	99.3	49.2
	Mean	37.8	29.4	5.72	40.3	72.1	71.1	68.8	43.9

First, the 2D-only case in Table 5 represents a feature set only consisting of the head rotation and joint coordinates (X and Y coordinates), and depth information is not used. As shown in Table 5, the model trained with 2D information achieves similar accuracy results compared with the model trained with the entire feature set (Table 3). The results indicate that depth information has very limited impact on the model classification task.

The second block in Table 5 illustrates driving task classification using only head pose information. Specifically, the three head rotation angles: yaw, pitch, and roll are used to construct the feature set. The classification accuracy using head pose is much less than the accuracy in previous cases. For the left mirror-checking and texting tasks, which have significantly different characteristics than other tasks, the detection is accurate. However, for the other tasks, using only head pose information is not sufficient for accurate detection. For example, the driver rear mirror-checking behaviour (T3) is similar to the task of using a video device. Moreover, without considering body information, the phone answering behaviour cannot be detected accurately since the driver is usually looking forward to the road and the head pose is very similar to normal driving. The confusion matrix for driver 5, which has the most accurate results among the five drivers, is shown in Fig. 12. In terms of driver 5, FFNN is not able to accurately distinguish tasks 1, 3, and 7, which are normal driving, rear mirror checking, and answering the phone. Approximately one-third of normal driving cases (224 samples) are classified into the phone answering task. For rear mirror checking, more samples are falsely detected as the video device using task. It is obvious from the confusion matrix that, without using body features and using only head pose features, it is difficult to identify the actual driver behaviour.

Confusion Matrix

Output Class	1	2	3	4	5	6	7	
1	420 (9.8%)	50 (1.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	7 (0.2%)	88.1% / 11.9%
2	129 (3.0%)	755 (17.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.0%)	195 (4.6%)	69.8% / 30.2%
3	0 (0.0%)	0 (0.0%)	162 (3.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	100% / 0.0%
4	0 (0.0%)	0 (0.0%)	0 (0.0%)	797 (18.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	100% / 0.0%
5	11 (0.3%)	0 (0.0%)	320 (7.5%)	3 (0.1%)	420 (9.8%)	0 (0.0%)	1 (0.0%)	55.6% / 44.4%
6	12 (0.3%)	2 (0.0%)	0 (0.0%)	0 (0.0%)	11 (0.3%)	509 (11.9%)	11 (0.3%)	93.4% / 6.6%
7	224 (5.2%)	22 (0.5%)	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	214 (5.0%)	46.4% / 53.6%
	52.8% / 47.2%	91.1% / 8.9%	33.5% / 66.5%	99.6% / 0.4%	97.4% / 2.6%	99.6% / 0.4%	50.0% / 50.0%	76.6% / 23.4%

Target Class

Fig. 12. Confusion matrix of the classification of the seven tasks for driver 5 using 3D head pose features only.

The third block indicates the behaviour detection using only body features. There are 30 total features used, containing the X, Y, and Z coordinates of the hand, wrist, elbow, and shoulder joints. As shown in Table 5, ignoring the 3D head pose features and the eyes and nose location information, the detector fails to identify the mirror-checking behaviours. By using only body features, the distraction behaviour can be detected with a certain degree of accuracy, while the detection accuracy for the four mirror checking behaviours is quite low. The worst case is the rear-mirror checking behaviour, which only achieved 5.72% accuracy in general.

Based on the above evidence, to obtain a better understanding of the tasks that the driver is undergoing, both the head and body features are necessary. From the feature comparison, head pose features are more useful than body features since the 3D head pose information leads to better detection results (62.1% average) compared with the 3D body features (43.9%). Fig. 13 shows the model classification results for driver 1 when the model is only trained with body features. The three distraction behaviours are accurately detected using body features, while the four mirror checking detections are difficult to identify.

Fig. 13. Confusion matrix of the classification of the seven tasks for driver 1 using only body features.

TABLE 6
CLASSIFICATION RESULT USING FFNN WITH 18 SELECTED FEATURES

| | Driving Tasks | | | | Non-Driving Tasks | | | |
	T1	T2	T3	T4	T5	T6	T7	Ave
D1	90.6	91.8	98.2	98.1	100	88.3	100	95.1
D2	1.4	20.9	25.5	91.1	97.3	60.2	93.5	49.3
D3	97.8	98.7	82.5	99.8	98.1	100	28.5	85.0
D4	65.2	78.1	96.7	99.9	98.3	97.9	80.4	86.2
D5	51.4	100	89.9	99.9	87.9	94.9	100	88.0
Mean	61.3	80.0	78.6	97.8	96.3	88.2	80.5	80.7

Finally, the important features for driver task classification are selected according to the integrated feature extraction technique in Section ☐ and these features are input to the FFNN model. In total, 18 features are selected as important features. The feature set contains the following features: {yaw, pitch, roll, nose (X,Y,Z), left hand (X,Y,Z), right hand (X,Y,Z), left shoulder (X,Y,Z), and right shoulder (X,Y,Z)}. The classification results are shown in Table 6. The overall accuracy of task detection is 80.7%, which is slightly less than the model trained with the entire feature set. However, the selected 18 features still yield an acceptable accurate detection, also, the time cost of the training and testing process is less than when using the entire feature set. Therefore, the driver tasks can be detected using the small feature set.

V. DISCUSSION AND FUTURE WORK

Based on the results shown in the previous section, driver task recognition can be achieved with a feedforward neural network. The FFNN could reasonably detect seven tasks for different drivers and achieved high-precision detection for secondary tasks. The FFNN has advantages for driver task detection over other machine learning methods. Classification for different tasks resulted in different detection accuracy. The results indicate that for tasks such as texting, left and right mirror checking, which have obvious distinct features, the detection results are accurate. However, for tasks that have similar postures, the model can be confused. In this study, normal driving behaviour has similar characteristics to rear mirror checking behaviour and phone-answering tasks; therefore, the detection results for these behaviours are slightly worse than for other behaviours. In addition to the similar characteristics of these behaviours, another reason for less accurate detection results is driving style. Although accurate detection results can be achieved for some drivers, the FFNN cannot obtain a universal accuracy for all drivers. For example, task detection for driver 2 is less accurate than for other drivers due to driving style and sensor noise. A driver has a unique driving and mirror checking style. Some drivers prefer to use significant head and body movement during mirror checking while others may try to use less body movement and use eye movement to capture information. Therefore, the following aspects are discussed and can be improved to achieve higher task detection accuracy.

First, the driver head and body signals captured with a Kinect are very noisy. Sometimes the detection is less precise and the detected joint positions are shifted and unreasonable. This phenomenon is particularly worse for the seated driver inside the vehicle. In this study, a simple integrated tracking and smoothing technique is used, which consists of a jitter removal filter and an exponential filter. Although the integrated filter can recover unreasonable detection and smooth the signals, important information can be lost and the filter can be further improved by using more advanced filters such as the Kalman filter or particle filter for joint position tracking. Therefore, the quality of Kinect signals, as well as the model detection results, can be further improved. Moreover, in this study, only colour and depth images are collected; however, Kinect also supports audio recording. Therefore, in the future, audio information in the cabin can be captured as another important data source to assist in the detection of non-driving-related tasks.

Second, in this study, the feature selection and extraction methods are constructed based on random forests and the maximal information coefficient technique. This integrated method estimates the importance of the driver body features and the FFNN using these features achieved accurate detection results for some drivers. However, detection accuracy decreased significantly for the second driver for a few reasons. To obtain universal accurate task detection results, more drivers must to be studied in the future. Increasing the dataset volume and data diversity is an efficient way to solve the aforementioned problem. Meanwhile, more driver features can be used. In the study, only the position and depth information for the eyes are used. The driver gaze movement and gaze tracking technique have been successfully adopted in some research on driver fatigue, inattention, and distraction monitoring. Gaze information can be very useful when the drivers prefer not to move their body when performing mirror-checking tasks.

Finally, on-road data collection can be performed in the future for the study of real-time driver behaviour detection within normal driving environments. Currently, for safety considerations, the drivers were asked to perform the experimental tasks without driving the vehicle because secondary tasks such as texting and playing a video device are extremely dangerous when driving and should be avoided. Therefore, the most naturalistic data is difficult to collect. However, in the future, with the help of ADAS and the mid-level automated vehicle technique, drivers are allowed to remove their hands from the steering wheel. Therefore, more distraction behaviours can be collected and the study for real-time driver distraction detection in a real vehicle can be performed. The real-time driver monitoring study will significantly improve the driving safety for both conventional vehicles and highly automated vehicles.

VI. CONCLUSIONS

In this study, driving behaviours for different drivers are studied. The driving behaviours are classified into two categories, normal driving tasks and distracting tasks. A feedforward neural network is trained to distinguish the four mirror-checking behaviours from the three secondary driving tasks. Both depth information and the 2D location of the body joints are collected using Kinect. The noisy data is processed with an integrated filtering system. Then, the importance of each driver feature to behaviour recognition is evaluated using random forests and maximal information efficiency. The feature importance prediction with these two feature evaluation techniques shows consistent results. The most important driver features for driver behaviour among all the drivers are determined. The FFNN has been proven to have advantages for behaviour detection tasks over other popular machine learning methods. The model achieved an average of greater than 80% accuracy for the five drivers. With the evaluation of feature importance and their influence to the classification task, the head pose feature, hand position, and shoulder positions for the driver are selected as the most important features. In addition, based on the evaluation of the depth, head, and body features, it is found that the depth information for the body joints and facial markers have very limited influence on behaviour

recognition. Meanwhile, the head and body features should be combined with a comprehensive driver behaviour understanding since only using the head or body features will lead to large false detection rates.

The conclusion is made that for future driving monitoring and behaviour understanding, the head and body signals are equally important and necessary. Future works will focus on the collection of more real world dataset and recognize more sophisticate driver behaviours. These study will benefit future ADAS design and improve driving safety by real-time driver status monitoring.

REFERENCES

[1] Wang, Fei-Yue. "Computational Social Systems in a New Period: A Fast Transition Into the Third Axial Age." IEEE Transactions on Computational Social Systems, 4(3): 52-53, 2017.
[2] Arbabzadeh, Nasim, and Mohsen Jafari. "A Data-Driven Approach for Driving Safety Risk Prediction Using Driver Behavior and Roadway Information Data." IEEE Transactions on Intelligent Transportation Systems (2017).
[3] Xing, Yang, et al. "Driver Workload Estimation using a Novel Hybrid Method of Error Reduction Ratio Causality and Support Vector Machine." Measurement, 2017, in press.
[4] Michon, John A. "A critical view of driver behavior models: What do we know, what should we do." Human behavior and traffic safety (1985): 485-520.
[5] Wang, F-Y., Zheng, N., Cao, D., et al. "Parallel driving in CPSS: a unified approach for transport automation and vehicle intelligence." IEEE/CAA Journal of Automatica Sinica, 2017.
[6] Castignani, German, et al. "Driver behavior profiling using smartphones: A low-cost platform for driver monitoring." IEEE Intelligent Transportation Systems Magazine 7.1 (2015): 91-102.
[7] Lv, C., Liu, Y., et al. Simultaneous Observation of Hybrid States for Cyber-Physical Systems: A Case Study of Electric Vehicle Powertrain, IEEE Transactions on Cybernetics, 2017.
[8] Foroutan, N.,Hamzeh, A. "Discovering the Hidden Structure of a Social Network: A Semi Supervised Approach." IEEE Transactions on Computational Social Systems, 3(4): 151-163, 2016.
[9] Tong, G., et al. "Effector Detection in Social Networks." IEEE Transactions on Computational Social Systems, 4(1): 14-25, 2017.
[10] Lv, C., Wang, H., and Cao, D. "High-Precision Hydraulic Pressure Control Based on Linear Pressure-Drop Modulation in Valve Critical Equilibrium State." IEEE Transactions on Industrial Electronics, 2017.
[11] Liang, Yulan, and John D. Lee. "A hybrid Bayesian Network approach to detect driver cognitive distraction." Transportation research part C: emerging technologies 38 (2014): 146-155.
[12] Lv, C., Cap, D., et al. Analysis of Autopilot Disengagements Occurring During Autonomous Vehicle Testing, IEEE/CAA Journal of Automatica Sinica, 2017.
[13] Sivak M, Schoettle B. Motion sickness in self-driving vehicles. UMTRI-2015-12. (2015). Ann Arbor: University of Michigan Transportation Research Institute.
[14] Jain, Ashesh, et al. "Car that knows before you do: Anticipating maneuvers via learning temporal driving models." Proceedings of the IEEE International Conference on Computer Vision. 2015.
[15] Murphy-Chutorian, Erik, and Mohan Manubhai Trivedi. "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness." IEEE Transactions on intelligent transportation systems 11.2 (2010): 300-311.
[16] Murphy-Chutorian, Erik, and Mohan Manubhai Trivedi. "Head pose estimation in computer vision: A survey." IEEE transactions on pattern analysis and machine intelligence 31.4 (2009): 607-626.
[17] Vicente, Francisco, et al. "Driver gaze tracking and eyes off the road detection system." IEEE Transactions on Intelligent Transportation Systems 16.4 (2015): 2014-2027.
[18] Fletcher, Luke, and Alexander Zelinsky. "Driver inattention detection based on eye gaze—Road event correlation." The international journal of robotics research 28.6 (2009): 774-801.
[19] Das, Nikhil, Eshed Ohn-Bar, and Mohan M. Trivedi. "On performance evaluation of driver hand detection algorithms: Challenges, dataset, and

metrics." Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on. IEEE, 2015.

[20] Cheng, Shinko Y., Sangho Park, and Mohan M. Trivedi. "Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis." Computer Vision and Image Understanding 106.2 (2007): 245-257.

[21] Du, Haiping, Weihua Li, and Nong Zhang. "Vibration control of vehicle seat integrating with chassis suspension and driver body model." Advances in Structural Engineering 16.1 (2013): 1-9.

[22] Tran, Cuong, Anup Doshi, and Mohan Manubhai Trivedi. "Modeling and prediction of driver behavior by foot gesture analysis." Computer Vision and Image Understanding 116.3 (2012): 435-445.

[23] Ohn-Bar, Eshed, et al. "Head, eye, and hand patterns for driver activity recognition." Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014.

[24] Rezaei, Mahdi, and Reinhard Klette. "Look at the driver, look at the road: No distraction! No accident!" Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.

[25] Mbouna, Ralph Oyini, Seong G. Kong, and Myung-Geun Chun. "Visual analysis of eye state and head pose for driver alertness monitoring." IEEE transactions on intelligent transportation systems 14.3 (2013): 1462-1469.

[26] Li, Nanxiang, and Carlos Busso. "Detecting drivers' mirror-checking actions and its application to maneuver and secondary task recognition." IEEE Transactions on Intelligent Transportation Systems 17.4 (2016): 980-992.

[27] Kondyli, Alexandra, et al. "Computer Assisted Analysis of Drivers' Body Activity Using a Range Camera." IEEE Intelligent Transportation Systems Magazine 7.3 (2015): 18-28.

[28] Jha, Sumit, and Carlos Busso. "Analyzing the relationship between head pose and gaze to model driver visual attention." Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on. IEEE, 2016.

[29] Morton, Jeremy, Tim A. Wheeler, and Mykel J. Kochenderfer. "Analysis of recurrent neural networks for probabilistic modeling of driver behavior." IEEE Transactions on Intelligent Transportation Systems 18.5 (2017): 1289-1298.

[30] Wollmer, Martin, et al. "Online driver distraction detection using long short-term memory." IEEE Transactions on Intelligent Transportation Systems 12.2 (2011): 574-582.

[31] Tango, Fabio, and Marco Botta. "Real-time detection system of driver distraction using machine learning." IEEE Transactions on Intelligent Transportation Systems 14.2 (2013): 894-905.

[32] Lee, Yi-Ching, et al. "Learning to Predict Driver Behavior from Observation." (2017). The AAAI 2017 Spring Symposium on Learning from Observation of Humans Technical Report SS-17-06.

[33] Harbluk, Joanne L., Y. Ian Noy, and Moshe Eizenman. The impact of cognitive distraction on driver visual behaviour and vehicle control. No. TP 13889 E. 2002.

[34] Mühlbacher-Karrer, Stephan, et al. "A Driver State Detection System—Combining a Capacitive Hand Detection Sensor With Physiological Sensors." IEEE Transactions on Instrumentation and Measurement 66.4 (2017): 624-636.

[35] Ranft, Benjamin, and Christoph Stiller. "The role of machine vision for intelligent vehicles." IEEE Transactions on Intelligent Vehicles 1.1 (2016): 8-1.

[36] Gaglio, Salvatore, Giuseppe Lo Re, and Marco Morana. "Human activity recognition process using 3-D posture data." IEEE Transactions on Human-Machine Systems 45.5 (2015): 586-597.

[37] Neto, Laurindo Britto, et al. "A Kinect-based wearable face recognition system to aid visually impaired users." IEEE Transactions on Human-Machine Systems 47.1 (2017): 52-64.

[38] Mehran Azimi. Skelton Joint Smoothing White Paper [Online]. Available: https://msdn.microsoft.com/en-us/library/jj131429.aspx.

[39] Darby, John, et al. "An evaluation of 3D head pose estimation using the Microsoft Kinect v2." Gait & posture 48 (2016): 83-88.

[40] Breiman, Leo. "Random forests." Machine learning 45.1 (2001): 5-32.

[41] Fernandez-Delgado, Manuel, et al. Do we need hundreds of classifiers to solve real world classification problems." J. Mach. Learn. Res 15.1. (2014): 3133-3181

[42] Reshef, David N., et al. "Detecting novel associations in large data sets." science 334.6062 (2011): 1518-1524.

[43] Cover, T.M.; Thomas, J.A. (1991). Elements of Information Theory (Wiley ed.). ISBN 978-0-471-24195-9.

[44] Duch, Włodzisław, and Norbert Jankowski. "Survey of neural transfer functions." Neural Computing Surveys 2.1 (1999): 163-212.

[45] Ohn-Bar, Eshed, et al. "On surveillance for safety critical events: In-vehicle video networks for predictive driver assistance systems." Computer Vision and Image Understanding 134 (2015): 130-140.

Yang Xing received his B.S. in Automatic Control from Qingdao University of Science and Technology, Shandong, China, in 2012. He then received his Msc. with distinction in Control Systems from the department of Automatic Control and System Engineering, The University of Sheffield, UK, in 2014. Now he is a Ph. D. candidate for Transport Systems, Cranfield University, UK. His research interests include driver behaviour modelling, driver-vehicle interaction, and advance driver assistance systems. His work focuses on the understanding of driver behaviours and identification of driver intentions using machine-learning methods for intelligent and automated vehicles.
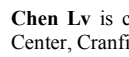
Chen Lv is currently a Research Fellow at Advanced Vehicle Engineering Center, Cranfield University, UK. He received the Ph.D. degree at Department of Automotive Engineering, Tsinghua University, China in 2016. From 2014 to 2015, he was a joint PhD researcher at EECS Dept., University of California, Berkeley. His research focuses on cyber-physical system, hybrid system, advanced vehicle control and intelligence, where he has contributed over 40 papers and obtained 11 granted China patents. Dr. Lv serves as a Guest Editor for IEEE/ASME Transactions on Mechatronics, IEEE Transactions on Industrial Informatics and International Journal of Powertrains, and an Associate Editor for International Journal of Electric and Hybrid Vehicles, International Journal of Vehicle Systems Modelling and Testing, International Journal of Science and Engineering for Smart Vehicles, and Journal of Advances in Vehicle Engineering. He received the Highly Commended Paper Award of IMechE UK in 2012, the National Fellowship for Doctoral Student in 2013, the NSK Outstanding Mechanical Engineering Paper Award in 2014, the Tsinghua University Graduate Student Academic Rising Star Nomination Award in 2015, the China SAE Outstanding Paper Award in 2015, the 1st Class Award of China Automotive Industry Scientific and Technological Invention in 2015, and the Tsinghua University Outstanding Doctoral Thesis Award in 2016.

Zhaozhong Zhang received his B.S. in Mechatronics from The University of Nottingham Ningbo China, Ningbo, China, in 2013. He then received in Msc. in Automotive Engineering from Cranfield University UK in 2015, where he is pursuing his Ph.D. degree currently. His research interests include driver attention and distraction detection. He is working on the topic of driver behavior and automated driving.

Huaji Wang received the B.S. degree in automotive engineering from Jilin University, China, in 2005, and the Ph.D. degree in engineering from the University of Cambridge, Cambridge, U.K., in 2016, concentrating on the study of driver/vehicle systems and driver-automation collaboration. He presently works as a Research Fellow in Automated Driving at Cranfield University, UK.
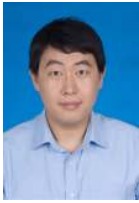
Xiaoxiang Na received the B.Sc. and M.Sc. degrees in automotive engineering from the College of Automotive Engineering, Jilin University, Changchun, China, in 2007 and 2009, respectively, and the Ph.D. degree in driver–vehicle dynamics from the Department of Engineering, University of Cambridge (CUED), Cambridge, U.K., in 2014. He is currently a Research Associate with CUED. His main research interests include driver–vehicle dynamics and vehicle in-service monitoring.

Dongpu Cao received the Ph.D. degree from Concordia University, Canada, in 2008. He is currently a Senior Lecturer at Advanced Vehicle Engineering Center, Cranfield University, UK. His research focuses on vehicle dynamics and control, automated driving and parallel driving, where he has contributed more than 100 publications and 1 US patent. He received the ASME AVTT'2010 Best Paper Award and 2012 SAE Arch T. Colwell Merit Award. Dr. Cao serves as an Associate Editor for IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE/ASME TRANSACTIONS ON MECHATRONICS and ASME JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT, AND CONTROL. He has been a Guest Editor for VEHICLE SYSTEM DYNAMICS, and IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS. He serves on the SAE International Vehicle Dynamics Standards Committee and a few ASME, SAE, IEEE technical committees.

Efstathios Velenis received the Ph.D. degree from the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2006. He is currently a Senior Lecturer with the Advanced Vehicle Engineering Centre, Cranfield University, Cranfield, U.K. His current research interests include vehicle dynamics and control, optimal control for active chassis systems, traction, braking, and handling control for electric/hybrid vehicles, lap-time optimization, and tire modeling. He obtained research funding from EPSRC, Innovate U.K., and the European Commission. Dr. Velenis received the Luther Long Award for best Ph.D. dissertation in the area of engineering mechanics at Georgia Tech.

Fei-Yue Wang (S'87-M'89-SM'94-F'03) received his Ph.D. in Computer and Systems Engineering from Rensselaer Polytechnic Institute, Troy, New York in 1990. He joined the University of Arizona in 1990 and became a Professor and Director of the Robotics and Automation Lab (RAL) and Program in Advanced Research for Complex Systems (PARCS). In 1999, he founded the Intelligent Control and Systems Engineering Center at the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Overseas Chinese Talents Program from the State Planning Council and "100Talent Program" from CAS, and in 2002, was appointed as the Director of the Key Lab of Complex Systems and Intelligence Science, CAS. From 2006 to 2010, he was Vice President for Research, Education, and Academic Exchanges at the Institute of Automation, CAS. In 2011, he became the State Specially Appointed Expert and the Director of the State Key Laboratory of Management and Control for Complex Systems.

Dr. Wang's current research focuses on methods and applications for parallel systems, social computing, and knowledge automation. He was the Founding Editor-in-Chief of the International Journal of Intelligent Control and Systems (1995-2000), Founding EiC of IEEE ITS Magazine (2006-2007), EiC of IEEE Intelligent Systems (2009-2012), and EiC of IEEE Transactions on ITS (2009-2016). Currently he is EiC of IEEE Transactions on Computational Social Systems, Founding EiC of IEEE/CAA Journal of Automatica Sinica, and Chinese Journal of Command and Control. Since 1997, he has served as General or Program Chair of more than 20 IEEE, INFORMS, ACM, and ASME conferences. He was the President of IEEE ITS Society (2005-2007), Chinese Association for Science and Technology (CAST, USA) in 2005, the American Zhu Kezhen Education Foundation (2007-2008), and the Vice President of the ACM China Council (2010-2011). Since 2008, he has been the Vice President and Secretary General of Chinese Association of Automation. Dr. Wang has been elected as Fellow of IEEE, INCOSE, IFAC, ASME, and AAAS. In 2007, he received the National Prize in Natural Sciences of China and was awarded the Outstanding Scientist by ACM for his research contributions in intelligent control and social computing. He received IEEE ITS Outstanding Application and Research Awards in 2009, 2011 and 2015, and IEEE SMC Norbert Wiener Award in 2014.