IET Intelligent Transport Systems

Special Issue: Recent Advancements on Electrified, Low Emission and Intelligent Vehicle-Systems

Statistical-based approach for driving style recognition using Bayesian probability with kernel density estimation

ISSN 1751-956X
Received on 20th November 2017
Revised 16th January 2018
Accepted on 7th February 2018
E-First on 23rd March 2018
doi: 10.1049/iet-its.2017.0379
www.ietdl.org

Wei Han[1,2], *Wenshuo Wang*[3,4] ✉, *Xiaohan Li*[5], *Junqiang Xi*[3]

[1]*Department of Computer Science, Tsinghua University, Beijing, People's Republic of China*
[2]*Institute of Microelectronics of Chinese Academy of Sciences, Beijing, People's Republic of China*
[3]*Department of Mechanical Engineering, Beijing Institute of Technology, Beijing, People's Republic of China*
[4]*Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA*
[5]*Chair of Human-Machine Systems, Faculty of Mechanical Engineering and Transport Systems, Technische Universität Berlin, Berlin, Germany*
✉ *E-mail: wwsbit@gmail.com*

Abstract: Driving style recognition plays a crucial role in eco-driving, road safety, and intelligent vehicle control. This study proposes a statistical-based recognition method to deal with driver behaviour uncertainty in driving style recognition. First, the authors extract discriminative features using the conditional kernel density function to characterise path-following behaviour. Meanwhile, the posterior probability of each selected feature is computed based on the full Bayesian theory. Second, they develop an efficient Euclidean distance-based method to recognise the path-following style for new input datasets at a low computational cost. By comparing the Euclidean distance of each pair of elements in the feature vector, then they classify driving styles into seven levels from normal to aggressive. Finally, they employ a cross-validation method to evaluate the utility of their proposed approach by comparing with a fuzzy logic (FL) method. The experiment results show that the proposed statistical-based recognition method integrating with the kernel density is more efficient and robust than the FL method.

1 Introduction

Driving style is very important for intelligent vehicle control, traffic systems, road safety and eco-driving [1–10]. For example, a moderate driver usually drives in a fuel-saving way, while an aggressive driver will drive in a fuel-consuming way. Driving style recognition can offer feedback information to vehicle control systems and enable the control systems to meet individual drivers' needs in time. However, recognising driver behaviour or driving style is a challenging task since feature parameters greatly vary over different driving behaviours and driving environments. Lots of approaches in existing research have been developed to recognise driving style [11, 12], which can be roughly categorised into two groups: model-based and learning-based.

One indirect way is to develop such driver model capable of characterising drivers' basic behaviours and utilise model parameters to represent driving styles. The hidden Markov model (HMM) has been widely utilised to model and predict the driver state and driving behaviour because of its powerful ability to describe a latent state in dynamic and stochastic processes. For example, researchers in [13, 14] applied a hidden Markov model (HMM) to identify the underlying relationship between observations and driver state. A driver–vehicle system was developed as a hybrid-state model and the HMM was then used to estimate the driver's decision when driving near intersections [14, 15]. Some authors [16, 17] also utilised an autoregressive exogenous (ARX) model and an extended probabilistic ARX (P-ARX) model to classify drivers. Shi *et al.* [18] made a comparison analysis for eco-driving based on a normalised driver model. To mimic and model the driver behaviour uncertainty, different kinds of stochastic models were also developed and adopted [19, 20]. However, the non-linearity and uncertainty of driving behaviour make it intractable to precisely identify these models' parameters.

The other is to directly analyse the driving data using pattern-recognition or data-analysis methods without establishing the specific driver models. For instance, Zhang *et al.* [21] applied three recognition methods to recognise driving skills, including the multilayer perception artificial neural networks, decision tree, and support vector machines (SVM). The coefficients of discrete Fourier transform of steering wheel angle were treated as the discriminant features. The authors in [22] investigated the relationships between driver state and driver's actions using a clustering method with eight state-action variables. According to different driving patterns, the state-action clusters segmented drivers into different styles. In addition, the learning-based methodologies such as Bayesian non-parametric techniques have also been directly applied to analyse driving style [23, 24]. For example, fuzzy logic (FL) methods [25] were selected to classify driving style for improving the cure speed model accuracy. Though the above mentioned works have made a great progress in driver behaviour modelling and driving style recognition, they did not consider the driver behaviour uncertainty which is usually caused by psychological/physical factors and driving environments. According to the above discussions, we found that there are two key issues existing in driving style recognition:

- *Feature selection*. It is difficult to select a pair of feature parameters that can fully represent or define all aggressive (or normal) drivers, though the rule-based strategies are able to classify most drivers into different categories. Two of the main reasons are (i) that an aggressive driver will not always drive vehicles in an aggressive way, which causes the overlapped data collected from drivers with different driving styles and (ii) that the threshold values of driving style are greatly different among individuals.
- *Driver behaviour uncertainty*. Driver behaviour could be affected by the disturbances of driving environments and physical/psychological factors, which will diversify driving style at the different time and different driving environments.

Therefore, it is difficult to recognise driving style (e.g. aggressive or normal) for individuals with uncertain factors considered [26]. Toward this end, we present a statistical-based recognition method to consider driver behaviour uncertainty, allowing us to classify drivers into two groups, i.e. aggressive and normal (typical). First, a conditional distribution function – kernel density function – is introduced to describe the driver behaviour uncertainty, which has shown its effectiveness of describing driving behaviour uncertainty [27] due to its ability to measure the variants of variables [28]. After that, according to the learned kernel density function and the full Bayesian theory, a posterior probability of each feature is computed with respect to each driving style. We also develop an efficient approach based on Euclidean distance to determine the driving style at a low computational cost. Finally, a series of cross-validation (CV) experiments are conducted by comparing with a FL method to show the effectiveness of our proposed statistical-based recognition method. In summary, the paper consists of the following three contributions:

i. Introducing a statistical-based approach to recognise driving styles considering driver behaviour uncertainty.
ii. Developing the Euclidean distance-based decision method to determine the driving style of specific driver behaviours.
iii. Verifying the effectiveness of our proposed method with comparison experiment.

Following the overview in Section 1 of this paper, Section 2 shows the procedure of feature selection. Section 3 presents the FL recognition algorithm and the proposed method. Section 4 describes data collection and experiment design in a driving simulator. Section 5 shows the experiment results and analysis. Finally, Section 6 gives a further discussion and conclusion.

2 Feature selection

The goal of feature selection is to allow pattern vectors belonging to different categories to occupy compact and disjoint regions as much as possible in a specified feature space. In general, the data using for driving style recognition can be grouped into three categories:

- Driver-dependent, including the physical signal (e.g. the steering angle, throttle opening, gesture, eyes related signal [29–31]) and physiological signal (e.g. the rate of heart beat, EEG, EMG [32]).
- Vehicle-dependent, including vehicle speed, acceleration, yaw angle [29, 30, 33] and so on.
- Driving environment-dependent, including road profile, surrounding vehicles, traffic flow and so on.

In this work, we mainly focus on the driver's longitudinal behaviour when tracking a given curvy road and the vehicle-dependent signal is preferred to characterise driving styles. However, the required feature is greatly different for different driving tasks, as shown in Table 1. Therefore, in order to select a feature that can describe the driver behaviour uncertainty when following a given curvy path, we make a distribution analysis for all feature parameters. Fig. 1 shows the time-series driving data (i.e. speed, throttle opening, and acceleration) collected from two

Table 1 Feature selection and the recognition methods with respect to different driving tasks

Driving task	Feature parameters	Method
car-following [22, 33]	• relative distance	• Gaussian mixture model
	• vehicle speed	• fuzzy clustering
	• vehicle position	
curve path-following	• vehicle speed	• model predictive control
driving styles	• acceleration	• P-ARX
[16, 29, 31, 34]	• yaw rate	• neural network
	• lateral displacement	• FL
	• vehicle speed	• HMM
	• steering angle	• SVM
	• physical signal	• OOBNs or BNs
	• physiological signal	• Bayesian filter

OOBNs – objected-oriented Bayesian networks

drivers with different driving styles and their distribution. With the aim of selecting the discriminative feature parameter, we make two assumptions as follows:

- *Statistical characteristic invariance*: In specified driving environments, the vehicle speed or throttle opening from a driver may change at a different time, but its statistical feature such as the distribution property is relatively invariance. Table 2 presents the statistical results from two drivers with distinct driving styles (aggressive and normal), which indicates that the mean value and standard deviations of statistical metrics vary greatly between drivers with different driving styles.
- *Maximum discrimination*: The selected feature parameters should maximise the discrimination of driving styles. From the statistical results in Fig. 1, it is obvious that the vehicle speed and throttle opening are the most discriminative features and therefore selected as the feature parameters.

2.1 Vehicle speed

When tracking a given curvy road, the vehicle speed is one of the parameters that can directly show and characterise driving preferences [25, 35] such as aggressive or normal. In Fig. 1 and Table 2, for example, it is obvious that the aggressive prefers to the vehicle speed of $v_x \in \{[20, 40] \cup [60, 100]\}$ km/h, while the normal driver prefers to the vehicle speed falling in $[40, 60]$ km/h. The authors of [9, 25, 34] also demonstrated that aggressive drivers usually prefer to drive with a high speed when following a curvy road.

2.2 Throttle opening

As one of the parameters directly controlled by a human driver, throttle opening can reflect the driver's preference. From Fig. 1, we know that the distribution of throttle opening (Table 2) is more suitable to recognise driving styles than acceleration. Wang *et al.* in [34] also demonstrated that the combination of vehicle speed and throttle opening can directly reflect the longitudinal acceleration. For instance, when driving on a flat road, a high vehicle speed with a small throttle opening will lead to a large deceleration.

Therefore, based on the aforementioned discussion, the vehicle speed (v) and the throttle opening (α) are selected as the feature vector $\boldsymbol{x} = (v, \alpha)$ to distinguish driving styles when following a curvy road, instead of involving the acceleration, which can reduce the information redundancy and the computational cost. Based on the selected feature parameter $\boldsymbol{x}$, our goal is to find a model $f : \boldsymbol{x} \to s$, capable of recognising driving styles $s \in \mathcal{S}$, where $\mathcal{S} = \{s | s = -3, -2, -1, 0, 1, 2, 3\}$ is the label set of driving styles. Here, the element of set $\{-3, -2, -1, 0, 1, 2, 3\}$ represents the aggressive or normal level. A larger value of s indicates a more aggressive driving style, and vice versa. For example, $s = -3$ and $s = 3$ represent the normal type and the aggressive type, respectively.

3 Methodology

3.1 FL method

To some extent, driving style is a *vague* concept that cannot precisely be divided into a specified category such as aggressive or normal because the aggressive or normal scales of driving style are difficult to be quantified. Our perception of the real driving styles is pervaded by concepts which do not have sharply defined boundaries. Therefore, a fuzzy-based mathematical tool – FL – has been widely introduced to recognise driver behaviour [36], driving profile [37], and driving styles [25], which provides a reasonable way to deal with imprecision and information granularity.

Therefore, in order to compare with our developed method, a fuzzy inference system (FIS) based on Mamdani rule is defined with two inputs (i.e. vehicle speed and throttle opening) and one output (i.e. level of driving styles). The membership function is defined according to prior knowledge and Chu *et al.* [25]. Corresponding fuzzy values of the first input – vehicle speed (v) –

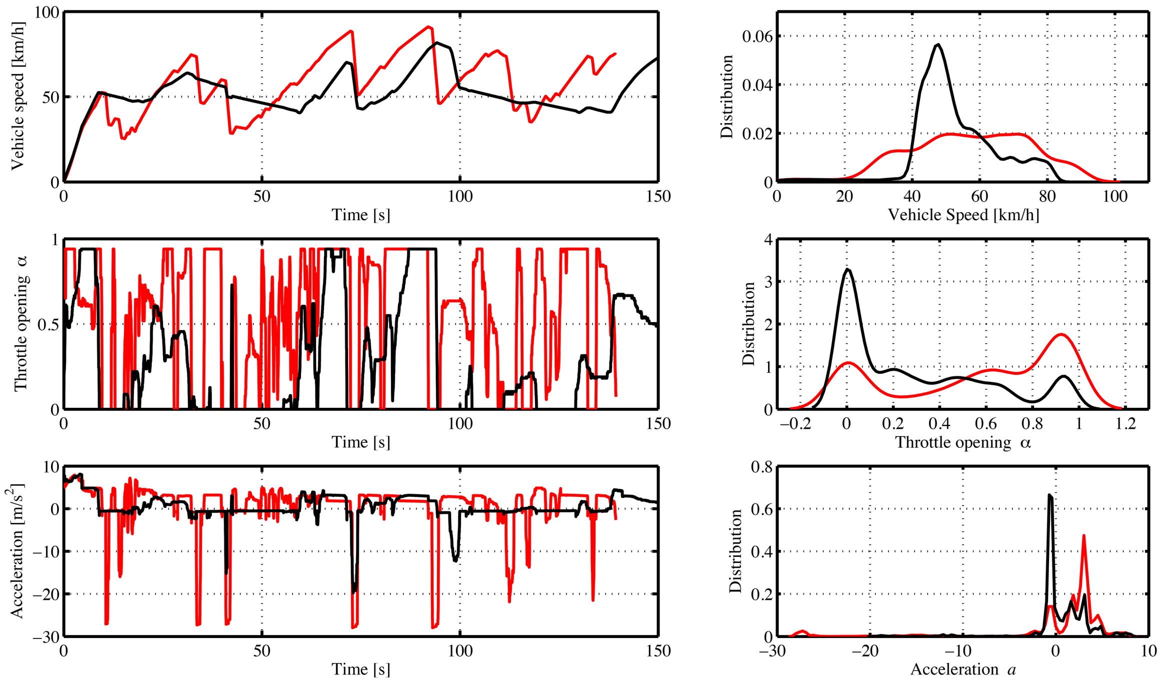

Fig. 1 *Driving data (left) with two distinguished driving styles and their distributions (right). Red line: the aggressive driver; black line: the normal driver*

Table 2 Means (standard deviations) of vehicle speed and throttle opening for two drivers with different driving styles

Aggressive drivers		Normal drivers	
Speed, km/h	Throttle	Speed, km/h	Throttle
56.85 (317.55)	0.57 (0.13)	52.49 (152.32)	0.285 (0.10)
61.33 (256.96)	0.61 (0.13)	48.53 (173.30)	0.259 (0.06)
61.93 (250.44)	0.65 (0.14)	52.93 (129.78)	0.235 (0.07)
61.34 (273.58)	0.63 (0.13)	50.57 (137.98)	0.238 (0.06)
64.45 (307.29)	0.60 (0.15)	50.67 (106.97)	0.195 (0.06)
64.24 (301.92)	0.69 (0.12)	50.53 (117.19)	0.213 (0.06)
63.15 (296.33)	0.61 (0.14)	49.49 (154.64)	0.284 (0.06)
61.93 (263.43)	0.65 (0.14)	46.34 (115.19)	0.156 (0.03)
64.15 (287.71)	0.61 (0.13)	48.25 (95.390)	0.153 (0.04)

are defined to be *lower* (L), *middle* (M), and *high* (H). The fuzzy values of second input – throttle opening ($\alpha \in [0, 1]$) – are defined to be *lower* (L), *middle* (M), and *high* (H). The fuzzy values of output – the level of driving styles – are defined to be *lower normal* (LN), *normal* (N), *middle* (M), *aggressive* (A), and *high aggressive* (HA). Here we encode the output sets LN and HA as -3 and 3, respectively. All membership functions are shown in Fig. 2 and Table 3.

3.2 Proposed method

3.2.1 Kernel density estimation: Kernel density estimation, as an unsupervised learning method, can estimate a probability density at a point x_0 given a random sample $x_1, x_2, \ldots, x_N$ from a probability density $f(x)$. For two classes of one-dimensional (1D) data sequences $X_1 = \{x_1^1, \ldots, x_i^1, \ldots, x_n^1\} \in \mathscr{C}_1$ and $X_2 = \{x_1^2, \ldots, x_j^2, \ldots, x_m^2\} \in \mathscr{C}_2$ with $x_i^1, x_j^2 \in \mathbb{R}^{1\times 1}$, $X_1 \in \mathbb{R}^n$, and $X_2 \in \mathbb{R}^m$, we can get two class-conditional probability density functions $f(x|\mathscr{C}_1)$ and $f(x|\mathscr{C}_2)$ [38]. In this work, the Gaussian kernel density at point x_0 is used to calculate the probability density $f(x|X)$

$$\begin{aligned} p(\boldsymbol{x}|\mathscr{C}_k) = f(x_0|X) &= \frac{1}{N\lambda}\sum_{i=1}^{N} K_\lambda(x_0, x_i) \\ &= \frac{1}{N(2\lambda^2\pi)^{\frac{d}{2}}}\sum_{i=1}^{N}\exp\left\{-\frac{1}{2}\left(\frac{\| x_i - x_0 \|}{\lambda}\right)^2\right\} \end{aligned} \tag{1}$$

where K_λ is the Gaussian kernel, λ is the kernel width and computed by $\lambda = 1.06 \cdot \hat{\sigma} \cdot N^{-1/5}$ [39] with the standard deviation $\hat{\sigma}$ of the training data $\{x_i\}_{i=1}^N$.

3.2.2 Bayesian decision: Suppose that the prior probabilities $P(\mathscr{C}_k)$ and the conditional-probabilities density $p(\boldsymbol{x}|\mathscr{C}_k)$ are known for the class $k = 1, 2, \ldots$. Based on the Bayes formula, we have

$$\begin{aligned} P(\mathscr{C}_k|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|\mathscr{C}_k)P(\mathscr{C}_k)}{p(\boldsymbol{x})} \\ p(\boldsymbol{x}) &= \sum_{k=1}^{l} p(\boldsymbol{x}|\mathscr{C}_k)p(\mathscr{C}_k) \end{aligned} \tag{2}$$

then, the posterior probability given random input $\boldsymbol{x}$ can be estimated by (2). Under (2), decisions about $\boldsymbol{x}$ can be made by

$$\text{Decide} \quad \mathscr{C}_k \text{ if } P(\mathscr{C}_k|\boldsymbol{x}) > P(\mathscr{C}_{\setminus k}|\boldsymbol{x}) \tag{3}$$

where $P(\mathscr{C}_{\setminus k}|\boldsymbol{x})$ are the left categories except for the kth category.

The key to calculate (2) and decide (3) is the conditional probability density $p(\boldsymbol{x}|\mathscr{C}_k)$. However, for a high-dimensional feature vector, computing covariances of each pair of the dependent components in the feature vector will suffer a huge computational cost. Therefore, instead of directly operating on the complicated covariances, a cost efficient method is developed based on the Euclidean distance in the following section.

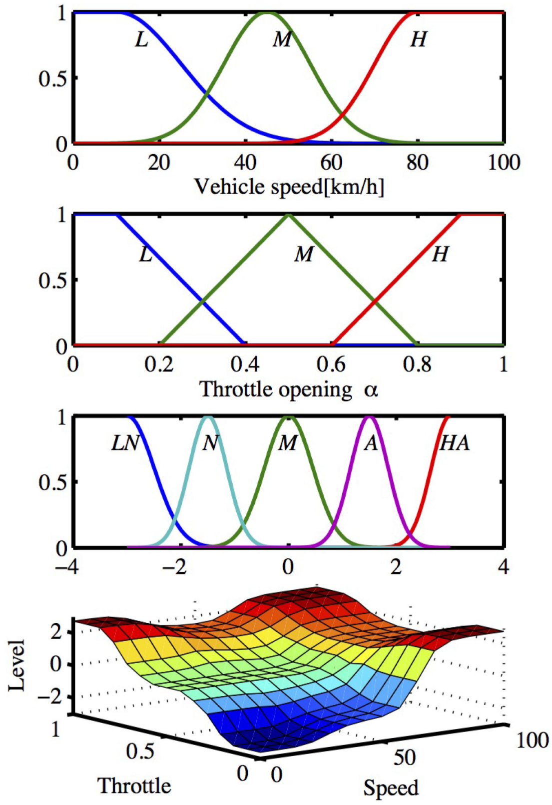

Fig. 2 *Membership function for inputs and output of the FIS. From top to bottom: membership functions of vehicle speed, membership functions of throttle opening, membership functions of output, and whole process mapping of the FIS*

Table 3 Fuzzy rules for definition of driving styles

No.	Input 1	Operator	Input 2	Weight	Output
1	L	and	L	1	LN
2	L	and	M	1	M
3	L	and	H	1	HA
4	M	and	L	1	N
5	M	and	M	1	M
6	M	and	H	1	A
7	H	and	L	1	HA
8	H	and	M	1	A
9	H	and	H	1	HA

3.2.3 Euclidean distance-based decision: It is very easy for Bayesian decision to deal with the case of 1D data, but not for the case of a d-dimension data sequence ($d \geq 2$) because elements in the feature vector are highly dependent, which is computationally expensive to calculate the conditional-probability $p(\boldsymbol{x} | \mathscr{C}_k)$ for high-dimensional data. In order to address this issue, we introduce the Euclidean distance with the Bayesian decision.

Take a 2D dataset with two classes (class A and class B), e.g. (Fig. 3), the posterior probabilities of an element in feature vector $\boldsymbol{x} = (x_1, x_2)$ falling in classes A and B are defined as $f_A(x_l)$ and $f_B(x_l)$ for $l = 1, 2$, respectively. Here, we set $f_A(x_l) = P(A | x_l)$, $f_B(x_l) = P(B | x_l)$. Given a random input $\boldsymbol{x}^* = (x_1^*, x_2^*)$, the relevant posterior probabilities $f_A(x_l^*)$ and $f_B(x_l^*)$ are calculated by (1) and (2), respectively. Then, projecting inputs and their corresponding posterior probability into the first quadrant in Fig. 3, we can get $A = (f_A(x_1^*), f_A(x_2^*))$ and $B = (f_B(x_1^*), f_B(x_2^*))$. The Euclidean distance between A and B is defined and computed by

$$\begin{aligned} d_A &:= \| f_A(x_1^*)^2 + f_A(x_2^*)^2 \|^{(1/2)} \\ d_B &:= \| f_B(x_1^*)^2 + f_B(x_2^*)^2 \|^{(1/2)} \end{aligned} \tag{4}$$

The joint density function $p((x_1^*, \ldots, x_d^*) | \mathscr{C}_k)$ of the d-dimensional feature vectors is decoupled into several simple densities of 1D feature scalar, and thereby, the Bayesian decision is transformed to the Euclidean distance-based decision. The classification rule on the basis of the Euclidean distance is defined by

- Decide class A if $(d_A > d_B) \wedge (|d_A - d_B| > \epsilon)$ (Fig. 3a),
- Decide class B if $(d_A < d_B) \wedge (|d_A - d_B| > \epsilon)$ (Fig. 3b),
- Decide class M if $|d_A - d_B| \leq \epsilon$ (Fig. 3c),

where M is the vague class between classes A and B, ϵ is the threshold with $\epsilon \in \mathbb{R}^+$. We should note that when $\boldsymbol{x}$ is in a d-dimensional Euclidean space $\mathbb{R}^d$ with $d = 3$, the Euclidean distance is the radius of a sphere. Therefore, the extended Euclidean distance in a d-dimensional space can be formulated as

$$d_{\mathscr{C}_k}(x_i^*) = \left\| \sum_{i=1}^{d} f_{\mathscr{C}_k}(x_i^*)^2 \right\|^{(1/2)}, \quad k = 1, 2, \ldots \tag{5}$$

for a new test input $\boldsymbol{x}^* = (x_1^*, \ldots, x_i^*, \ldots, x_d^*)$.

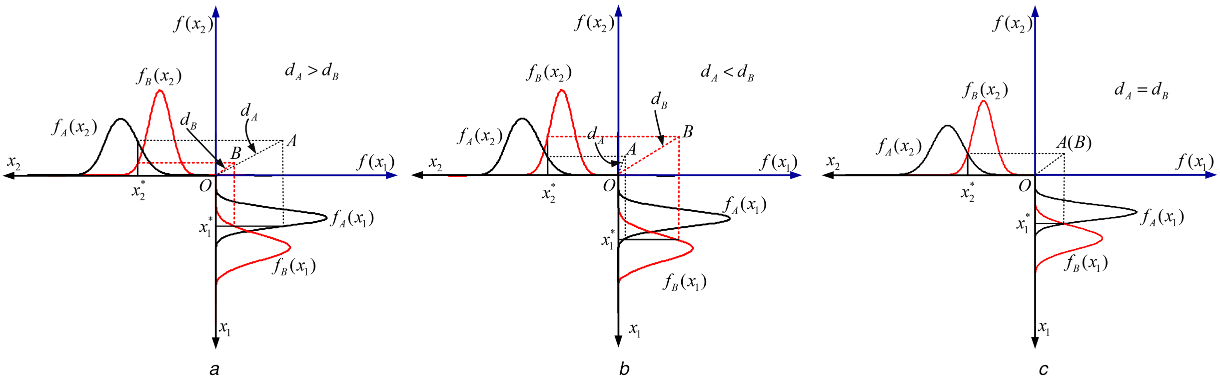

Fig. 3 *Schematic diagram of the proposed method for driving style recognition with kernel density estimation and Euclidean distance. Here,* $f_{(\cdot)}(x_k) = \boldsymbol{P}(\cdot \mid x_k)$ *represents the posterior probability*
(a) Decide input data $(x_1^*, x_2^*) \in \boldsymbol{A}$, ***(b)*** Decide input data $(x_1^*, x_2^*) \in \boldsymbol{B}$, ***(c)*** Decide input data $(x_1^*, x_2^*) \in$ fuzzy class $\boldsymbol{M}$

Training
1: Input training data sequence $\mathcal{X}^k = \{\boldsymbol{x}_i^k\}$ for $k = 1, 2$, $i = 1, \cdots, n$, $\boldsymbol{x}_i^k = \{x_{i,l}^k\}$, $l = 1, \cdots, d$.
2: Get $f(x_{i,l}^k | \mathcal{C}_k)$ from (1) for each element in feature vectors.
Testing
1: Input new data $\boldsymbol{x}_i^* = \{x_{i,l}^*\}$, $i = 1, 2, 3, \cdots$.
2: **for** i, l
3: Get $f(x_{i,l}^* | \mathcal{C}_k)$ from $f(x_{i,l}^k | \mathcal{C}_k)$ for $k = 1, 2$
4: Get $P(\mathcal{C}_k | x_{i,l}^k)$ under (2) and set $f_{\mathcal{C}_k}(x_{i,l}^*) = P(\mathcal{C}_k | x_{i,l}^k)$ for $k = 1, 2$
5: Get $d_{\mathcal{C}_k(x_{i,l}^*)} \Leftarrow (5)$
6: **if** $d_{\mathcal{C}_k(x_{i,l}^*)} > d_{\mathcal{C}_{\setminus k}(x_{i,l}^*)}$
7: ***if*** $d_{\mathcal{C}_k(x_{i,l}^*)} - d_{\mathcal{C}_{\setminus k}(x_{i,l}^*)} \in (\underline{\epsilon}, \bar{\epsilon})$ (see Table 5)
8: then $\boldsymbol{x}_i^* \in$ level $s = \mathcal{S}$
9: ***end if***
10: **else** $d_{\mathcal{C}_k(x_{i,l}^*)} \le d_{\mathcal{C}_{\setminus k}(x_{i,l}^*)}$
11: ***if*** $\| d_{\mathcal{C}_k(x_{i,l}^*)} - d_{\mathcal{C}_{\setminus k}(x_{i,l}^*)} \| \in (\underline{\epsilon}^\star, \bar{\epsilon}^\star)$
12: then $\boldsymbol{x}_i^* \in$ level $s = \mathcal{S}$
13: ***end if***
14: **end if**
15: **end for**
16: Output the results for data sequence $\{\boldsymbol{x}_i^k\}$

Fig. 4 *Algorithm of our proposed statistical-based approach*

Table 4 Threshold values of $(\underline{\epsilon}, \bar{\epsilon})$ and $(\underline{\epsilon}^\star, \bar{\epsilon}^\star)$

$(\underline{\epsilon}, \bar{\epsilon})$	Aggressive level	$(\underline{\epsilon}^\star, \bar{\epsilon}^\star)$	Normal level
(0.5, —)	3	(0.5, —)	−3
(0.2, 0.5]	2	(0.1, 0.5]	−2
(0.02, 0.2]	1	(0.02, 0.1]	−1
(0, 0.02]	0^+	[0, 0.02]	0^-

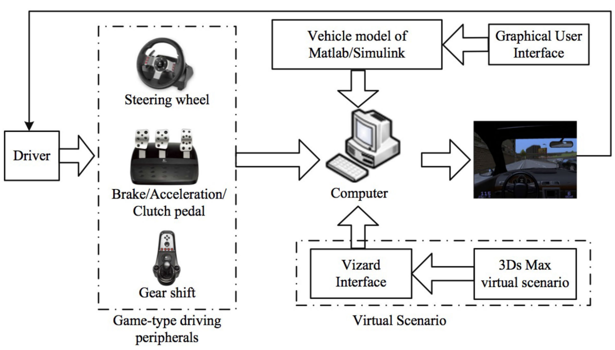

Fig. 5 *Driving simulator for data collection [40]*

3.2.4 Classification algorithm: Based on the above description, a classification method based on the conditional-kernel density function $f_{\mathscr{C}_k}(x)$ and the Euclidean distance $d_{\mathscr{C}_k}$ is developed, which allows computation of decision-making with high-dimensional data efficient. In order to represent different levels of driving styles easily, a numerical set is defined as $\mathscr{S} = \{s \mid -3, -2, -1, 0, 1, 2, 3\}$. A lager value of s indicates a more aggressive driving style. The classification algorithm is shown in Fig. 4, where the values of threshold $(\underline{\epsilon}, \bar{\epsilon})$ and $(\underline{\epsilon}^\star, \bar{\epsilon}^\star)$ are listed in Table 4. For training step 3, the prior probability $P(\boldsymbol{x})$ is set to $1/k$, k is the number of categories of training data. In this work, two typical driving styles, i.e. aggressive and normal, are considered and we set $k = 2$.

4 Experiment and data collection

4.1 Driving simulator

All the experiment data were obtained through a driving simulator as shown in Fig. 5. The driving simulator consists of four main parts: vehicle dynamics model, game-type driving peripherals, virtual driving environment, and human driver. The inputs applied by a human driver, including the steering angle, throttle opening, and braking force, were recorded through the game-type driving peripherals. The vehicle-dependent data such as vehicle speed and vehicle position were recorded from MATLAB. An eight degree-of-freedom vehicle model in [10] was used (Table 5) and validated in Carsim [41]. The data collection and processing systems were developed using MATLAB/Simulink (2015b, 64-bit version) and Vizard 5.0 software. The virtual driving environment was designed by 3Ds Max software and saved as FILENAME.IVE files that the Vizard can read.

4.2 Driving environment

In this work, we fixed our attention on the longitudinal behaviour when following a curvy road. The road factors have a big influence on the performance of driving style recognition. The road model must have the same scale as the road in the real driving environment. Therefore, the requirements of road model were subject to the following criteria: continuity of the path, continuity of the curvature, and differentiability of the set path.

Drivers were instructed to drive in their lane and follow the reference path designed using Carsim data, as shown in the top figure of Fig. 6. The length of the curvy road was 2,247 m and the lane width was 3.70 m. The path consists of a set of simple path elements that have curvature including straight segments (zero curvature), arcs (constant curvature), and clothoid (linearly varying curvature). Due to the limitations of experimental equipment, the effects of road slope, weather condition, and traffic flow were not considered in this work.

Table 5 Vehicle model parameters

Symbol	Meaning	Value	Unit
m	car mass	2100	kg
a	font axis from CoG	1256	mm
b	rear axis from CoG	1368	mm
I_z	car inertia	2549	kg·m^2
C_{α_f}	corner stiffness of front tires	107,850	N/rad
C_{α_r}	corner stiffness of rear tires	106,510	N/rad

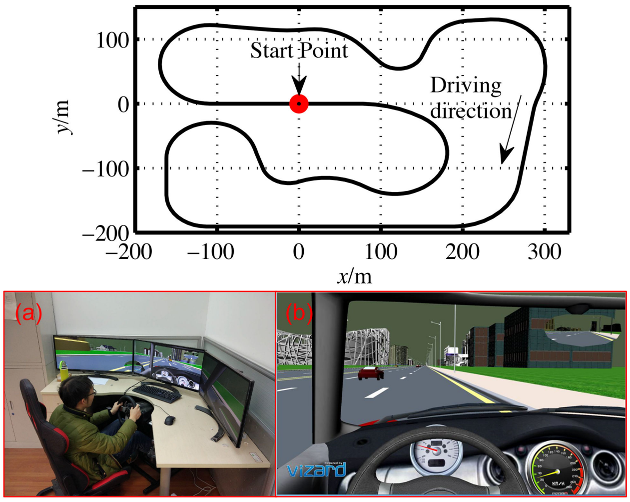

Fig. 6 *Top: The road profile. Bottom*
(a) One of the driver participants, ***(b)*** Screen shot of driving scenarios

4.3 Data collection procedure

All the driving data were collected at a sampling frequency at 50 Hz in the driving simulator, including vehicle speed (v), throttle opening (α), acceleration, vehicle position, steering angle, and yaw angle. Eight driver participants were selected in our experiment, four of them were aggressive drivers and the other four were normal drivers. Each participant should be labelled as aggressive or normal before running an experiment. Each subject driver was asked to drive in the simulator for ten runs from the start point shown in the bottom figure of Fig. 6. During the experiment procedure, all driver participants followed the specified rules: (i) all driver participants took about 20 min to be familiar with the driving simulator before collecting data; (ii) all participants were in mentally and physically normal states; (iii) the secondary tasks were forbidden, for example, reading a message or answering a phone while driving; (iv) each participant had a rest (about 1 min) before the next run; and (v) each driver driven a car in their own driving style.

5 Recognition performance evaluation

5.1 Cross-validation

The CV method, as one of the most popular evaluation schemes, is used to evaluate the recognition performance of the proposed approach. To do CV, we divide the available training data set into q ($q > 1$) folds evenly. All except one folds are randomly used to train the model and the *hold-out set* or *validation set* is used to assess the trained model. In this work, the driving datasets were evenly divided into nine folds – five folds for training and four folds for testing. The CV assessment makes sure that the training datasets are disjoint from the validation datasets.

In order to evaluate the proposed recognition method, the validation datasets were grouped as aggressive and normal styles to test how well the trained recogniser can identify them from those provided by the aggressive drivers [21]. The correction recognition rate (CRR) of driving style recogniser is defined as

$$\mathrm{CRR}_{\mathrm{agg}} = \frac{N_{\mathrm{agg,agg}}}{\sum_{\star\,\in\,\{\mathrm{agg,norm}\}} N_{\mathrm{agg},\,\star}} \tag{6}$$

for an aggressive driver, and

$$\mathrm{CRR}_{\mathrm{norm}} = \frac{N_{\mathrm{norm,norm}}}{\sum_{\star\,\in\,\{\mathrm{agg,norm}\}} N_{\mathrm{norm},\,\star}} \tag{7}$$

for a normal driver. The first and second subscriptions of $N_{\star,\star}$ represent the real driving style and the driving style recognised by the proposed method, respectively; *agg* and *norm* represent 'aggressive' and 'normal', respectively. Take $N_{\mathrm{agg,norm}}$, e.g. it represents the number of runs that are grouped as aggressive drivers but classified to be normal style.

5.2 Results and analysis

Figs. 7 and 8 show the recognition results for the aggressive drivers and the normal drivers using the proposed recognition approach and the FL approach, respectively. Fig. 9 presents an example of the computed Euclidean distance for drivers with respect to different driving styles using the proposed statistical-based method. We found that for the aggressive driver (top in Fig. 9), most of the Euclidean distance with respect to the aggressive driving style is greater than that with respect to the normal driving style, thus demonstrating that the driver is subject to an aggressive class. For the normal driver (bottom in Fig. 9), most of the Euclidean distance with respect to the aggressive driving style is smaller than that with respect to the normal driving style, thus demonstrating that the driver prefers to drive in a normal style. In what follows, we will analyse and discuss the experiment results from three aspects: feature analysis, efficiency analysis, and robustness analysis.

5.2.1 Feature analysis: From Figs. 7*a* and 8*a*, it can be concluded that the developed method is able to correctly classify aggressive drivers into an aggressive class in most of the time. In addition, we found that the aggressive driver also performs normally or moderately before entering a road curve, for example,

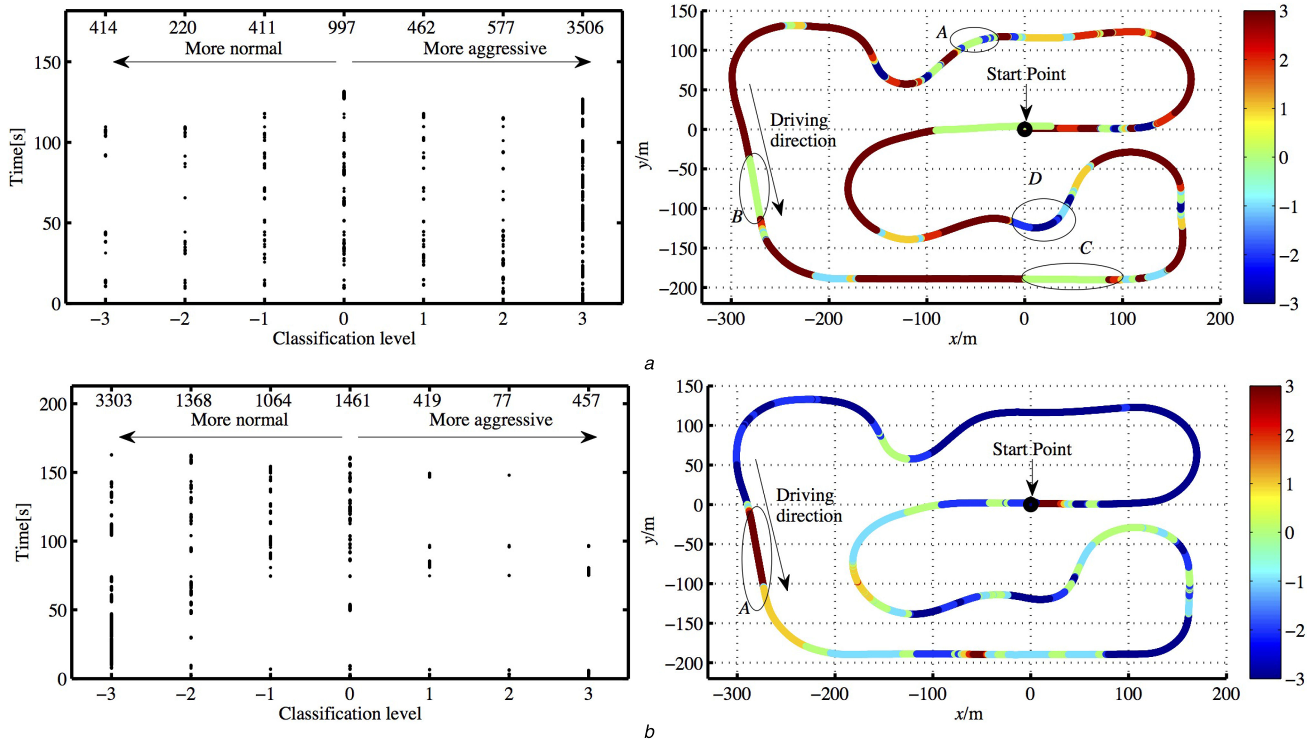

Fig. 7 *Example of the recognition results for*
(a) Aggressive driver, ***(b)*** Normal driver using the developed statistical-based method. Left: classification level; right: classification results

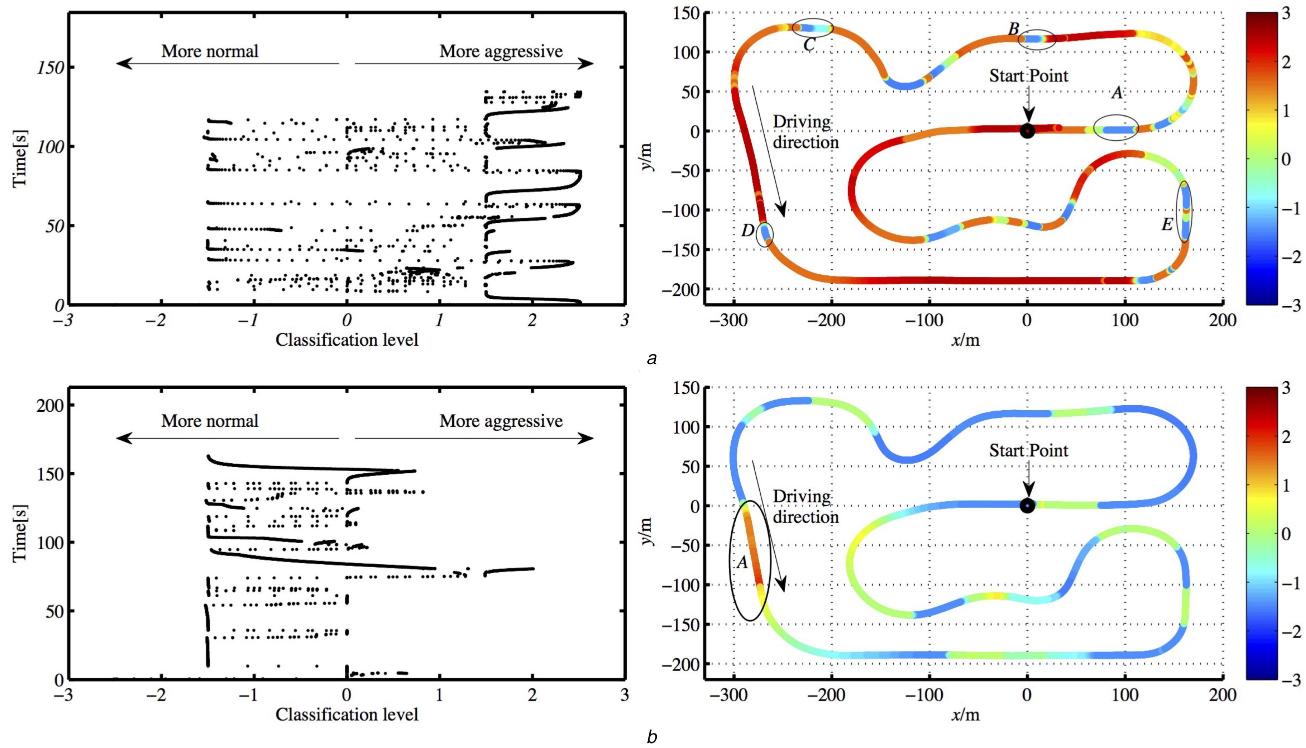

Fig. 8 *Example of the recognition results for*
(a) Aggressive driver, ***(b)*** Normal driver using the FL algorithm. Left: classification level; right: classification results

at parts *A*, *B*, *C*, and *D* of Fig. 7*a* and parts *A*, *B*, *C*, *D*, and *E* of Fig. 8*a*, but after entering the curvy road, they will drive in an aggressive style. The experiment results also consequentially demonstrate the driving behaviour uncertainty, e.g. an aggressive driver could not always drive in an aggressive way and the same results for a normal driver.

For normal drivers in Figs. 7*b* and 8*b*, they barely drive in an aggressive style. Namely, most of the driving data are classified into a normal class. In addition, we found that the normal drivers will act aggressively when driving into a straight road from a curvy road such as the regions *A* in Figs. 7*b* and 8*b*, but they will drive in a normal style after entering a curve.

By comparing Figs. 7 and 8, we found that both FL and our proposed methods could recognise driving styles, however, the FL method highly depends on their membership function design. Notice that the FL could not reach the most aggressive level (i.e. 3) or the most normal level (i.e. −3), but our proposed method could reach the most aggressive or normal level. In addition, we found that from the left plots in Figs. 7 and 8, the FL algorithm could not obtain the driving styles in a form of fixed levels, but falling in a

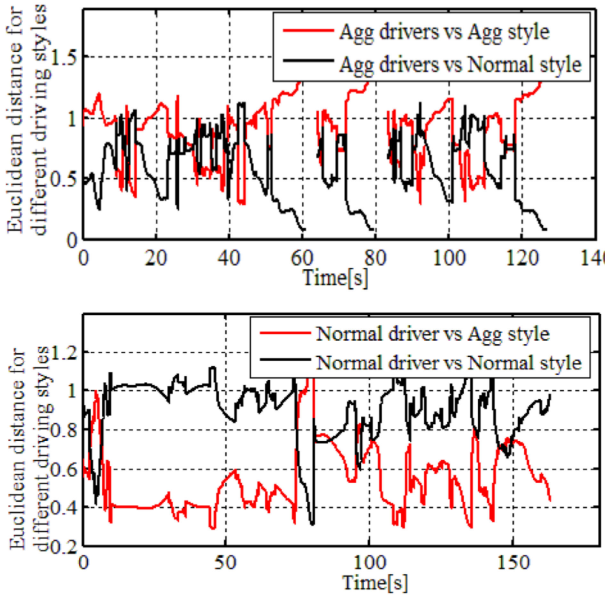

Fig. 9 *Euclidean distance of an aggressive driver (top) and a normal driver (bottom) with respect to two kinds of labelled drivers*

Table 6 Experiment results for all driver participants using the proposed statistical-based recognition algorithm

Types	Driver no.	Levels							CRR_{agg}	CRR_{norm}
		−3	−2	−1	0	1	2	3		
aggressive	1	397	204	292	1177	332	538	3485	0.861	—
	2	467	271	348	830	237	623	3955	0.839	—
	3	414	220	411	997	462	577	3506	0.841	—
	4	158	89	372	1301	283	869	3488	0.906	—
normal	5	4458	644	626	1062	257	554	578	—	0.914
	6	3303	1368	1064	1461	419	77	457	—	0.883
	7	5162	1334	550	1552	77	129	112	—	0.964
	8	3775	1823	640	2193	70	61	37	—	0.980
average									0.862	0.935

Table 7 Comparison of recognition results using the FL approach and the proposed approach

Type	Algorithm		
	FL	Proposed	
CRR_{agg}	0.812	0.861	
	0.882	0.906	
	0.819	0.841	
	0.809	0.839	
Average	0.831	0.862	↑ 3.79%
CRR_{norm}	0.834	**0.883**	
	0.870	0.964	
	0.602	**0.980**	
	0.682	0.914	
Average	0.747	0.935	↑ 22.36%

range of levels. For example, the recognised results of the normal driver by using the FL method in Fig. 8 are mostly falling in the range of [−1.5, 0.5], rather than being fixed at integer values.

5.2.2 Accuracy analysis: Table 6 shows the recognition performance for all drivers using our proposed method. It is obvious that the statistical-based method can obtain a good performance, with CRR_{norm} of 0.935 and CRR_{agg} of 0.862 in average for normal and aggressive drivers, respectively.

From Table 7, we know that the proposed recognition approach is more efficient than the FL approach. More specifically, compared to the FL method, the proposed statistical-based recognition algorithm improves the recognition accuracy by 3.79 and 22.36% for aggressive drivers and normal drivers, respectively.

5.2.3 Robustness analysis: From Table 7, we found that for normal drivers, the results of using the FL algorithm suffer a large variance (CRR_{norm} ranges from 0.602 to 0.870), while the developed approach obtains a small variance (CRR_{norm} ranges from 0.883 to 0.980), which demonstrates that the statistical-based recognition method has a stronger robustness than the FL algorithm. To some extent, the experiment results also indicate that the statistical-based recognition approach can change the problematic recognition task involved with uncertainty into a tractable case.

6 Conclusions

This paper presented a statistical pattern-recognition method by introducing kernel density estimation and Euclidean distance to deal with the driver behaviour uncertainty. We applied the full Bayesian theory to estimate the probability of being aggressive or normal. In addition, the Euclidean distance of each pair of elements in the feature vector was computed to decide the driving style for the test datasets, which can reduce the computational cost. Then, a CV method was used to show the benefits of our developed algorithm by comparing with the FL method. The results show that our developed statistical-based approach shows a strong robustness and can improve the recognition correctness by 3.79 and 22.36% for aggressive drivers and normal drivers, respectively, compared with the FL algorithm.

7 Acknowledgments

The first two authors, Wenshuo Wang and Wei Han, contributed equally to this work.

8 References

[1] Zhou, M., Jin, H., Wang, W.: 'A review of vehicle fuel consumption models to evaluate eco-driving and eco-routing', *Transp. Res. D, Transp. Environ.*, 2016, **49**, pp. 203–218

[2] Martinez, C.M., Heucke, M., Wang, F., *et al.*: 'Driving style recognition for intelligent vehicle control and advanced driver assistance: a survey', *IEEE Trans. Intell. Transp. Syst.*, 2017, doi: 10.1109/TITS.2017.2706978

[3] Li, Y., Wang, J., Chan, C.-Y., *et al.*: 'Develop right-turn real-time crash warning system at arterial access considering driver behaviour', *IET Intell. Transp. Syst.*, 2017, **11**, (1), pp. 44–52

[4] Li, L., Liu, Y., Wang, J., *et al.*: 'Human dynamics based driver model for autonomous car', *IET Intell. Transp. Syst.*, 2016, **10**, (8), pp. 545–554

[5] Wang, J., Wang, J., Wang, R., *et al.*: 'A framework of vehicle trajectory replanning in lane exchanging with considerations of driver characteristics', *IEEE Trans. Veh. Technol.*, 2017, **66**, (5), pp. 3583–3596

[6] Schnelle, S., Wang, J., Su, H., *et al.*: 'A driver steering model with personalized desired path generation', *IEEE Trans. Syst. Man Cybern., Syst.*, 2017, **47**, (1), pp. 111–120

[7] Sagberg, F., Giulio, S., Piccinini, F.B., *et al.*: 'A review of research on driving styles and road safety', *Hum. Factors*, 2015, **57**, (7), pp. 1248–1275

[8] Themann, P., Bock, J., Eckstein, L.: 'Optimisation of energy efficiency based on average driving behaviour and driver's preferences for automated driving', *IET Intell. Transp. Syst.*, 2015, **9**, (1), pp. 50–58

[9] Wang, W., Xi, J.: 'Study of semi-active suspension control strategy based on driving behaviour characteristics', *Int. J. Veh. Des.*, 2015, **68**, (1-3), pp. 141–161

[10] Wang, W., Xi, J., Liu, C., *et al.*: 'Human-centered feed-forward control of a vehicle steering system based on a driver's path-following characteristics', *IEEE Trans. Intell. Transp. Syst.*, 2017, **18**, (6), pp. 1440–1453

[11] Wang, W., Xi, J., Chen, H.: 'Modelling and recognizing driver behavior based on driving data: a survey', *Math. Probl. Eng.*, 2014, **2014**, p. 20

[12] Candamo, J., Shreve, M., Goldgof, D.B., *et al.*: 'Understanding transit scenes: A survey on human behavior-recognition algorithms', *IEEE Trans. Intell. Transp. Syst.*, 2010, **11**, (1), pp. 206–224

[13] Tadesse, E., Sheng, W., Liu, M.: 'Driver drowsiness detection through HMM based dynamic modeling'. IEEE Int. Conf. Robotics & Automation, Hong Kong, China, 2014, pp. 4003–4008

[14] Gadepally, V., Krishnamurthy, A., Özgüner, Ü.: 'A framework for estimating driver decisions near intersections', *IEEE Trans. Intell. Transp. Syst.*, 2014, **15**, (2), pp. 637–646

[15] Akita, T., Inagaki, S., Suzuki, T., *et al.*: 'Hybrid system modeling of human driver in the vehicle following task'. SICE Annual Conf., Kagawa University, Japan, 2007, pp. 1122–1127

[16] Sundbom, M., Falcone, P., Sjöberg, J.: 'Online driver behavior classification using probabilistic ARX models'. The 16th Int. IEEE Annual Conf. Intelligent Transportation Systems, Hague, Netherlands, 2013, pp. 1107–1112

[17] Sekizawa, S., Inagaki, S., Suzuki, T., *et al.*: 'Modeling and recognition of driving behavior based on stochastic switched ARX model', *IEEE Trans. Intell. Transp. Syst.*, 2007, **8**, (4), pp. 593–606

[18] Shi, B., Xu, L., Jiang, H., *et al.*: 'Comparing fuel consumption based on normalised driving behaviour: a case study on major cities in China', *IET Intell. Transp. Syst.*, 2017, **11**, (4), pp. 189–195

[19] Angkititrakul, P., Terashima, R., Wakita, T.: 'On the use of stochastic driver behavior model in lane departure warning', *IEEE Trans. Intell. Transp. Syst.*, 2011, **12**, (1), pp. 174–183

[20] Wang, W., Zhao, D.: 'Evaluation of lane departure correction systems using a regenerative stochastic driver model', *IEEE Trans. Intell. Veh.*, 2017, **2**, (3), pp. 221–232

[21] Zhang, Y., Lin, W.C., Chin, Y.S.: 'A pattern-recognition approach for driving skill characterization', *IEEE Trans. Intell. Transp. Syst.*, 2010, **11**, (4), pp. 905–916

[22] Higgs, B., Abbas, M.: 'Segmentation and clustering of car-following behaviour: recognition of driving patterns', *IEEE Trans. Intell. Transp. Syst.*, 2015, **16**, (1), pp. 81–90

[23] Qi, G., Du, Y., Wu, J., *et al.*: 'Leveraging longitudinal driving behaviour data with data mining techniques for driving style analysis', *IET Intell. Transp. Syst.*, 2015, **9**, (8), pp. 792–801

[24] Wang, W., Xi, J., Zhao, D.: 'Driving style analysis using primitive driving patterns with Bayesian nonparametric approaches'. 2017, arXiv:1703.09744

[25] Chu, D., Deng, Z., He, Y., *et al.*: 'Curve speed model for driver assistance based on driving style classification', *IET Intell. Transp. Syst.*, 2017, **11**, (8), pp. 501–510

[26] Schubert, B.: 'Evaluating the utility of driving: toward automated decision making under uncertainty', *IEEE Trans. Intell. Transp. Syst.*, 2012, **13**, (1), pp. 354–364

[27] Khushaba, R.N., Kodagoda, S., Lal, S., *et al.*: 'Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm', *IEEE Trans. Biomed. Eng.*, 2011, **58**, (1), pp. 121–131

[28] Klir, G.J.: '*Uncertainty and information: foundations of generalized information theory*' (John Wiley & Sons, New York, 2005)

[29] Quintero, M.C.G., López, J.O., Pinilla, A.C.C.: 'Driver behaviour classification model based on an intelligent driving diagnosis system'. IEEE Int. Conf. Intelligent Transportation Systems, Anchorage, AK, 2012, pp. 894–899

[30] Higgs, B., Abbas, M.: 'A two-step segmentation algorithm for behavioural clustering of naturalistic driving styles'. IEEE Annual Conf. Intelligent Transportation Systems, Hague, Netherlands, 2013, pp. 857–862

[31] Kasper, D., Weidl, G., Dang, T.: 'Object-oriented Bayesian networks for detection of lane change maneuvers', *Intell. Transp. Mag.*, 2012, **4**, (1), pp. 1–10

[32] Healey, J.A., Picard, R.W.: 'Detecting stress during real-world driving tasks using physiological sensors', *IEEE Trans. Intell. Transp. Syst.*, 2005, **6**, (2), pp. 156–166

[33] Miyajima, C., Nishiwaki, Y., Ozawa, K., *et al.*: 'Driver modeling based on driving behavior and its evaluation in driver identification', *Proc. IEEE*, 2007, **95**, (2), pp. 427–437

[34] Wang, W., Xi, J., Chong, A., *et al.*: 'Driving style classification using a semi-supervised support vector machine', *IEEE Trans. Human-Mach. Syst.*, 2017, **47**, (5), pp. 650–660

[35] Richard, C.M., Campbell, J.L., Lichty, M.G., *et al.*: 'Motivations for speeding'. Volume I: Summary report, Report No. DOT HS 811 658, National Highway Traffic Safety Administration, Washington, DC, 2012

[36] Hülnhagen, T., Dengler, I., Tamke, A., *et al.*: 'Maneuver recognition using probabilistic finite-state machines and fuzzy logic'. IEEE Intelligent Vehicles Symp., San Diego, CA, USA, 2010, pp. 65–70

[37] Wahab, A., Quek, C., Tan, C.K., *et al.*: 'Driving profile modeling and recognition based on soft computing approach', *IEEE Trans. Neural Netw.*, 2009, **20**, (4), pp. 563–582

[38] Duda, R.O., Hart, P.E., Stork, D.G.: 'Pattern classification', 2nd, 2000

[39] Wang, W., Liu, C., Zhao, D.: 'How much data are enough? a statistical approach with case study on longitudinal driving behavior', *IEEE Trans. Intell. Veh.*, 2017, **2**, (2), pp. 85–98

[40] Wang, W., Xi, J., Wang, J.: 'Human-centred feed-forward control of a vehicle steering system based on a driver's steering model'. IEEE American Control Conf., Chicago, IL, USA, 2015, pp. 3361–3366

[41] Xi, J., Zong, Y., Wang, W.: 'Research on virtual experimental teaching platform of vehicle electronic control', *Lab. Res. Explor.*, 2015, **34**, (4), pp. 79–83