Accident Analysis and Prevention 129 (2019) 289–298

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Quantifying visual road environment to establish a speeding prediction model: An examination using naturalistic driving data

Bo Yu[a,b,*], Yuren Chen[a], Shan Bao[b]

[a] Key Laboratory of Road and Traffic Engineering of the Ministry of Education, School of Transportation Engineering, Tongji University, 4800 Cao'an Highway, Shanghai, 201804, China

[b] University of Michigan Transportation Research Institute, 2901 Baxter Road, Ann Arbor, MI, 48109, USA

ARTICLE INFO

Keywords:
Speeding prediction
Drivers' visual road environment
Naturalistic driving data
Random forests
Pre-warning system

ABSTRACT

Speeding is one of the major contributors to traffic crashes. To solve this problem, speeding prediction is recognized as a critical step in a pre-warning system. While previous studies have shown that speeding is affected by road environmental design, research in predicting speeding behavior through road environment features has not yet been conducted. Furthermore, there is a large discrepancy between actual and perceived road environmental information given that a driver's visual perception plays a crucial role as the dominant source of information in determining driver's behavior. Thus, this paper aims to establish a speeding prediction model based on quantifying the visual road environment to improve the design of pre-waring systems, which can predict whether drivers are going to speed and provide them with visual or/and audio warnings about their current driving speed and the speed limit prior to the occurrence of speeding behavior. Twenty input variables derived from three categories including visual road environment parameters, vehicle kinematic features, and driver characteristics were considered in the proposed speeding prediction model. Especially, the road environmental design factors consisting of the visual road geometry and visual roadside environment as perceived by the driver's eyes were quantified using a visual road environment model. Field experiments were conducted to collect naturalistic driving data concerning speeding behavior on the typical two-lane mountainous rural highways in five provinces of China. Random Forests, an ensemble learning method for regression and classification, were applied to build the speeding prediction model and variable importance was calculated. Additionally, logistic regression was used as a supplement to further investigate factors impacting on speeding behavior. A speeding criterion was defined with two levels in this study: a lower level (exceeding the posted speed limit) and a higher level (10% above the posted speed limit). Under both levels of the speeding criterion, the speeding prediction model performed well with high accuracy (over 85%). This model could use the value of the variables obtained from the current position to predict drivers' speeding behavior at the future position located a sighting distance away. This interval was sufficient for a pre-warning system to give a speeding warning that a driver with normal perception-reaction time (around 2.5 s) could respond to. Findings in this study can be used to effectively predict speeding in advance and help to reduce speeding-related traffic accidents.

1. Introduction

Speeding is recognized as one of the major contributors to traffic crashes. It has been observed that about one-third of the fatal road accidents occur because of speeding in many countries (Mehmood, 2009). For example, in 2016, speeding resulted in 27% of total traffic fatalities in the United States, with the loss of 10,111 lives (NHTSA, 2018). In Europe, the European Road Safety Observatory (ERSO) identified speeding as a contributing factor for approximately 30% of fatal crashes (Viallon and Laumon, 2013). Similar results were reported in Australia by the Australian Transport Council (2011). Speeding lengthens stopping distances and reduces drivers' response time in emergency situations, resulting in increased crash risks (Broughton et al., 2009). Even small increases in speed substantially raise the risk and severity of accidents (Chevalier et al., 2016; Elvik et al., 2009).

Speeding is a complex problem consisting of multiple interacting

* Corresponding author at: Key Laboratory of Road and Traffic Engineering of the Ministry of Education, School of Transportation Engineering, Tongji University, 4800 Cao'an Highway, Shanghai, 201804, China.

E-mail address: 1yubo@tongji.edu.cn (B. Yu).

https://doi.org/10.1016/j.aap.2019.05.011
Received 6 November 2018; Received in revised form 13 May 2019; Accepted 14 May 2019
Available online 06 June 2019
0001-4575/

factors. From the perspective of road environmental design, Friedman (2006) demonstrated that road attributes such as lane width, road alignments, the number of lanes, and the availability of shoulders affected drivers' speed choices. If road design is very forgiving, such as wide shoulders, wide lanes and no curves, drivers will be more confident, resulting in more speeding (Shinar, 2017). Under different landscape types, drivers have different visual experiences which can significantly affect their speed choice (Antonson et al., 2009). The low illumination, monotonous environment and few references in tunnels reduced drivers' speed perception ability and may induce speeding behaviors (Wan et al., 2018). In an open landscape, drivers were more likely to generate speeding behavior than in a landscape with obstacles (Antonson et al., 2014). Human elements were also closely associated with drivers' speeding behavior. Personality traits such as sensation seeking and driving anger appeared to be correlated with a greater likelihood of speeding (Tay et al., 2003; Cristea et al., 2013). Drivers were more willing to comply with the speed limit as their ages and education level increased (Mannering, 2009). Compared with male drivers, females had odds ten times lower of being repeatedly high-range speeding offenders (Watson et al., 2015). Besides, as the number of years holding a driving license increased, the probability of speeding went down (Warner et al., 2010).

Many strategies and countermeasures have been taken to reduce speeding-related traffic accidents. In terms of traffic laws, setting lower speed limits and enhancing the enforcement of punishments were generally suggested and adopted to reduce speeding (Matírnez et al., 2013; Richter et al., 2006). From the perspective of road environmental design, speed bumps and channelization (like channelized islands) were the two main methods to solve the speeding issue (Antić et al., 2013; Yuan et al., 2012; Yu et al., 2019). Additionally, warning systems were developed to improve speed limit compliance. For example, intelligent speed adaption/assistant (ISA) systems compared the vehicle's current speed with the speed limit to generate feedback (Warner and Åberg, 2008). However, ISA belongs to the post-warning system, and sometimes drivers cannot take actions to avoid crashes in time since ISA provides warnings based on the driver's current driving speed (Zhao and Wu, 2013). Pre-warning systems refer to systems that can predict whether drivers are going to speed and provide them with visual or/and audio warnings about their current driving speed and the speed limit prior to the occurrence of speeding behavior. Pre-warning systems can offer driver more time to respond and control vehicles, so this system can help to reduce speeding effectively and make drivers more willing to comply with speed limits (Zhao and Wu, 2013). A pre-warning system can provide real-time monitoring for speeding behavior during the whole driving process, while a speed limit sign is generally erected at intervals along the road or some locations where the speed limit changes. Thus, pre-warning systems can be used as a good supplement to speed limit signs for better mitigate the speeding issue.

Speeding prediction is a critical step in a pre-warning system, with many speeding prediction models presented in prior studies. The theory of planned behavior (TPB) was used widely to predict drivers' intention of exceeding the posted speed limit (Cristea et al., 2013). The TPB mostly focused on driver characteristics, including drivers' attitudes, subjective norms, perceived behavioral control, moral norms, anticipated regret, demographic data, past speeding behavior, etc. All these measures were taken by self-report questionnaires (Conner et al., 2007). Nevertheless, the TPB cannot predict speeding behavior in real time. The task capability interface model was established to predict unintentional speeding behavior by comparing capacity with task demand (Brandenburg et al., 2010). However, the validity of the method in predicting speeding behavior was not tested by data. An intelligent speeding prediction system (ISPS) was developed by Zhao et al. (2013) based on a mathematical model of driver speed control. This speed control model could provide drivers with warnings in advance, since many factors were taken into consideration, such as speed, acceleration, throttle, the post speed limit, the density of outside texture, traffic flow, drivers' personality, etc. The effectiveness and acceptance of the ISPS were validated by a driving simulator experiment (Zhao and Wu, 2013).

Currently, speeding behavior prediction mainly focuses on driver characteristics and vehicle variables. While previous studies have shown that speeding is impacted by road environmental design (Friedman, 2006; Shinar, 2017), research in predicting speeding behavior through road environment features has not yet been conducted. Besides, some scholars have presented that there was a large discrepancy between actual and perceived road environmental information (Bidulka et al., 2002; Hassan and Sarhan, 2012). Although road environmental design satisfies the requirements of design indicators, sometimes drivers may make erroneous speed choices according to the information perceived by their eyes. During the process of driving, drivers go through a continuous perception-decision-action loop, so drivers' visual perception as the dominant source of information plays a crucial role in determining driver's behavior (Peters and Nilsson, 2007; Cheng et al., 2016). Thus, in this paper, we combine road environment perceived by drivers' eyes with driver characteristics and vehicle kinematic features to predict speeding. A drivers' visual road environment model was established to quantify the road information from drivers' visual perception. We hope that our research results could help to provide a new perspective for predicting speeding behavior and improve the design of pre-warning systems with high accuracy and sufficient advance time.

The remaining parts of this paper are organized as follows: The next section presents a description of the driver's visual road environment model, naturalistic driving studies, and a Random Forests algorithm. In Section 3, the Random Forests algorithm is applied to establish a speeding prediction model and calculate the importance scores of input variables. Logistic regression is used as a supplement to investigate factors impacting on speeding behavior. Then, Section 4 gives the discussion on those contributors to speeding behavior, and conclusions are drawn in Section 5.

2. Methodology

This Methodology section consists of three parts: (1) A visual road environment model is established to quantify the road environmental design perceived by drivers' eyes; (2) A description about naturalistic driving studies is provided, including the speeding criterion, a definition of speeding prediction, input variables, experimental sites, and data collection and extraction; (3) A Random Forests Algorithm is used to establish a speeding prediction model.

2.1. Drivers' visual road environment model

Road geometry and roadside environment are two crucial components of road environment design, which greatly affect driving behavior and traffic safety. Prior studies found that road environment obtained from drivers' visual perception was significantly different from the actual one, especially when there were constrained topography and unreasonable combinations of alignments (Hassan and Sarhan, 2012; Yu et al., 2018). For example, Bidulka et al. (2002) investigated that overlapping vertical alignments led to the erroneous perception of horizontal curvature, and their results showed that when there were overlapping crest curves, the perceived horizontal curvature would be shaper than the actual value, while overlapping sage curves made the horizontal curvature appear less shaper in drivers' eyes. Hassan and Sarhan (2012) also presented that drivers could misperceive the road alignments when horizontal and vertical curves were combined, and this might result in erroneous drivers' speed choice. Driving behavior was not directly influenced by the objective road environment but rather by the perceived road environment with additional properties (Weller et al., 2008). Drivers' wrong perception of road environment was partly due to the current road design practice that focused on two-dimensional analyses and did not take three-dimensional visual quality

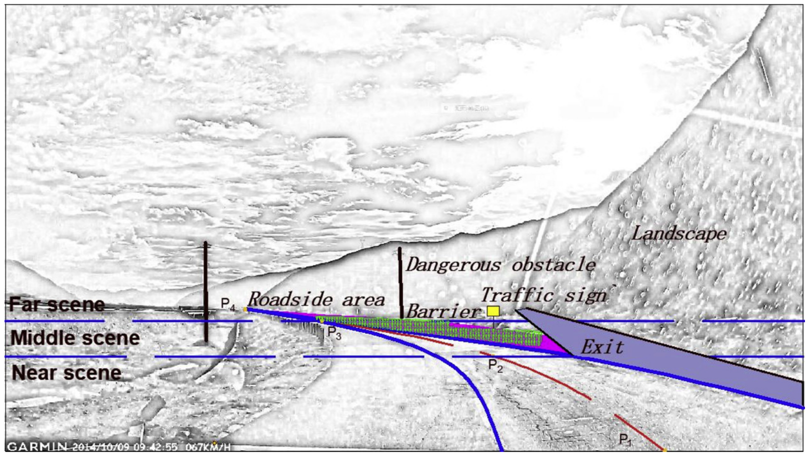


Fig. 1. Drivers' visual environment model.

into consideration (Hassan and Easa, 2003).

Thus, this study tries to establish a visual road environment model to describe and quantify the road environmental design perceived by drivers' eyes. As shown in Figs. 1 and 2, this model is composed of two parts, including the visual road geometry and visual roadside environment. This study extends our previous studies on visual road geometry (Yu et al., 2016, 2018) to a more comprehensive visual road environment model considering roadside information from drivers' visual perception, such as roadside areas, landscapes, barriers, road traffic signs, etc.

Fig. 2 illustrates that visual road geometry is composed by the visual lane and roadside areas. During the driving process, there is always a "lane" the visual field of the driver, and drivers can perceive the "lane" according to the specific road environment. This "lane" is called drivers' visual lane. A drivers' visual lane model was developed based on the Catmull-Rom spline in our previous study (Yu et al., 2016). The Catmull-Rom spline is a kind of cubic interpolating splines and can form arbitrary shapes, so it can be used to detect road geometry alignments and lane boundaries. Comparing with quadratic curves used in most models and other cubic curves such as a Cubic Bezier spline, a Cubic B spline, etc., a Catmull-Rom spline showed higher accuracy in fitting different road alignments and lane boundaries, even those complex conditions like a S-curve which is hard to be quantified by most of the lane models (Yu et al., 2016; Wang et al., 2000). Owing to the advantage of Catmul-Rom splines, the visual lane can be used to depict road alignments in drivers' field of vision, and the validity of the visual lane model has been demonstrated by us (Yu et al., 2016, 2018).

Visual road geometry in this study is modeled by combining the previous visual lane model with the quantification of roadside areas. As shown in Fig. 2, a visual road geometry model sets the bottom-left corner of the drivers' visual field as the origin, and then builds pixel coordinate system. The visual lane is fitted by the Catmull-Rom spline, and the matrix equation of Catmull-Rom splines is as follows:

$$P(t) = [1\ t\ t^2\ t^3]\begin{bmatrix} 0 & 1 & 0 & 0 \\ -0.5 & 0 & 0.5 & 0 \\ 1 & -2.5 & 2 & -0.5 \\ -0.5 & 1.5 & -1.5 & 0.5 \end{bmatrix}\begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} \quad (1)$$

where t denotes the interpolation variable, $t \in [0,1]$; (P_1, P_2, P_3, P_4) are control points of a Catmull-Rom spline.

There are four control points (P_1, P_2, P_3, P_4) in a Catmull-Rom spline. Locations of control points are determined by the shape of a Catmull-Rom spline, so positions of control points are diverse when drivers' visual lane models change. The information at control points (P_1, P_2, P_3, P_4) can be denoted by (S_i, X_i, Y_i) (i = 1,2,3,4). Specifically, S_i denotes the cumulative length of visual lane centerline at P_i point (pixels), and (X_i, Y_i) are image coordinates of P_i point (pixels). In addition, f_i is the tangential angle of the visual lane centerline at each control point (radians).

Four horizontal lines pass through four control points and divide drivers' visual field into three different regions - namely, "near scene", "middle scene", and "far scene". Curve length and curvature of the visual lane centerline in three feature regions are regarded as shape

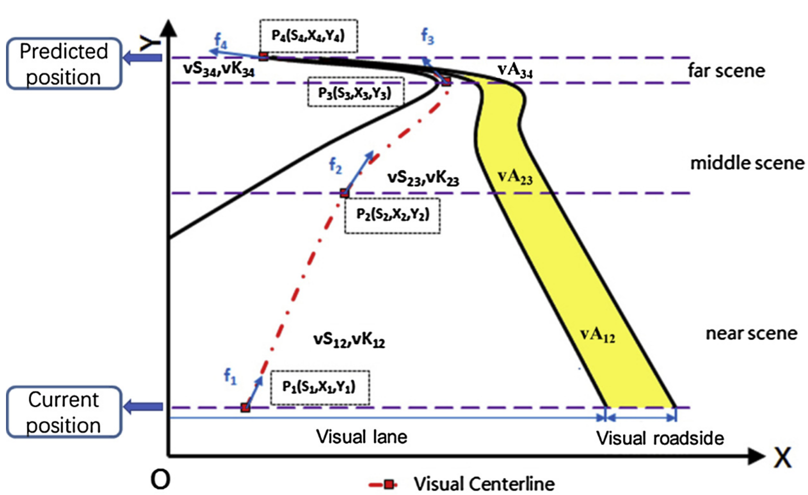


Fig. 2. Drivers' visual road geometry model.

parameters of the visual lane, denoted by [$vS_{i(i+1)}$, $vK_{i(i+1)}$] (i = 1,2,3). They can be calculated by Eqs. (2) and (3):

$$vS_{i(i+1)} = S_{i+1} - S_i \quad (2)$$

$$vK_{i(i+1)} = \frac{f_{i+1} - f_i}{vS_{i(i+1)}} \quad (3)$$

Where: $i = 1,2,3$; $vS_{i(i+1)}$ is the visual curve length between control point P_i and P_{i+1} (pixels);$vK_{i(i+1)}$ is the visual curve curvature between control point P_i and P_{i+1}, namely, the unit rate of change of tangential angle (radians); f_i is the tangential angle at control point P_i(radians); S_i is the visual curve cumulative length at control point P_i(pixels).

The visual roadside represents the strip of land along a road in the drivers' visual field, and its horizontal width ranges from the right-side lane marker of drivers' visual lane to the edge of the clear zone. As Fig. 2 demonstrates, the area of roadside in three feature regions describes the geometric information of visual roadside, denoted by $vA_{i(i+1)}$ ($i = 1,2,3$) (pixel^2). In total, there are nine shape parameters for a visual road geometry model, namely, [$vS_{i(i+1)}$, $vK_{i(i+1)}$, $vA_{i(i+1)}$] ($i = 1,2,3$).

As shown in Fig. 1, drivers' visual roadside environment consists of landscapes, barriers, exits, dangerous obstacles, and traffic signs. Roadside landscapes are divided into five types, including trees or other plants, mountains, houses or streets, tunnels, and open fields. Barriers are categorized into three different groups, namely, no barriers, discontinuous concrete barriers for a warning purpose, and continuous barriers. Whether there are dangerous obstacles or exits will be coded in the model. Traffic signs erected at the roadside will also be coded, including speed limits, warning signs with suggested speed (such as dangerous road sections), etc., because they give instructions or offer information to drivers, leading to a significant impact on driving behavior.

2.2. Naturalistic driving studies

2.2.1. Speeding criterion and definition of speeding prediction

In a previous study, Zhao and Wu (2013) defined a speeding criterion with two levels: lower (the posted speed limit plus 1 mph) and higher (the posted speed limit plus 5 mph) speed threshold. In China, if drivers exceed the posted limit but less than 10%, three points will be deducted from driving licenses, but they do not need to pay fines. If the driving speed reaches 10% above the posted speed limit, drivers must pay fines and their driving license will also be deducted corresponding points. Therefore, in our study, we present a new definition for the speeding criterion, including a lower level (exceeding the posted speed limit) and a higher level (10% above the posted speed limit).

The speed prediction model, in this study, use input variables that are obtained at the current position to predict whether drivers will generate speeding behavior at the future position located a sighting distance away. Sighting distance is the length of roadway ahead that is visible to a driver (Transportation Officials, 2011). That is, the prediction position is at the end of the driver's visual field, while the current position is at the beginning of this visual field, as shown in Fig. 2.

2.2.2. Input variables

In total, 20 input variables are used to predict speeding, which can be divided into three categories:

(1) **visual road environment parameters**, representing road environmental design perceived by driver's eyes that includes the visual road geometry parameters and roadside environment parameters as follows:
   1) *visual road geometry parameters*: visual curve length ($vS_{i(i+1)}$ ($i = 1,2,3$)), curvature ($vK_{i(i+1)}$ ($i = 1,2,3$)) and area of visual roadside ($vA_{i(i+1)}$ ($i = 1,2,3$)) in the three feature regions;
   2) *visual roadside environment parameters*: landscapes, barrier types, exits, dangerous obstacles, and the posted speed limit;

(2) **vehicle kinematic features**: driving speed, acceleration, the presence of speeding at the current position;

(3) **driver characteristics**: age, gender, years of driving experience.

2.2.3. Experimental site

Data were acquired from naturalistic driving studies in five provinces of China, including Shandong, Anhui, Tibet, Zhejiang, and Jiangxi. There were more than 50,000 km of driving data in total, and each driver had at least 1,000-km driving data. Road sections used in our study were typical two-lane mountainous rural highways with different road conditions and varied facility layouts, where the traffic volume was relatively low, so traffic flow had relatively low impact on driving behavior. The speed limit was 40 km/h or 50 km/h for general road sections, while there were many sharp turns and steep slopes on two-lane mountainous rural highways, where the speed limits were 20 km/h or 30 km/h.

2.2.4. Participants and vehicles

Forty-two drivers (33 males and 9 females) participated in driving experiments with an average age of 32.9 (SD = 7.1, range 23~50). To avoid the influence of novice drivers on the model, all the drivers had over three years of driving experience (mean = 11.3, SD = 5.8, range 3~22). These participants all had normal or corrected-to-normal vision. The vehicle types used in our driving experiments among participants were similar, belonging to the four-door passenger sedan category.

2.2.5. Data collection and extraction

A driving recorder (GARMIN GDR35) was set in the line of the driver's sight. There was always an electromagnetic interference between GPS signal reception and video recording, but this well-designed driving recorder could greatly mitigate this problem and accurately match GPS positions with driving video information in time. Driving position, speed, three-axis acceleration and a video of the driver's eye view were recorded simultaneously. The sampling frequency was 1 Hz, since we hoped our research result could be used for real-time prediction. Videos from driving recorders were preliminarily processed on the basis of what Wang et al. (2000) put forward. A video data processing system was developed based on the HALCON software and MFC libraries to effectively establish a visual road environment model. Fig. 3 demonstrates the video data processing system, which consisted of two parts, including video image processing and data record. In this self-developed system, parameters of visual road environment were calculated automatically in the left interface and then were recorded in the right interface after matching with vehicle kinematic features.

2.3. Random forests algorithm

This study used Random Forests to establish a speeding prediction model and identify key factors that impact speeding. Since a single decision tree cannot perform well in classification, Random Forests are used to improve classification performance. Random Forests are an ensemble learning method for regression and classification which combines bootstrapping with boosting and bagging (Harb et al., 2009). Bootstrapping refers to random sampling with replacement. Boosting is a method that can calculate many models at the same time and these models have the same weight, so the result of each model can be regarded as a vote and the final result for this overall method is the one with majority votes from all models. Bagging is an approach to reduce the variance of prediction and avoiding overfitting (refers to a model error which occurs when the model performs well to a particular set of data but fails to fit additional data or predict future observations reliably) by generating additional data with some repeated observations for training which have the same size as the original data.

Random Forests have many features superior to other methods

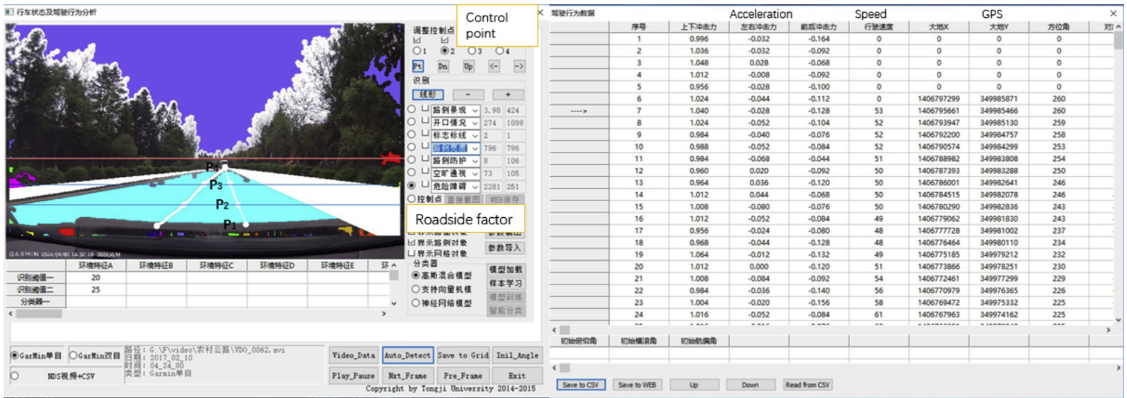


Fig. 3. Self-developed video data processing system.

(Breiman, 2001), for example, 1) it has a high accuracy among current algorithms; 2) it runs faster than other methods and can efficiently handle a large number of variables without dimensionality reduction (i.e., there is no need to decrease the size of input variables); 3) it prevents overfitting by creating random subsets of the features and building smaller trees with these subsets; 4) it provides an internal unbiased estimate of the generalization error as the forest grows; 5) it can evaluate the importance of each variable and mitigate the multicollinearity problem; etc.

In this model, there are 20 independent variables from three categories, such as acceleration, visual curve length, curvature, area of visual roadside, landscapes, barrier types, age, gender, etc., while speeding is chosen as the dependent variable. Random Forests require a limited number of parameters, including two main parameters: the number of trees in the forest (denoted by n_{tree}) and the number of input variables tried at each split (denoted by m_{try}). By using bootstrapping in the tree induction process, only approximately two-thirds of the samples in the training dataset are used. The left-out samples are called "out-of-bag (OOB)" and are used to estimate the prediction error (Khoshgoftaar et al., 2007). The process of Random Forests generally consists of four steps, demonstrated as follows.

Step 1: Build a bootstrapped sample D_i from the original observations D, where $|D_i| = |D|$ and samples are drawn with replacement from D; namely, each bootstrapped sample randomly used two-thirds sample from original dataset, and the rest will be applied to test the model effectiveness.

Step 2: Grow an un-pruned tree C_i using D_i as the training dataset with the standard decision tree algorithm. At each node in the tree, restrict the set of candidate attributes to a randomly selected subset (x_1, x_2, …, x_l), where $l = m_{\text{try}}$.

Step 3: Repeat Steps 1 and 2 for $i = 1, \ldots, n_{\text{tree}}$, to construct a forest of trees C_i, derived from different bootstrap samples. In the n_{tree} different bootstrap samples, they contain different original data, since they randomly select original observations.

Step 4: Classify an observation x by aggregating the votes over all trees in the forest, where $C_i(x)$ denotes the class of x determined by tree C_i. The predicted class of x is the class receiving the most votes.

After the final Random Forests classifier has been built, the estimation of variable importance can be calculated. The calculation is performed by testing how much the OOB error increases when the value of the target variable is permuted while others remain unchanged (Bureau et al., 2005). This process is carried out by the following equations:

$$T_i = \sum_{j=1}^{T} t_{ij} \tag{4}$$

$$\text{mg}(X_i, y_i) = \frac{1}{T_i}\sum_{j=1}^{T} 1(V_j(X_i) = y_i)t_{ij} - \max_{k \neq y_i}\left\{\frac{1}{T_i}\sum_{j=1}^{T} 1(V_j(X_i) = k)t_{ij}\right\} \tag{5}$$

$$\text{IMP(B)} = \frac{1}{N}\sum_{i=1}^{N}\left[\text{mg}(X_i, y_i) - \text{mg}(X_i^{(B)}, y_i)\right] \tag{6}$$

Where: IMP(B) denotes the importance index for variable B; $\text{mg}(X_i, y_i)$ denotes the margin of votes, namely, the difference between the proportion of votes for the true class and the largest proportion of votes among the other classes for a given individual; T_i denotes the number of trees where individual i is OOB; X_i represents the vector of predictor variable, and y_i denotes its true class; t_{ij} represents an indicator taking value 1 when individual i is OOB for tree j and 0 otherwise; $V_j(X_i)$ denotes the vote of tree j; $1(V_j(X_i) = y_i)$ denotes the indicator function taking value 1 when $V_j(X_i) = y_i$ and 0 otherwise; $X_i^{(B)}$ represents the vector of predictor variables with the value of variable B randomly permuted among the OOB individuals; N is the total number of individuals in the sample.

3. Results

3.1. The distribution of input variables

In total, 7540 valid samples were in the training group, obtained from two-lane mountain highways in Shandong, Anhui, Tibet, and Jiangxi Province, while 1724 valid samples derived from road sections in Zhejiang Province are selected as the testing group. There are 20 input variables coded for each sample, derived from three categories, namely visual road environment parameters, vehicle kinematic features, and driver characteristics. Since a Random Forests model can mitigate the multicollinearity problem and efficiently deal with many variables at the same time, there is no need to select and preprocess variables at the beginning stage of analysis. Visual road environment parameters consist of 14 variables extracted from the visual road geometry and visual roadside environment, which describe the features of road design perceived by drivers' vision. Vehicle kinematic features include three variables, representing the vehicle's current state. In addition, to explore how differently the various driver characteristics affect speeding behavior, three variables were used. The detailed definitions of the input variables from 7540 samples, together with their codes and distributes, are presented in Table 1.

3.2. Speeding prediction model

3.2.1. Prediction using a lower speeding criterion

Using a lower level of the speeding criterion (exceeding the posted speed limit), there are 5328 speeding samples and 2212 non-speeding samples in the training group, which are extracted from road sections in

Table 1
Distributions of input variables.

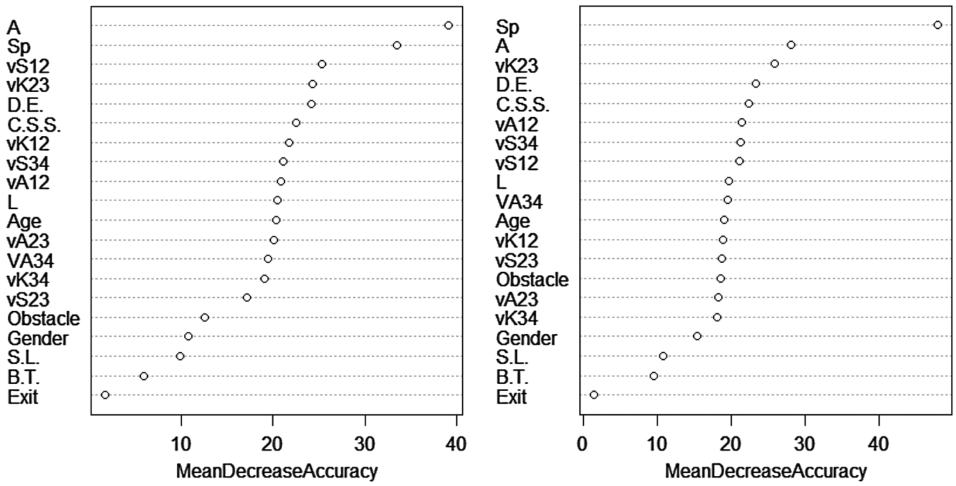
Input variables	Variable code	Description	Min	Max	Mean	S.D.
Visual road environment parameters						
Curve length of visual lane (pixel)	vS_{12}	Curve length in "near scene"	77	997	394	165
	vS_{23}	Curve length in "middle scene"	45	786	203	111
	vS_{34}	Curve length in "far scene"	20	770	111	78
Curve curvature of visual lane	vK_{12}	Curvature in "near scene"	0	0.04	4e-4	2e-3
	vK_{23}	Curvature in "middle scene"	0	0.03	2e-3	4e-3
	vK_{34}	Curvature in "far scene"	0	0.11	3e-3	0.01
Area of visual roadside(pixel^2)	vA_{12}	Roadside area in "near scene"	3300	3e5	5e4	4e5
	vA_{23}	Roadside area in "middle scene"	1200	2e5	2e4	1e4
	vA_{34}	Roadside area in "far scene"	200	7e4	5630	6410
Landscapes	L	Roadside landscapes	trees or other plants (36%) mountains (21%) houses or streets (33%) tunnels (2%) open fields (8%)			
Barrier types	B.T.	Roadside barrier types	no barriers (53%) discontinuous concrete barriers (5%) continuous barriers (42%)			
Dangerous obstacles	Obstacle	Roadside obstacles	with dangerous obstacles (38%) without dangerous obstacles (62%)			
Exits	Exit	Roadside exits	with exits (18%) without exits (82%)			
Posted speed limit (km/h)	S.L.	Speed limit at the predicting position	50 km/h (47%), 40 km/h (34%), 30 km/h (14%), 20 km/h (5%)			
Vehicle kinematic features						
Driving speed (km/h)	Sp	Driving speed at the current position	19	82	54	11
Acceleration (g·m/s^2)	A	Acceleration at the current position	−0.5	0.5	−0.05	0.08
Current speeding status	C.S.S.	Speeding or not at the current position	speeding (66%) non-speeding (34%)			
Driver characteristics						
Age (years old)	Age	Age of drivers	23	50	32.9	7.1
Gender	Gender	Gender of drivers	male (79%) or female (21%)			
Driving experience (years)	D.E.	Years of driving experience	3	22	11.3	5.8

Table 2
Prediction results of Random Forests under two levels of speeding criterion.

Using lower speeding criterion							
	Training set				Testing set		
Predicted	Observed		OOB error rate	OOB accuracy	Observed		Accuracy
	Speeding	Non-speeding			Speeding	Non-speeding	
Speeding	4587	361	14.6%	85.4%	810	124	84.2%
Non-speeding	741	1851			148	642	
Using higher speeding criterion							
	Training set				Testing set		
Predicted	Observed		OOB error rate	OOB accuracy	Observed		Accuracy
	Speeding	Non-speeding			Speeding	Non-speeding	
Speeding	2921	532	13.6%	86.4%	661	135	85.2%
Non-speeding	493	3594			120	808	

four provinces of China, namely, Shandong, Anhui, Tibet, and Jiangxi Province. The data in the testing group are obtained from another province in China, Zhejiang Province, including 958 speeding samples and 766 non-speeding samples. A Random Forests model is built by inputting 20 independent variables, which is implemented using R software. As the number of trees in the forest increases, the OOB error initially goes down before becoming stable after 500 trees. With the growth of the number of input variables tried at each split, the OOB error decreases at first and then increases, which reaches the bottom when $m_{\text{try}} = 4$. Thus, n_{tree} is set to 500 and m_{try} is equal to 4. Table 2 shows the result of Random Forests. The final OOB error rate is 14.6%, i.e., the OOB accuracy is 85.4%. In the testing set, the overall accuracy reaches 84.2% (95% CI: 81.1% ˜ 87.3%). It indicates that this model performs well in predicting whether drivers will exceed the speed limit at the future position that is a sighting distance away.

Variable importance is demonstrated in Fig. 4(a) and Table 3 which tests how worse the model performs without each variable when a lower speeding criterion (over the speed limit) is used. Current acceleration and driving speed rank top two with importance scores larger than 30, indicating that these two variables are the most important in the speeding prediction model. The followings are visual environment parameters such as visual curve length in “near scene” and visual curve



(a) Using a lower speeding criterion (over the speed limit)   (b) Using a higher speeding criterion (10% over the speed limit)

Fig. 4. Variable importance of Random Forests (variable abbreviations are shown in Tables 1 and 3).
(a) Using a lower speeding criterion (over the speed limit) (b) Using a higher speeding criterion (10% over the speed limit).

Table 3
Variable importance rankings and scores in Random Forests.

Variables (variable code)		lower speeding criterion (over the speed limit)		higher speeding criterion (10% over the speed limit)	
		Ranking	Score	Ranking	Score
Visual road environment parameters					
visual curve length	vS_{12}	3	25.3	8	21.1
	vS_{23}	15	17.2	13	18.7
	vS_{34}	8	21.2	7	21.2
visual curvature	vK_{12}	7	21.8	12	18.9
	vK_{23}	4	24.3	3	25.9
	vK_{34}	14	19.1	16	18.1
visual roadside area	vA_{12}	9	20.9	6	21.3
	vA_{23}	12	20.1	15	18.3
	vA_{34}	13	19.5	10	19.5
landscapes (L)		10	20.5	9	19.6
barrier types (B.T.)		19	5.9	19	9.5
dangerous obstacles (Obstacle)		16	11.3	14	18.5
exits (Exit)		20	1.8	20	1.4
posted speed limit (S.L.)		18	9.9	18	10.7
Vehicle kinematic features					
driving speed (Sp)		2	33.5	1	47.9
acceleration (A)		1	39.1	2	28.1
current speeding status (C.S.S.)		6	22.5	5	22.3
Driver characteristics					
age (Age)		11	20.3	11	19.1
gender (Gender)		17	10.8	17	15.4
driving year (D.E.)		5	24.2	4	23.4

curvature in “middle scene”. In general, shape parameters of visual lane and area of visual roadside in three feature regions have a large impact on speeding behavior, since their importance values are approximately larger than 20. Drivers’ age and driving experiences also significantly affect the performance of speeding prediction with importance values between 20 and 30. Additionally, different roadside landscapes may lead to distinct speeding behavior. The low-ranking variables are exits, barrier types, the posted speed limit, and their importance values are less than 10, which means that they have less effects on the speeding behavior.

3.2.2. Prediction using a higher speeding criterion

When a higher level (10% above the posted speed limit) is chosen as the speeding criterion, 3414 speeding samples and 4126 non-speeding samples are obtained in the training group from four provinces of China. In the testing group, there are 781 speeding samples and 943 non-speeding samples, derived from another province in China. To predict drivers’ speeding behavior at the future position that is a sighting distance away, 20 input variables obtained at the current position are used to establish a Random Forests model. According to the stable and minimum value of the OOB error rate, n_{tree} and m_{try} are set to 300 and 4 respectively. The prediction result of Random Forests is shown in Table 2. In the training group, the OOB error rate 13.6%, namely, the OOB accuracy is 86.4%. In addition, the overall testing accuracy is 85.2% (95% CI: 81.6% ˜ 88.3%). This Random Forests model shows high accuracy in speeding prediction.

Fig. 4(b) and Table 3 illustrate the variable importance which represents the statistical prioritization of variables regarding their contribution to the predictive model with using a higher speeding criterion (10% over the speed limit). Similar to the result of the lower speeding criterion, current acceleration and driving speed are still the top two critical factors, but the importance value for driving speed (equal to 47.9) is much larger than any other factors. Years of driving experience and current speeding status are also significant in the prediction, rank fourth and fifth respectively. As for visual environment parameters, all the shape parameters of the visual lane (i.e., visual curve length and curvature) and area of visual roadside in “near scene”, “middle scene”, and “far scene” are influential features, with their importance values around 20. Roadside landscapes also show a high impact on speeding prediction. However, exits, barrier types, the posted speed limit still stay at the bottom of the importance score ranking and their importance values are less than 11, indicating that these variables did not significantly affect the performance of speeding prediction.

3.2.3. Impacting factors analysis by logistic regression models

Logistic regression models were used to further investigate the effects of independent variables on speeding behavior, as supplements to Random Forests. Logistic regression is one of the most commonly used method in analyzing categorical dependent variables and exploring impacting factors. Thus, the introduction of logistic regression modeling is not provided in this paper for brevity. For more information on logistic regression, please see Agresti (2018). Similar to Random

Table 4
Results of logistic regression under a lower speeding criterion (over the speed limit).

Parameters	Estimate	Std. Error	Z value	Pr (> |z|)
Visual curve length				
vS_{12} (near scene)	0.029	1.2e-3	2.528	0.011
vS_{34} (far scene)	0.019	1.1e-3	2.013	0.036
Visual curve curvature				
vK_{12} (near scene)	−10.86	4.059	−2.694	0.007
vK_{23} (middle scene)	18.47	8.435	3.111	0.002
Visual roadside area				
vA_{12} (near scene)	6.36e-6	5.0e-7	4.178	< 0.001
vA_{23} (middle scene)	6.62e-6	1.3e-7	3.429	< 0.001
vA_{34} (far scene)	8.67e-6	2.1e-7	3.711	< 0.001
Landscape				
open fields*	0			
trees or other plants	−0.117	0.054	−2.181	0.029
tunnels	−0.163	0.063	−1.956	0.040
mountains	−0.328	0.027	−4.156	< 0.001
houses or streets	−2.408	0.733	−6.718	< 0.001
Barrier				
no barriers*	0			
discontinuous concrete barriers	0.035	0.385	0.921	0.357
continuous barriers	−0.288	0.135	−3.153	0.002
Posted speed limit	−0.132	0.025	−5.202	< 0.001
Driving speed	0.100	0.013	7.771	< 0.001
Acceleration	14.63	1.104	13.25	< 0.001
Current speeding status				
non-speeding*	0			
speeding	0.530	0.223	2.380	0.017
Age	−0.229	0.088	−2.603	0.009
Driving year	−0.134	0.075	−4.253	< 0.001

Null deviance: 9702.7 on 7539 degrees of freedom.
Residual deviance: 6516.0 on 7515 degrees of freedom.
AIC: 6566.
Overall prediction accuracy for the testing set: 71.4%.
Note: * denotes reference group for categorical variables.

Table 5
Results of logistic regression under a higher speeding criterion (10% over the speed limit).

Parameters	Estimate	Std. Error	Z value	Pr (> |z|)
Visual curve length				
vS_{12} (near scene)	0.011	1.2e-3	3.464	< 0.001
vS_{34} (far scene)	0.018	1.2e-3	3.537	< 0.001
Visual curve curvature				
vK_{12} (near scene)	−11.50	4.893	−2.715	0.006
vK_{23} (middle scene)	26.26	9.343	2.351	0.019
Visual roadside area				
vA_{12} (near scene)	4.14e-6	5.3e-7	2.635	0.008
vA_{34} (far scene)	3.26e-6	2.8e-7	2.685	0.007
Landscape				
open fields*	0			
trees or other plants	−0.116	0.021	−2.371	0.018
tunnels	−0.241	0.042	−5.041	< 0.001
mountains	−0.346	0.021	−3.894	< 0.001
houses or streets	−0.847	0.078	−4.831	< 0.001
Barrier				
no barriers*	0			
discontinuous concrete barriers	−0.345	0.355	−0.972	0.331
continuous barriers	−0.681	0.176	−3.867	< 0.001
Dangerous obstacles				
without dangerous obstacles*	0			
with dangerous obstacles	−0.307	0.188	−2.511	0.012
Posted speed limit	−0.149	0.023	−6.456	< 0.001
Driving speed	0.173	0.016	11.15	< 0.001
Acceleration	13.47	1.146	11.75	< 0.001
Current speeding status				
non-speeding*	0			
speeding	0.481	0.221	2.180	0.029
Age	−0.215	0.085	−2.511	0.012
Driving year	−0.115	0.072	−4.370	< 0.001

Null deviance: 10,371.0 on 7539 degrees of freedom.
Residual deviance: 6344.1 on 7515 degrees of freedom.
AIC: 6394.1.
Overall prediction accuracy for the testing set: 72.1%.
Note: * denotes reference group for categorical variables.

Forests, logistic regression models are established with 20 input variables while speeding is treated as dependent variables. There are 7540 training data from four provinces in China, while 1724 testing samples are derived from road sections in another province. After eliminating insignificant factors by the stepwise selection, results of the final models under lower and higher speeding criteria are illustrated in Tables 4 and 5 respectively. Comparing with Random Forests, logistic regression models show similar results of the relationship between input variables and speeding. In terms of prediction accuracy, the Random Forests algorithm (over 85%) is superior to logistic regression (around 70%), indicating that Random Forests models can perform well in improving the pre-warning systems. The detailed explanations of impacting factors and comparison between Random Forests and logistic regression are given in the discussion section.

4. Discussion on impacting factors

4.1. Visual road environment parameters

The drivers' visual road environment model quantified road environmental information from two aspects: visual road geometry and roadside environment. Drivers' visual field was divided into three feature regions - "near scene", "middle scene", and "far scene", and in these three scenes drivers have different characteristics of visual perception. Several parameters were extracted to represent the features of visual road environment, including curve length and curvature of the visual lane, area of the visual roadside, dangerous obstacles, landscapes, exits, barrier types and posted traffic signs.

4.1.1. Visual road geometry

According to the variable importance provided by Random Forests, visual curve length in "near scene", curvature in "middle scene" and visual roadside area in "near scene" are top three important visual road geometry factors for speeding behavior. The results of logistic regression models show that under both speeding criteria, visual curve length in "near scene" and "far scene" has a significantly positive effect on speeding. Visual curve curvature in "near scene" negatively affects speeding behavior while curvature in "middle scene" have a positive impact on speeding. Several reasons might be able to explain the relative complex association between visual curve curvature and speeding. When visual curve curvature is very small, the likelihood of speeding is large owe to the less challenging road condition, but when driving on sharp curves, drivers also have a high likelihood of speeding, since they may be less willing or do not have enough time to greatly reduce driving speed to satisfy the posted speed limit on the sharp curve segment. As the visual roadside areas in three feature regions increase, drivers are more like to be involved in speeding behavior. Previous studies also have found some relevant results. Yang et al. (2007) proposed spatial sight distance to quantify the visual information, which was similar to visual curve length in our study, and they found that driving speed went up with the growth of spatial sight distance. As horizontal curvature increases, driving speed had a downward trend (Cafiso and Cerni, 2012) while having a wide roadside led to the rise of speeding (Shinar, 2017).

4.1.2. Visual roadside environment

Among these variables of visual roadside environment, roadside landscapes have the largest importance score calculated by Random Forests. As shown in the logistic regression models under both levels of the speeding criterion, drivers are most likely to generate speeding behavior in an open field, followed by in a landscape with trees or other plants. The likelihood of speeding when driving through tunnels and mountains ranks third and fourth respectively. When passing by houses or streets, drivers are more conservative with a least probability of speeding. Antonson et al. (2009) conducted a driving simulator study and also found that in an open landscape, drivers' experienced speed is much higher than others, and driving speed in a forested landscape was the second largest. When encountering a dangerous obstacle, drivers will be more careful and less likely to drive 10% over the speed limit. On road segments with a lower posted speed limit, drivers have a larger probability of speeding, especially when they suddenly enter an exceedingly speed-limited segment. Although barriers rank the second lowest on the variable importance score, relative to no barriers, continuous barriers can significantly reduce the likelihood of speeding while discontinuous concrete barriers for a warning purpose do not show any significant effects according to the results of logistic regression.

4.2. Vehicle kinematic features

Vehicle kinematic features obtained at current position, including acceleration, driving speed, and current speeding status, play great roles in predicting whether drivers will exceed the speed limit (or drive 10% over the speed limit) at the future position a sighting distance away. Especially, according to the variable importance calculated by Random Forests, acceleration and driving speed ranks top two under both speeding criteria. Besides, logistic regression models show that all these three variables have significantly positive effects on the likelihood of speeding.

4.3. Driver characteristics

Driver characteristics like driving experience and age also have obvious contributions to the speeding prediction model, as demonstrated by Random Forests in Fig. 4. Previous studies demonstrated that drivers' age had a significantly negative impact on speeding behavior (Ogle, 2005). As driving experience increased, drivers were more likely to comply with the speed limit (Warner et al., 2010). The results of logistic regression models in our studies also show the same trends, illustrated in Tables 4 and 5. Under both levels of the speeding criterion, the likelihood of speeding goes down with the increase of their age and driving experience. The association between gender and speeding behavior is not statistically significant in this study, while several prior studies found that male drivers exceeded the speed limit more frequently than female drivers (Ogle, 2005; Warner et al., 2010).

4.4. Other factors

Weather and light conditions are also closely associated with the likelihood of speeding. Owing to reduced visibility caused by bad weather conditions, nighttime, etc., drivers are more likely to choose a lower driving speed (Jägerbrand and Sjöbergh, 2016). Although sometimes this speed is less than the posted speed limit in normal conditions, it may be still not safe under such limited conditions, namely, driving too fast under adverse weather or light conditions should not be defined by exceeding the same speed limit as normal conditions. Thus, many countries have applied lower general speed limits for bad weather and light conditions. For example, in China, when inclement weather conditions like heavy rain, snow, and fog were detected, the speed limits were downgraded in steps of 20 km/h corresponding to different levels of the visibility (Regulation on the implementation of the road traffic safety law of the People's Republic of China, 2019). In France, the speed limits on rural roads decreased from 90 km/h to 80 km/h under severe rain and snow, and visibility less than 50 m in fog changed the speed limit on all types of roads to 50 km/h (Mobility and transport, 2019). As for nighttime, a calculation model for the speed limits at night was establish based on response-braking distance and limited visibility (Han, 2011). In this study, speeding defined by exceeding the speed limit or driving over 10% the speed limit, and this definition can also be applied to adverse weather and light conditions by using the speed limits in reduced conditions.

Besides, the reason why this study used driving recorders located in the line of the sight of the drivers rather than eye trackers could be explained from three aspects: (1) Driver recorders could obtain clearer and more comprehensive road environment information than eye trackers; (2) Driver recorders were more convenient for acquiring large-scale samples such as naturalistic driving data; (3) After an investigation of participants, we found that eye trackers always had a disturbing effect on driving behavior, making the results unconvincing.

5. Conclusion

This study aimed to establish a speeding prediction model which combined the visual road environment with vehicle kinematic features and driver characteristics, hoping to improve the pre-warning systems and mitigate speeding problems. Previous research has demonstrated that the road environment had a significant impact on speeding, but few speeding prediction models considered road environmental design. Models also failed to take into consideration that road environmental information perceived by drivers' eyes had significant differences from actual road design under some circumstances. Thus, this paper provided a more comprehensive visual road environment model by extending our previous drivers' visual lane model.

Random Forests were used to establish the speeding prediction model with naturalistic driving data. The 20 input variables consisted of visual road environment parameters, vehicles kinematic features, and driver characteristics. A speeding criterion was defined with two levels in this study: a lower level (exceeding the posted speed limit) and a higher level (10% above the posted speed limit). Under both levels of the speeding criterion, the speeding prediction model performed well with high OOB accuracy (over 84%). This model could use the value of the variables obtained from the current position to predict drivers' speeding at the future position that was a sighting distance away. The sight distance and driving speeding decided how far ahead of time this model could predict speeding. In our study, on two-lane mountainous rural highways, the average time prior to the occurrence of speeding was 6.3 s (SD = 1.3 range 4 ˜13). This interval was sufficient for a pre-warning system to give a warning that a driver with normal perception-reaction time (around 2.5 s) could respond to (Triggs and Harris, 1982). Reaction time of drivers to road stimuli.). In the current intelligent speeding prediction system (ISPS), speeding can be predicted about 4 s in advance (Zhao et al., 2013). Compared with existing methods used to predict speeding, such as TPB, TCI, ISPS, etc., our model showed its unique advantages in which it took road environmental design from drivers' visual perception into consideration. Moreover, this model could be used for real-time speeding prediction and prevention. Additionally, logistic regression was used as supplements to explore the impacting factors. In terms of prediction accuracy, Random forests showed a huge advantage over logistic regression.

Findings in our study can be used to effectively predict speeding in advance and help to reduce speeding-related traffic accidents. This prediction model can facilitate the improvement of speeding pre-warning systems. Although this model is only applied under free-flow or nearly free-flow conditions in this study, it can also be used in other traffic flow conditions, since visual road environment parameters will change if there are other vehicles in front of the host vehicle. Besides, the visual road environment model has quantified the road information perceived by drivers' eyes, so it can provide automatic driving vehicles with the road information from drivers' visual perception. To make this speeding prediction model more reliable, more factors of driver personality need to be considered in follow-up research. In the future, with the continuous optimization of visual roadside information, a more elaborate visual road environment model will be presented.

Acknowledgements

The authors acknowledge the support from National Science & Technology Support Program of China (2014BAG01B06) and National Science Foundation of China (51238008).

References

Agresti, A., 2018. An Introduction to Categorical Data Analysis. Wiley.
Antić, B., Pešić, D., Vujanić, M., Lipovac, K., 2013. The influence of speed bumps heights to the decrease of the vehicle speed–Belgrade experience. Saf. Sci. 57, 303–312.
Antonson, H., Mårdh, S., Wiklund, M., Blomqvist, G., 2009. Effect of surrounding landscape on driving behaviour: A driving simulator study. J. Environ. Psychol. 29 (4), 493–502.
Antonson, H., Ahlström, C., Mårdh, S., Blomqvist, G., Wiklund, M., 2014. Landscape heritage objects' effect on driving: a combined driving simulator and questionnaire study. Accid. Anal. Prev. 62, 168–177.
Australian Transport Council, 2011. National Road Safety Strategy 2011–2020. Australian Transport Council, Canberra.
Bidulka, S., Sayed, T., Hassan, Y., 2002. Influence of vertical alignment on horizontal curve perception: phase I: examining the hypothesis. Transp. Res. Rec.: J. Transp. Res. Board 1796, 12–23.
Brandenburg, S., Drewitz, U., Krems, J., Petzold, T., Henning, M., 2010. An integrative model for the prediction of overspeeding. European Conference on Human Centred Design for Intelligent Transport Systems.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Broughton, P.S., Fuller, R., Stradling, S., Gormley, M., Kinnear, N., O'dolan, C., Hannigan, B., 2009. Conditions for speeding behaviour: a comparison of car drivers and powered two wheeled riders. Transp. Res. Part F: Traff. Psychol. Behav. 12 (5), 417–427.
Bureau, A., Dupuis, J., Falls, K., Lunetta, K.L., Hayward, B., Keith, T.P., Van Eerdewegh, P., 2005. Identifying SNPs predictive of phenotype using random forests. Genet. Epidemiol. 28 (2), 171–182.
Cafiso, S., Cerni, G., 2012. New approach to defining continuous speed profile models for two-lane rural roads. Transp. Res. Rec.: J. Transp. Res. Board 2309, 157–167.
Cheng, Y., Gao, L., Zhao, Y., Du, F., 2016. Drivers' visual characteristics when merging onto or exiting an urban expressway. PLoS One 11 (9), e0162298.
Chevalier, A., Coxon, K., Chevalier, A.J., Wall, J., Brown, J., Clarke, E., Keay, L., 2016. Exploration of older drivers' speeding behaviour. Transp. Res. Part F Traffic Psychol. Behav. 42, 532–543.
Conner, M., Lawton, R., Parker, D., Chorlton, K., Manstead, A.S., Stradling, S., 2007. Application of the theory of planned behaviour to the prediction of objectively assessed breaking of posted speed limits. Br. J. Psychol. 98 (3), 429–453.
Cristea, M., Paran, F., Delhomme, P., 2013. Extending the theory of planned behavior: the role of behavioral options and additional factors in predicting speed behavior. Transp. Res. Part F Traffic Psychol. Behav. 21, 122–132.
Elvik, R., Vaa, T., Hoye, A., Sorensen, M. (Eds.), 2009. The Handbook of Road Safety Measures. Emerald Group Publishing.
Friedman, S., 2006. Is counter-productive policy creating serious consequences? The case of highway maintenance. Syst. Dyn. Rev. 22 (4), 371–394.
Jägerbrand, A.K., Sjöbergh, J., 2016. Effects of weather conditions, light conditions, and road lighting on vehicle speed. SpringerPlus 5 (1), 505.
Han, J., 2011. Speed Limit on Freeway at Night. Harbin Institute of Technology University.
Harb, R., Yan, X., Radwan, E., Su, X., 2009. Exploring precrash maneuvers using classification trees and random forests. Accid. Anal. Prev. 41 (1), 98–107.
Hassan, Y., Easa, S.M., 2003. Effect of vertical alignment on driver perception of horizontal curves. J. Transp. Eng. 129 (4), 399–407.
Hassan, Y., Sarhan, M., 2012. Operational effects of drivers' misperception of horizontal curvature. J. Transp. Eng. 138 (11), 1314–1320.
Khoshgoftaar, T.M., Golawala, M., Van Hulse, J., 2007. An empirical study of learning from imbalanced data using random forest. ICTAI 2007. 19th IEEE International Conference on Tools with Artificial Intelligence, 2007, vol. 2, 310–317.
Mannering, F., 2009. An empirical analysis of driver perceptions of the relationship between speed limits and safety. Transp. Res. Part F Traffic Psychol. Behav. 12 (2), 99–106.
Matírnez, A., Mántaras, D.A., Luque, P., 2013. Reducing posted speed and perceptual countermeasures to improve safety in road stretches with a high concentration of accidents. Saf. Sci. 60, 160–168.
Mehmood, A., 2009. Determinants of speeding behavior of drivers in Al Ain (United Arab Emirates). J. Transp. Eng. 135 (10), 721–729.
Mobility and transport, 2019. Current Speed Limit Policies. Retrieved January 31, 2019, from. https://ec.europa.eu/transport/road_safety/specialist/knowledge/speed.
NHTSA, 2018. 2016 Speeding Traffic Safety Fact Sheet. National Highway Traffic Safety Administration, Washington, DC.
Ogle, J.H., 2005. Quantitative Assessment of Driver Speeding Behavior Using Instrumented Vehicles. Diss. Georgia Institute of Technology.
Peters, B., Nilsson, L., 2007. Modelling the driver in control. Modelling Driver Behaviour in Automotive Environments. Springer, London, pp. 85–104.
Regulation on the implementation of the road traffic safety law of the People's Republic of China, 2019. Chapter IV Provisions on Road Passage. Retrieved January 31, 2019, from. http://en.pkulaw.cn/display.aspx?cgid=52565&lib=law.
Richter, E.D., Berman, T., Friedman, L., Ben-David, G., 2006. Speed, road injury, and public health. Annu. Rev. Public Health 27, 125–152.
Shinar, D., 2017. Traffic Safety and Human Behavior. Emerald Publishing Limited.
Tay, R.S., Champness, P.G., Watson, B.C., 2003. Personality and speeding: some policy implications. IATSS Res. 27 (1), 68–74.
Transportation Officials, 2011. A Policy on Geometric Design of Highways and Streets, 2011. AASHTO.
Triggs, T.J., Harris, W.G., 1982. Reaction Time of Drivers to Road Stimuli.
Viallon, V., Laumon, B., 2013. Fractions of fatal crashes attributable to speeding: evolution for the period 2001–2010 in France. Accid. Anal. Prev. 52, 250–256.
Wan, H., Du, Z., Yan, Q., Chen, X., 2018. Evaluating the effectiveness of speed reduction markings in highway tunnels. Transport 33 (3), 647–656.
Wang, Y., Shen, D., Teoh, E.K., 2000. Lane detection using spline model. Pattern Recognit. Lett. 21 (8), 677–689.
Warner, H.W., Åberg, L., 2008. The long-term effects of an ISA speed-warning device on drivers' speeding behaviour. Transp. Res. Part F Traffic Psychol. Behav. 11 (2), 96–107.
Warner, H.W., Özkan, T., Lajunen, T., 2010. Can the traffic locus of control (T-LOC) scale be successfully used to predict Swedish drivers' speeding behaviour? Accid. Anal. Prev. 42 (4), 1113–1117.
Watson, B., Watson, A., Siskind, V., Fleiter, J., Soole, D., 2015. Profiling high-range speeding offenders: investigating criminal history, personal characteristics, traffic offences, and crash history. Accid. Anal. Prev. 74, 87–96.
Weller, G., Schlag, B., Friedel, T., Rammin, C., 2008. Behaviourally relevant road categorisation: A step towards self-explaining rural roads. Accid. Anal. Prev. 40 (4), 1581–1588.
Yang, Z., Guo, Z., Du, X., 2007. Visual information-based freeway operating speed prediction model. J. Tongji Univ. (Nat. Sci.) 35 (7), 929–934.
Yu, B., Chen, Y., Wang, R., Dong, Y., 2016. Safety reliability evaluation when vehicles turn right from urban major roads onto minor ones based on driver's visual perception. Accid. Anal. Prev. 95, 487–494.
Yu, B., Chen, Y., Bao, S., Xu, D., 2018. Quantifying drivers' visual perception to analyze accident-prone locations on two-lane mountain highways. Accid. Anal. Prev. 119, 122–130.
Yu, B., Bao, S., Chen, Y., Chen, Y., 2019. Using 3D mobile mapping to evaluate intersection design through drivers' visual perception. IEEE Access 7, 19222–19231.
Yuan, A., Li, C., Zhang, G., 2012. GRSP/MOT speed management pilot project in China. In Australasian Road Safety Research Policing Education Conference, 2012.
Zhao, G., Wu, C., 2013. Effectiveness and acceptance of the intelligent speeding prediction system (ISPS). Accid. Anal. Prev. 52, 19–28.
Zhao, G., Wu, C., Qiao, C., 2013. A mathematical model for the prediction of speeding with its validation. IEEE Trans. Intell. Transp. Syst. 14 (2), 828–836.