WILEY | Hindawi

Research Article

Study on Driver Behavior Pattern in Merging Area under Naturalistic Driving Conditions

Yan Li,[1,2,3] **Han Zhang,**[2,3,4] **Qi Wang** ⓘ **,**[2,3] **Zijian Wang,**[1,3] **and Xinpeng Yao**[1,3]

[1]*Shandong Hi-Speed Construction Management Group Co. Ltd., Jinan 250001, China*
[2]*School of Qilu Transportation, Shandong University, Jinan 250002, China*
[3]*Shandong Provincial Key Laboratory of Intelligent Transportation, Jinan, China*
[4]*Shandong Hi-Speed Group Innovation Institute, Jinan 250001, China*

Correspondence should be addressed to Qi Wang; 202285009024@sdu.edu.cn

To reduce the risk of traffic conflicts in merging area, driver's behavior pattern was analyzed to provide a theoretical basis for traffic control and conflict risk warning. The unmanned aerial vehicle (UAV) was used to collect the videos in two different types of merging zones: freeway interchange and service area. A vehicle tracking detection model based on YOLOv5 (the fifth version of You Only Look Once) and Deep SORT was constructed to extract traffic flow, speed, vehicle type, and driving trajectory. Acceleration/deceleration distribution and vehicle lane-changing behavior were analyzed. The influence of different vehicle models on vehicle speed and lane-changing behavior was summarized. Based on this data, the mean and standard deviation of velocity, acceleration, and variable acceleration were selected as the characteristic variables for driving style clustering. To avoid redundant information between features, principal component dimensionality reduction was performed, and the dimensionality reduction data was used for K-means and K-means++ clustering to obtain three driving styles. The results show that there are obvious differences in the driving behaviors of vehicles in different types of merging areas, and the characteristics of different areas should be fully considered when conducting traffic conflict warnings.

1. Introduction

With the increase of traffic flow, traffic pressure in the highway merging area is heavy. Drivers need to finish complex driving tasks such as acceleration, deceleration, and lane change in merging area. Although the merging area account for a small proportion of the highway, traffic accidents occurred in this area account for a large percentage. For example, during the period from 2014 to 2016, the total length of interchanges in a province in eastern China accounted for about 9.5% of highway in this province but the number of accidents accounted for 34.5% of the total number of highway accidents, and it accounted for 24.95% of the total number of fatalities in the province [1]. In a statistical study in the United States, it also showed that accidents occurred in the highway interchanges accounted for 18% of the total number of accidents and the percentage of fatal accidents is 21.8% [2]. As an important part of the highway, the merging area is an important "valve" for auxiliary vehicle steering. In the merging area, as vehicles continue to merge into the mainline lanes, driving behaviors such as vehicle acceleration, deceleration, and lane changes occur more frequently, which is likely to cause drivers to take improper driving actions and causes traffic accidents. Therefore, it is important to study speed changes, trajectory changes, and other driving behaviors in merging area to identify and resolve traffic conflicts and improve traffic safety.

Due to drivers' personalities, driving styles are diverse, which can be broadly classified as aggressive, steady, and conservative [3]. Early classification of driving styles relied on questionnaire surveys, which were subjective [4–6]. In recent years, with the update of driving simulators and sensors, the evaluation of driving styles through specific

parameter evaluation indicators (e.g., speed, acceleration, etc.) has become more convincing [7]. Many scholars used simulations to obtain data such as time headway, acceleration, and steering wheel angle to analyze driving styles or behavioral characteristics. Shi et al. took the lane-changing behavior on highways as the research object and obtained subjective and objective driving data through questionnaires and driving simulators. The experimental analysis showed that the driving style had a great influence on drivers' lane-changing behaviors [8]. Orfanou et al. studied driving style in traffic congestion and used neural networks to analyze traffic flow parameters. The results showed that the parameters with the greatest influence on driving style are distance and acceleration [9]. Yang et al. constructed a real-time traffic crash risk prediction model considering the temporal effect difference and explored the relationship between dynamic traffic flow characteristics and real-time traffic crash risk under different temporal conditions [10]. Zhao et al. collected pedal and speed data under different traffic flow densities based on a driving simulator. Feature parameters were extracted, and a driving style recognition model was established considering the influence of traffic flow density [11]. Aguilar et al. established a driving style recognition system based on machine learning algorithms, aiming to improve driving safety [12]. In addition, aerial photography technology has been widely used in transportation research and application fields, and some scholars tend to use UAVs to collect real vehicle data [13, 14]. Constantinescu et al. extracted real-time vehicle motion parameters and proposed a driving style evaluation model based on real vehicle driving data [15]. Rodriguez Gonzalez et al. investigated the effects of vehicle speed and lateral and longitudinal acceleration on driving behavior based on data collected from real vehicles and identified driving behaviors with speed, acceleration, and fuel consumption evaluation indexes [16]. Wu et al. utilizes roadside radar to identify and detect targets and classifies them using a two-phase method with a detection accuracy of 89.5% [17]. Li et al. took longitudinal vehicle speed and vehicle speed error as input, and took acceleration and brake pedal opening as output. A driver model that can reflect driving styles was established based on a neural network algorithm [18]. Zhang et al. proposed an improved driver clustering framework by accounting for road types and average speed. The clustering results were compared with those without considering traffic conditions. The improved clustering framework performs better in both intraclass aggregation and interclass separation [19]. Yang et al. constructed a multidimensional multilevel system for traffic crash analysis, this system was capable of accurately and efficiently capturing the mechanics of high-consequence (and possibly low support) highway crashes [20]. Wu and Xu analyzed the effects of five types of driving behaviors on accidents using random forests. These included driver age, road alignment, traffic density, road environment, and whether the driver had both hands on the steering wheel [21]. Other scholars have analyzed lane-changing behavior. Li et al. established a grouped random parameter logit model with heterogeneity in means and variances (GRPMV) and its baseline models, the GRMPV

model has the best model performance and can better capture the unobserved heterogeneity [22]. Li et al. investigated the characteristics of discretionary lane change (LC) duration on freeways, established accelerated failure time (AFT) models with fixed parameters, latent classes, and random parameters [23].

It can be found that the current studies related to driving behavior mostly focus on the analysis of driving behavior characteristics of urban roads [24–36]. There is a lack of studies focusing on the driving behavior for highway merging areas. The ability to acquire microscopic driving behaviors such as speed changes and lane changing is insufficient. Therefore, in this paper, a UAV was used to collect data of highway merging areas, data of vehicle speed and vehicle type was extracted based on YOLOv5 and Deep SORT. The effects of different types of merging areas, different lanes, and different vehicle types on driving behavior were analyzed.

2. Materials and Methods

2.1. Data Acquisition.
This paper adopted the method of fixed-point shooting by UAV and monitoring to obtain real-time data, which can record the whole process of following and changing lanes on the main road and ramp within the merging area of the highway. The use of drones to collect data can avoid interference with the driver and truly reflect the driver's driving behavior.

The videos collected by DJI UAV were divided into two parts. One part was the video data from the actual merging area of the highway, as shown in Figure 1, and the other part was the speed verification video data of non-highway sections. The videos include two types of interchanges and service areas, and the collecting time period for each site was one hour. The shooting height of the UAV was fixed in the range of 15–20 m, and the shooting area of the merging area was about 450 m.

2.2. Vehicle Recognition Detection Model

2.2.1. YOLOv5-Based Vehicle Recognition Model.
Firstly, the dataset is classified and given labels. As shown in Figure 2, the vehicle types in the vehicle recognition dataset are classified into 3 categories: truck, bus, and car. Classification was based on body length, with large vehicles being over 6 m, medium vehicles being 3.5 to 6 m, and small vehicles being under 3.5 m. The dataset was divided into 3 parts: Train, Test, and Valid, where the Train dataset includes 1488 sample images, the Test dataset includes 31 sample images, and the Valid dataset includes 507 sample images.

YOLOv5 integrates some features of YOLOv3-spp and YOLOv4, with the advantages of small size, extreme speed, and high accuracy. Meanwhile, YOLOv5 adds adaptive anchor frame calculation and mosaic data enhancement, the detection effect of small targets is improved and the overall application effect is better. Therefore, the optimal best weight was derived in this project using YOLOv5 model training. The weights are brought into the recognition program, and the recognition effect is shown in Figure 3.

(a)

(b)

FIGURE 1: Data collection site. (a) Qingdao–Lanzhou expressway, Shandong Province, China; (b) Jinan service area, Shandong Province, China.

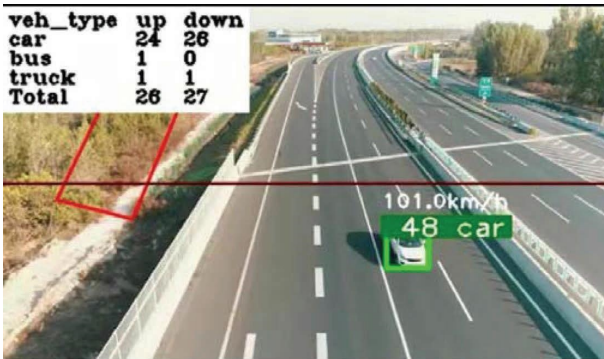FIGURE 2: Datasets for vehicle identification.

FIGURE 3: Effectiveness of video vehicle recognition detection.

2.2.2. Vehicle Tracking Detection Model Based on Deep SORT.

The Deep SORT model can track the relevant attributes of the target of the current frame, including the center point x coordinate, center point y coordinate, target serial no., vehicle type, vehicle pixel width, etc. From the classical formula $V = S/T$, the speed can be obtained by dividing the distance travelled by the time in one direction in a certain period. t_0 is assumed to be the initial time. s_0 is seen as the initial position. t_1 is supposed to the stopping time and

s_1 is regarded as the ending position, then, we have the following equation:

$$v = \frac{\Delta s}{\Delta t} = \frac{(s_1 - s_0)}{(t_1 - t_0)}, \quad (1)$$

d_0 and d_1 are represented as the pixel positions corresponding to the target vehicle at moments t_0 and t_1 in the continuous video sequence, respectively. Then, $(d_1 - d_0)$ is represented as the pixel distance that the target vehicle moves from t_0 to t_1. If $f(x)$ is a single mapping function that represents the mapping relationship between the actual traffic road coordinate system and the image pixel coordinate system, then the formula can be transformed as shown in the following equation:

$$v = \frac{\Delta s}{\Delta t} = \frac{(s_1 - s_0)}{(t_1 - t_0)} = \left(\frac{f(d_1) - f(d_0)}{(t_1 - t_0)} \right). \quad (2)$$

Currently, we already know the transformation relation $f(x)$ between the image coordinate system and the world coordinate system. Then, this can be used to calculate the actual moving distance of the target vehicle in the video. Further combined with the time parameter, the actual travelling speed of the vehicle is calculated.

2.3. Vehicle Identification Detection Model Validation

2.3.1. Validation of Traffic Flow Detection Results. A total of 2 videos of each of the 5 different congestion levels were selected for traffic flow verification, and the length of the videos were all 3 minutes. At first, the traffic flow of each video was recorded by manually counting and then compared with the calculation results of the detection model, and the detection accuracy is shown in Table 1.

As shown in Table 1, when the road traffic is running smoothly, the detection accuracy can reach 100%. However, as the level of traffic congestion increases, there was serious occlusion between vehicles, resulting in a decrease in detection accuracy.

2.3.2. Validation of Vehicle Speed Detection. A total of 5 fixed speed driving videos were used for speed verification, including 10 km/h, 20 km/h, 30 km/h, 40 km/h, and 50 km/h. The data acquisition scenario is shown in Figure 4(a). The vehicle speed was read by extracting the speed value every 2 frames. Since the videos with different speeds have different number of frames, there was a difference in the amount of speed value data available for each video. In addition, the vehicle needs an acceleration process to drive to a fixed speed, so the middle 15 values of each video speed data are selected. The overall average error rate of the vehicle speed was calculated to be 10.03%, as shown in Figure 4(b). The detection error was about 22.67% at the maximum and 0% at the minimum. The speed fluctuates more when the vehicle is farther away from the filming location and stabilises the closer it gets.

2.4. Data Preprocessing. Model and speed data of 2450 vehicles were extracted using the YOLOv5 + Deep SORT model. Due to problems such as video shooting angle and vehicle occlusion, some vehicle speed changes were incompletely recorded, so the raw detection data needed to be sorted and screened. Finally, 2000 vehicles were selected as sample data for driving behavior analysis, of which 1000 were vehicle data from interchanges and 1000 from service areas.

3. Variation of Vehicle Speed in Different Types of Merging Areas

3.1. Traffic Flow and Speed Distribution of Different Lanes. Speed percentage can be more obvious to see the concentration of speed distribution. To better analyze the distribution of vehicle operating speed in the merging area, so as to determine the distribution of speed in each lane, the percentage is used as an indicator for analysis, and the corresponding math expression is as shown in the following equation:

$$P_j = \frac{N_j}{\sum_{i=1}^{13} N_j}. \tag{3}$$

In the equation, P_j represents the percentage of vehicles in a lane whose speed is in the i speed interval. N_j represents the number of vehicles in a lane whose speed lies in the i speed interval.

From vehicle speed data extracted by statistics, it is found that the speed distribution of each lane of the interchange is obviously different from that of each lane in the service area merging area, and most speeds are distributed between 60 km/h and 120 km/h. For this reason, the speed range of 0 to 60 km/h is used as the first speed interval, and then 12 speed intervals are generated from 60 km/h to 120 km/h at an interval of 5 km/h. According to the formula (3), the percentage of vehicle speed of each lane in the corresponding speed interval is counted separately for two different types of merging areas.

Both the interchange and the merging area of the service area have four lanes, with three lanes on the main road and a single lane on the ramp, as shown in Figure 5. The interchange and the merging area of the service area both have four lanes, the main road is three lanes, including the inner lane, middle lane, and outer lane, and the entrance ramp is a single lane.

From the statistical analysis of the 2,000 samples, it can be seen from Figure 6(a) that the middle lanes of the interchange merging area have the highest traffic flow, about 42%. This is followed by the inner lanes, and the outer lanes and ramps have similar traffic flows. It can be seen from Figure 6(b) that the service area merge zone has the highest traffic flow in all lanes, about 49%. This is followed by the middle lane, and the entrance ramp has a slightly higher traffic flow than the outer lane. Based on the statistical data, the percentage of vehicle travelling speed for each lane in the merged area was plotted.

According to the statistical data, the percentage speed of vehicles travelling in each lane of the merging area is plotted.

For the two types of merging areas, speeds from the inside lanes to the entrance ramps all showed a gradual downward trend. As shown in Figure 7, the average speed in the inner lanes of the mainline tends to be concentrated at 105 km/h or more. The average speed in the middle lane was mostly concentrated above 90 km/h. The average speed distribution of the outer lanes of the mainline of the interchange is more dispersed, and the average speed in the service area merging zone was mainly concentrated in 80–100 km/h. In contrast, the biggest difference between the two lies in the speed distribution of the entrance ramps. The average speed of vehicles on the entrance ramps of the interchanges was mostly below 60 km/h, which is obviously lower than the average speed of the entrance ramp area in the service area.

3.2. Distribution of Vehicle Acceleration and Deceleration in Each Lane. As a factor in the change of vehicle travelling speed, acceleration is one of the main parameters reflecting the driving status of the vehicle. It is also a core indicator reflecting driving stability, comfort, safety, and driving behavior analysis, which can objectively reflect the driver's driving behavior pattern. In this section, the average vehicle acceleration was divided into 10 intervals to analyze the acceleration and deceleration of vehicles in each lane.

The acceleration distribution showed a trend of high, middle, and low at both ends, mainly distributed in the $(-2, 2)$ m/s^2 interval. The acceleration and deceleration of vehicles in each lane were basically stable; however, due to lane

TABLE 1: Accuracy of traffic flow detection.

Congestion level	Car		Bus		Truck		Traffic flow accuracy (%)
	Actual	Testing	Actual	Testing	Actual	Testing	
Very smooth	68	68	1	1	0	0	100
	43	43	2	2	2	2	100
Unobstructed	99	99	6	3	0	3	100
	86	86	1	2	2	1	100
Mild congestion	144	143	3	2	3	3	98.67
	112	110	2	1	0	0	96.90
Moderate congestion	125	121	2	1	2	3	96.90
	110	106	1	0	0	1	96.36
Heavy congestion	54	50	2	1	1	1	91.23
	43	38	4	3	0	1	89.36

alignment, driver status, etc., the frequency of sharp acceleration and deceleration of vehicles in the service area merging zone was higher than that in the interchange merging zone. From Figure 8(a), it can be seen that the speed of the middle lane is more stable, and the frequency of acceleration and deceleration in the inner lane is relatively high. The outer lanes are more closely connected to the entrance ramps, and the acceleration and deceleration frequencies of vehicles were similar and higher. Figure 8(b) shows that the middle lane has a more stable speed, the entrance ramp has a higher frequency of acceleration and deceleration, and the frequency of acute acceleration and deceleration is more frequent.

4. Analysis of Vehicle Lane Change Behavior of Different Lanes

Lane-changing behavior requires a combination of the surrounding traffic scenario and driving characteristics. By finding the right time to change lanes, the vehicle reaches its driving destination. Lane-changing behavior is one of the most common behaviors of road vehicles. Especially in complex areas such as highway merging zones, it has a great impact on the road safety and stable operation of vehicles. Therefore, detailed analysis of the characteristics of vehicle lane changes in the merger area is important for improving traffic safety.

To better analyze the frequency of vehicle lane changes in different lanes in the merging area, the lane change rate is used as an analysis index. Lane change rate is defined as the ratio of the number of vehicle lane changes in a certain area to the total number of vehicles in that area, as shown in the following equation:

$$\beta_{lcr} = \frac{t_{lc}}{N_{tv}}, \qquad (4)$$

where β_{lcr} is the vehicle lane change rate, t_{lc} represents the number of vehicle lane changes in an area, and N_{tv} represents the total number of vehicles in a region.

It is stipulated that the behavior of a vehicle entering from one lane to another adjacent lane is a lane change. The merging area is divided into 10 regions, each with a range of about 50 m, as shown in Figure 9. Vehicle lane change data for each lane in different regions is then extracted from it.

4.1. Frequency of Vehicle Lane Change in the Main Road Lane.
To study the impact of merge zone entrance ramp vehicle convergence on main road vehicles, this section summarized the frequency of vehicle lane changes for each lane of the arterial. The results were obtained as shown in Table 2. A total of 115 vehicles changed lanes on the main road in the interchange merging area, and 259 vehicles changed lanes on the main road in the service area merging area.

Since the lane lines between the middle lane and the outer lane in Areas III to VIII are solid lines, lane changing is prohibited for vehicles. So, mainline vehicles change lanes very frequently in Areas I and II to avoid conflicts with the merging ramp traffic and the mainline traffic. When the lane lines of the middle lane and the outer lane will become dashed, the lane change rate between the middle lane and the inner lane in Area VIII increased. And the overall lane change rate of each lane in the merging area of the service area was slightly higher than that of the interchange.

4.2. Vehicle Change Location on the Entrance Ramp.
To summarize the pattern of where vehicles from the ramps merge onto the main roadway, the number of vehicles merging from the ramps in zones IV–VII was counted. As shown in Table 3, the number of converging vehicles in the interchange merging area ramp is 70, and the number of converging vehicles in the merging area of service area is 63.

It can be seen that 20% of the vehicles in the interchange merging area have merged onto the main road within 50 m of the acceleration lane. This is related to the aggressive driving behavior of drivers or the low traffic volume on the mainline. In addition, when the traffic volume of the mainline is too large or the speed increase is slow, it will lead to vehicles not being able to merge into the mainline in time. In this dataset, 10% of vehicles merge onto the main roadway late. The location of vehicles from service area ramps joining the main road is similar to that of interchanges, with the highest number of vehicles joining within 50–100 m, accounting for 38.65%.

Using the data in Tables 2 and 3, it can be clearly seen that the situation of vehicle lane changing in each lane area, which provides sufficient data support for the subsequent vehicle driving law, conflict analysis and early warning, and clarifies the focus research area of vehicle conflict location.
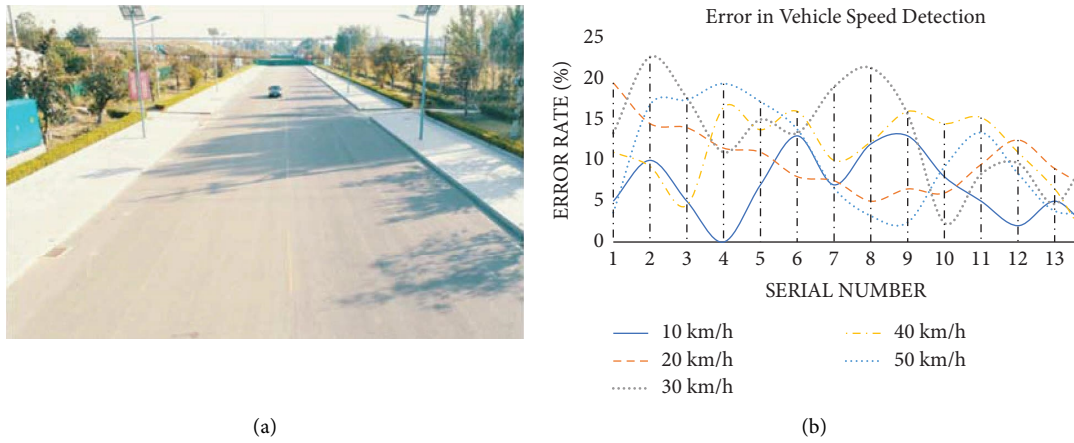
(a)

(b)

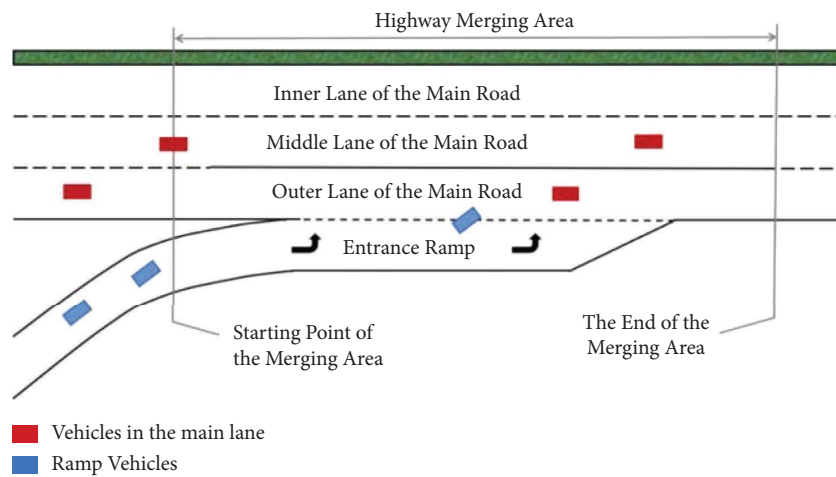FIGURE 4: (a) The speed verification video data. (b) Error in vehicle speed detection.

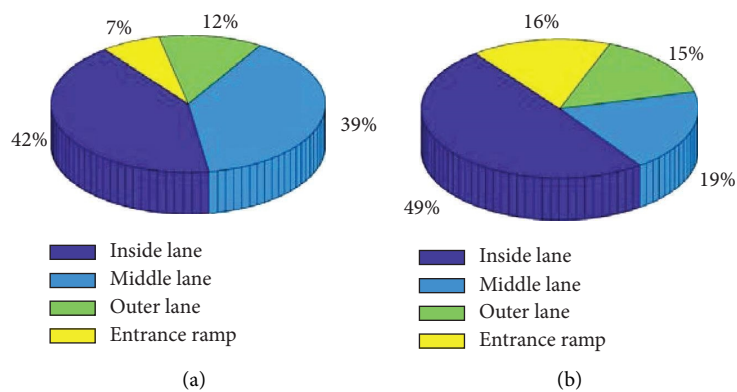FIGURE 5: Highway merging area.

(a)

(b)

FIGURE 6: (a) The traffic volume distribution of each lane in the merging area of the interchange; (b) the traffic volume distribution of each lane in the merging area of the service area. (a) Interchange and (b) Rest area.
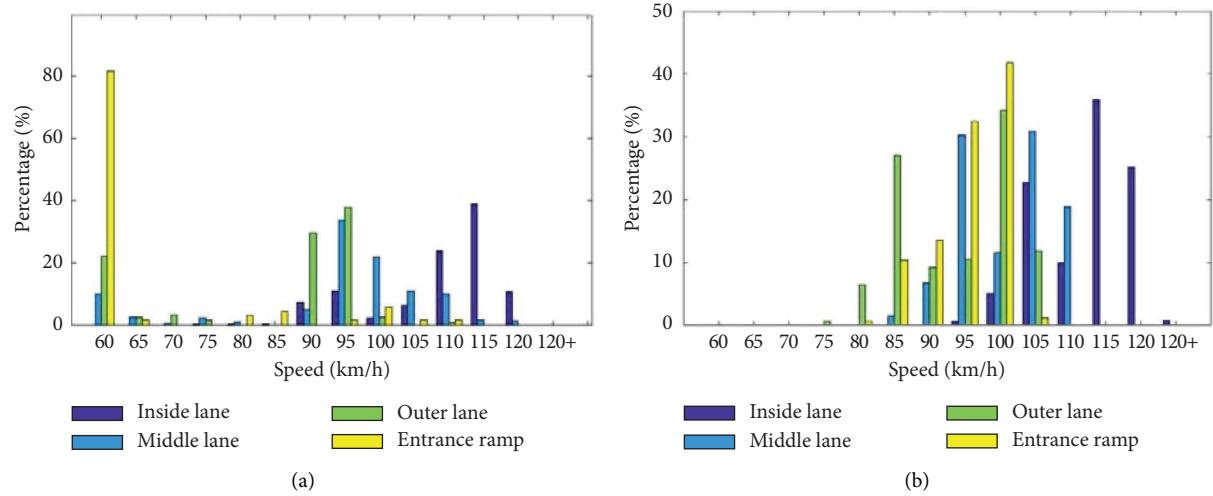
FIGURE 7: (a) The average speed distribution of each lane in the merging area of the interchange; (b) the average speed distribution of each lane in the merging area of the service area.
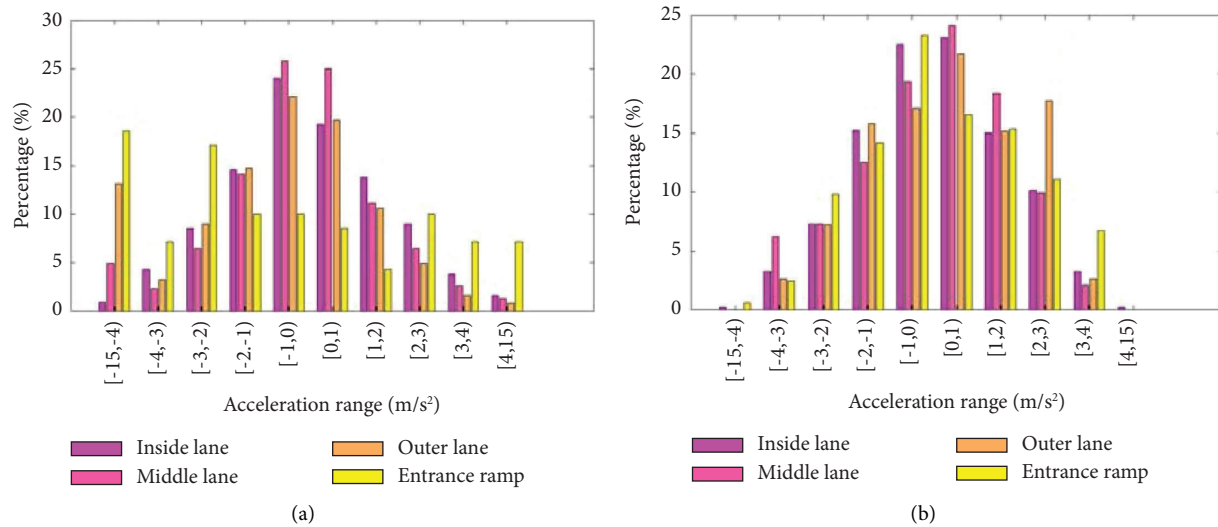

FIGURE 8: (a) The average acceleration distribution of each lane in the merging area of the interchange; (b) the average acceleration distribution of each lane in the merging area of the service area.
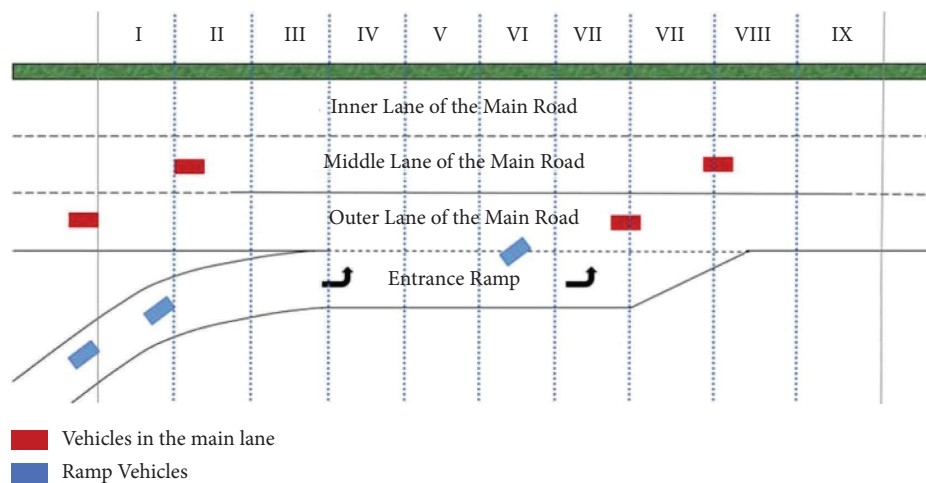

FIGURE 9: Zoning of the highway merging area.

TABLE 2: Statistics on the frequency of vehicle lane change in different areas.

Area		β_I (%)	β_{II} (%)	β_{III} (%)	β_{IV} (%)	β_V (%)	β_{VI} (%)	β_{VII} (%)	β_{VIII} (%)	β_{IX} (%)
Interchange	Middle-inner	7.2	12.2	1.4	0.0	0.0	4.3	9.4	13.7	2.9
	Outer-middle	23.0	18.7	0.0	0.0	0.0	0.0	0.0	0.0	15.1
Service area	Middle-inner	25.6	12.2	3.7	1.2	1.2	0.0	6.1	22.0	32.9
	Outer-middle	20.7	3.7	0.0	0.0	0.0	0.0	0.0	0.0	18.3

TABLE 3: Lane change rates for entrance ramps in each region.

Area		IV	V	VI	VII
Interchange	Range	0–50	50–100	100–150	150–200
	Number of vehicles	14	34	15	7
	Percentage (%)	20.00	48.57	21.43	10.00
Service area	Range	0–50	50–100	100–150	150–200
	Number of vehicles	30	63	41	29
	Percentage (%)	18.40	38.65	25.15	17.79

5. The Effect of Different Vehicle Type on Speed and Lane Change Behavior

There were 899 cars and 101 trucks in the interchanges in the sample data set. There were 816 cars and 184 trucks in the service area. Because of the small number of buses in the sample and their large size, they are categorized among the trucks. Due to the characteristics of different car models, there will be a significant difference in driving speed and lane changing behavior in the merging zone. This section selected special points or sections of the merging zone to analyze the driving patterns of different car models. The merging zones at the collection sites are all single-lane parallel acceleration lanes, and the zoning schematic is shown in Figure 10. It mainly included the front end of ramp confluence, the road segments for acceleration, the confluence point, the segment used for waiting, and the segment used for gradients.

5.1. The Effect of Different Vehicle Type on Speed Variation. To compare the effect of different vehicle type on the speed change, this study summarized the average speed change of vehicles of different vehicle type on the ramp, as shown in Figure 11. The abscissa coordinate 1 in the figure indicated the front end of ramp confluence, 2 indicated the road segments for acceleration, 3 indicated the confluence point, 4 indicated the segment used for waiting, and 5 indicates the segments used for gradients. Each point data is the average of the sample.

As can be seen from Figure 11, in the three mainline lanes, the vehicle speeds in the inner and middle lanes change less. Vehicle speed in the middle lane in the acceleration zone decreased, mainly due to the influence of the outer lane and converging vehicles. On the entrance ramp, both cars and trucks driving past the confluence point decelerate. After a period of acceleration, speeds stabilise at about the same level as the outer lanes.

Compared to interchanges, the service area ramps have different speed trends for different vehicle types. As shown in Figure 12, vehicles in the outer lanes accelerate after

passing the confluence point. Vehicle speed change differently on service area merge zone ramps and interchange merge zone ramps. In this case, vehicle speed on the interchange ramps is decreasing, while vehicle speed in the service area is increasing.

5.2. Effects of Different Vehicle Types on Lane Change Behavior. The merge area merge section was divided into 4 segments of 50 m. The sample consisted of 179 cars and 54 trucks merging from the ramp to the main road. In this study, the trajectories of vehicles merging into the main road from the ramps in the sample data were recorded. The trajectories are shown in Figure 13.

According to the statistics, the lane change positions of cars are concentrated in the interval of 0, 50 m, accounting for 76% of the total number of cars. The lane change positions of trucks are concentrated in the interval of 50–150 m, accounting for 85% of the total number of trucks. Influenced by body length, vehicle acceleration, and other factors, we can see that the overall position of cars changing lanes to the ramp is more forward, from the lane change trajectory diagram, the trajectory of trucks changing lanes is more smooth.

6. Driving Style Clustering

Driving style not only affects the interaction behavior between vehicles in different traffic scenarios, but also has a subtle impact on the decision-making, planning, and control of future autonomous vehicles. Different driving styles may perform differently in different traffic scenarios. Different types of driving styles are also inconsistent in terms of traffic safety hazards. Studying driving styles can to some extent reduce the likelihood of traffic accidents.

6.1. Characteristic Variables of Driving Style Clustering. To cluster driving style, it is first necessary to obtain features that can represent driving style. This study mainly extracted features from the perspective of representing driving safety
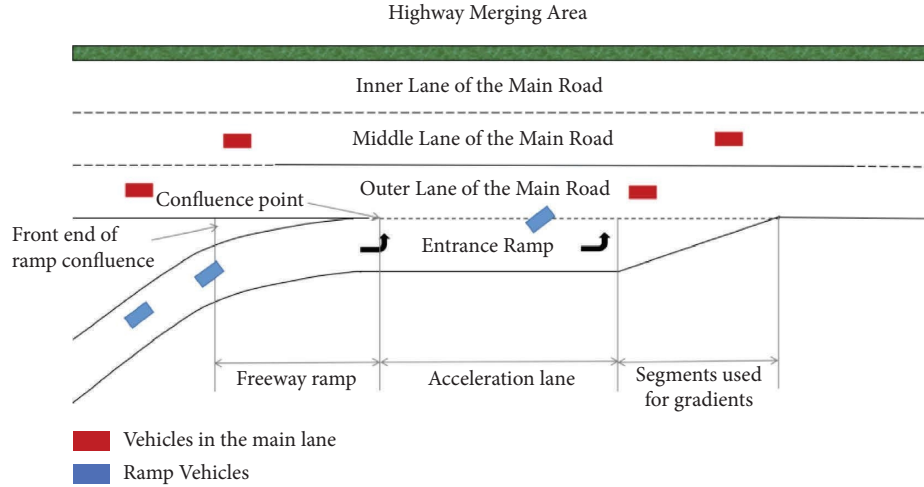
FIGURE 10: Special point and segment distribution in the merging area.
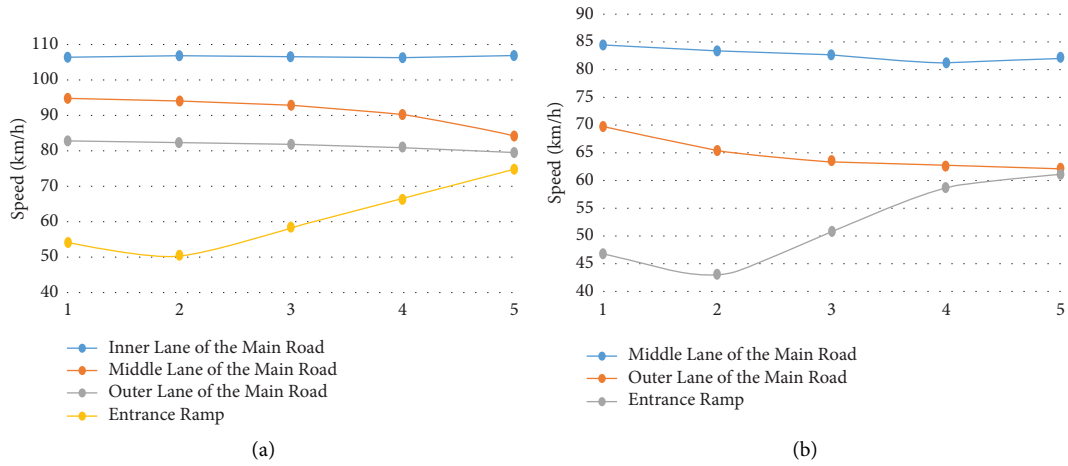

FIGURE 11: Changes in vehicle speed for different types in the interchange merging area. (a) The speed change of the car, which contains 4 lanes; (b) the speed change of the truck. (a) Car. (b) Truck.
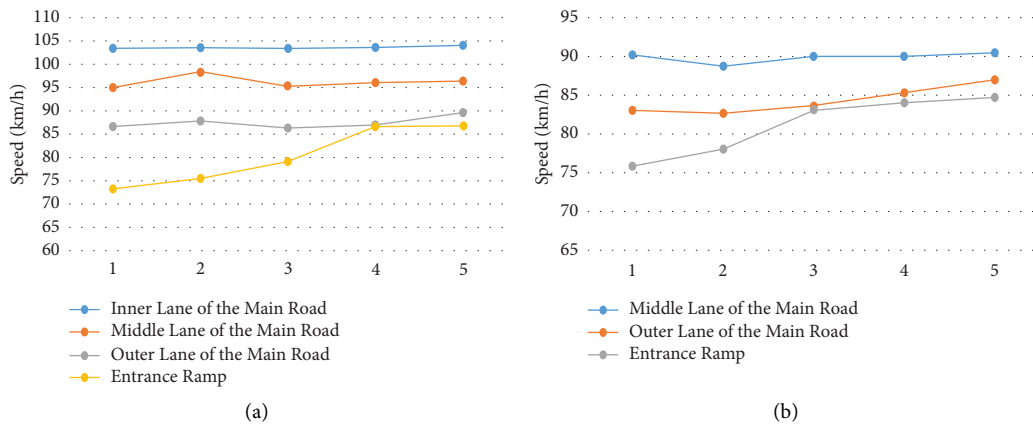

FIGURE 12: Changes in vehicle speed for different types in the interchange merging area. The abscissa is consistent with Figure 11. (a) Car. (b) Truck.

to cluster driving style. We used a sample of 1,000 vehicles in an interchange versus 1,000 vehicles in a service area. The main extracted features include speed, acceleration, and variable acceleration. The commonly used statistical values of these indicators are calculated as features, and a total of 6 features are obtained, as shown in Table 4.
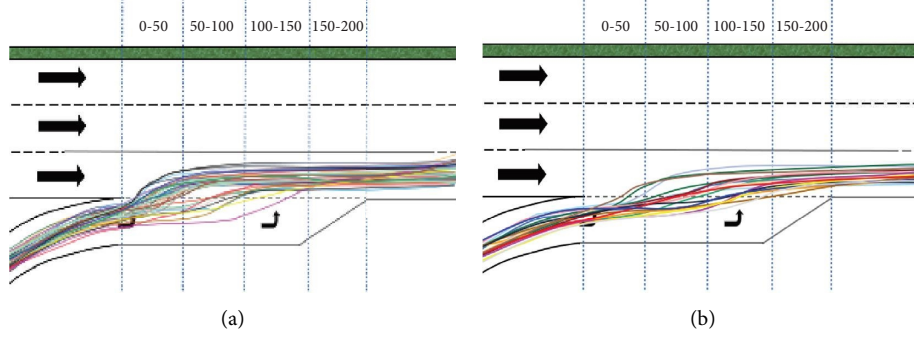
Figure 13: Vehicle lane change trajectory of different vehicle types in the merging area. (a) Car. (b) Truck.

Table 4: Characteristics of driving style clustering.

Symbol	Unit	Meaning
V_{mean}	m/s	Mean of absolute speed values
V_{std}	m/s	Standard deviation of absolute speed value
A_{mean}	m/s^2	Mean value of absolute acceleration
A_{std}	m/s^2	Standard deviation of absolute acceleration value
J_{mean}	m/s^3	Mean value of absolute value of variable acceleration
J_{std}	m/s^3	Standard deviation of absolute value of variable acceleration

6.2. Driving Style Feature Dimensionality Reduction. Given that there is a certain correlation between the above extracted features, and there is redundant information between the features. If all the features are directly brought into the clustering model processing, it will increase the model training cost. It even affects the comprehensive performance of the model and reduces the accuracy. Therefore, in this paper, the proposed features are first downscaled before clustering the driving styles. In view of the excellent performance of PCA in this field, this paper firstly carries out normalisation and PCA dimensionality reduction on the features. Normalisation eliminates the misjudgment of feature importance by the model, which is caused by the nonuniformity of the feature scale. PCA can generate new features through a series of linear transformations, and these new features are principal components. The principal components are generated based on the importance of the original feature information, i.e., the cumulative contribution rate. In practice, it is usually enough to extract the first few principal components whose cumulative contribution rate reaches more than 80%.

Based on Python for feature normalisation and PCA dimensionality reduction, the cumulative contribution rate of principal components can be obtained, as shown in Figure 14. It can be seen that the first three principal components can already represent more than 85% of the original feature information. Therefore, this article selected the first three principal components for subsequent analysis.

6.3. Driving Style Clustering Model. Clustering algorithms can divide a series of data into meaningful clusters according to certain laws. Combined with the PCA dimensionality reduction results above, it can be found that the results obtained only have features without labels, so unsupervised algorithms can be used for subsequent

analysis. The K-means algorithm is the most commonly used unsupervised classification algorithm. The traditional K-means algorithm uses a random method to draw samples from the sample points as the initial center of mass, which has some limitations. The improved K-means++ algorithm based on the traditional K-means algorithm can make the initial centers of mass far away from each other, so as to obtain more reliable results than the random initialisation. In this section, the practical effects of the two algorithms are compared and evaluated using evaluation metrics.

6.4. Cluster Assessment Indicators. Compared with the silhouette coefficient, the Calinski Harabasz Score (CH Score) were calculated quickly. It is better for more concentrated samples with clearer "separation boundaries." The CH score was used to evaluate the clustering effect, generally when the value of the CH score is larger the clustering effect is better. The equation is as follows:

$$\text{CH}(K) = \frac{T_r(B_k)}{T_r(W_k)} * \frac{N - K}{K - 1}, \tag{5}$$

where K is the number of clusters, N is the number of samples, $T_r(B_k)$ and $T_r(W_k)$ are the traces of the intercluster deviation matrix (i.e., covariance matrix) and the traces of the intraclass deviation matrix, respectively, and W_k and B_k are computed as shown in the following equations:

$$W_k = \sum_{q=1}^{K} \sum_{x \in C_q} (x - C_q)(x - C_q)^T, \tag{6}$$

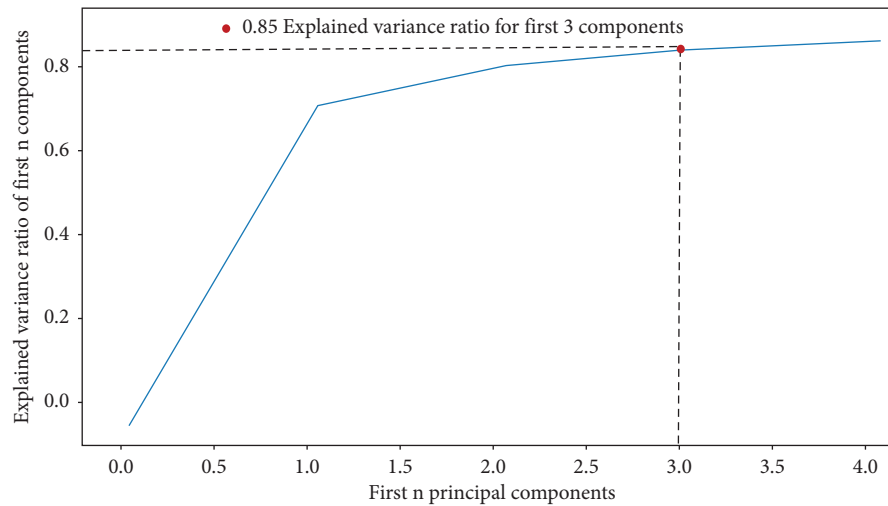$$B_k = \sum_{q=1}^{K} n_q (C_q - c)(C_q - c)^T, \tag{7}$$

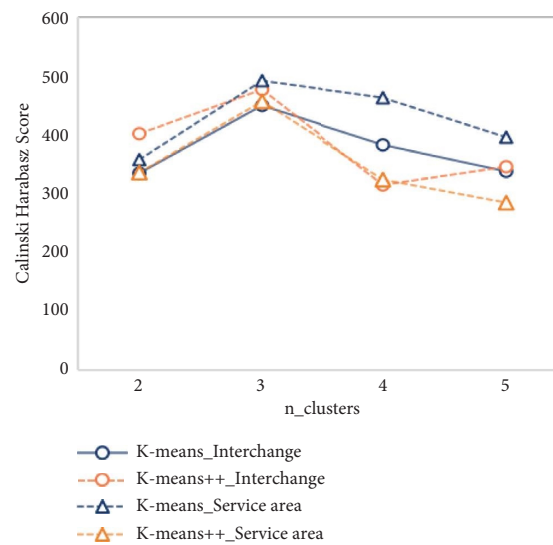FIGURE 14: Schematic diagram of PCA cumulative contribution rate.

FIGURE 15: CH assessment scores for different numbers of clustered clusters.
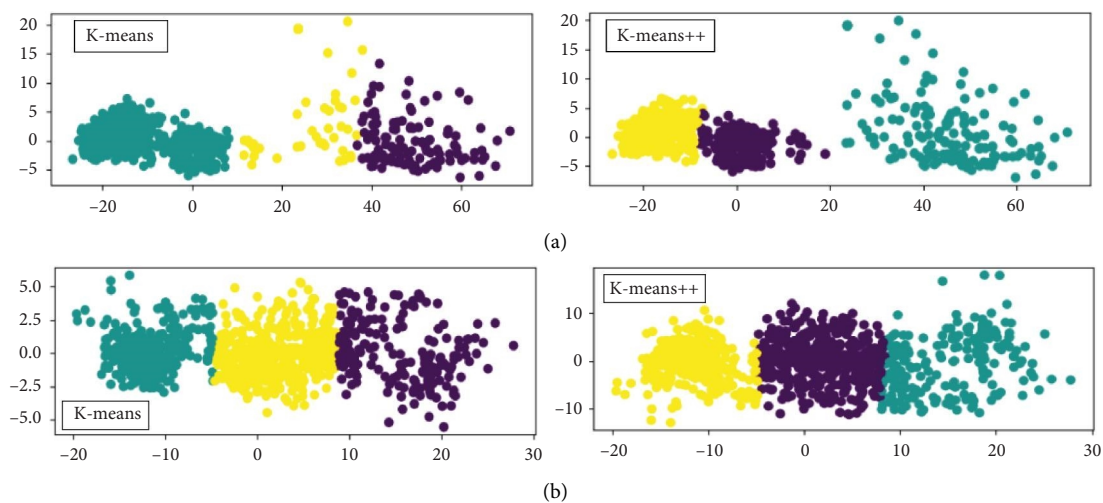

FIGURE 16: Clustering results of vehicle driving styles in the merging area. (a) The clustering effect of interchange merging area; (b) the clustering effect of service area merging area.

where C_q is the sample points in a particular cluster q, c is the center of all datasets, and n_q is the total number of samples in a particular cluster q.

6.5. Clustering Results and Analyses. For both k-means and k-means++ algorithms, CH scores were computed for different numbers of clusters. The results are shown in Figure 15. As can be seen from the CH score, clustering is best when the number of clusters is 3.

On this basis, the dimensionality reduction feature data were clustered. The results after dimensionality reduction were clustered using k-means and k-means++ clustering, as shown in Figure 16, The axes in the figure are the coordinates of the center of mass points of the eigenvalues.

Compared to manually labeled driving styles, the accuracy rate of k-means clustering results is 86.45%. The accuracy rate of k-means++ is 97.76%. It can be seen that there was a significant difference between the driver style clustering between the service area and the interchange and merging area. The driver style classification in the service area confluence area was more balanced, and the driver style of the interchange merging area was more aggressive.

7. Conclusion

This paper analyzed the driving behavior patterns of vehicles in two types of merging zones. The results are summarized as follows:

(1) Compared to interchange merging areas, vehicle speed distribution was more decentralized across the lanes in the service area. In interchange merging area, about 80% of vehicles were travelling at speeds of ≤60 km/h. In the service area merging area, the average speeds of the vehicles in each lane were concentrated in the area of (85, 120) km/h. The speed gap was mostly in the range of 20 km/h within the same lane.

(2) For two types of merging areas, the acceleration distribution of vehicles showed a trend of high in the middle and low at both ends. Acceleration was mainly distributed within $-2 \sim 2$ m/s^2. Vehicles in the interchange merge area accelerate and decelerate sharply more frequently.

(3) Vehicle lane change rates were higher in the service area merge area. It exceeded the interchange lane change rate by 21.7%. The average speed of large vehicles in each lane is lower than that of small vehicles, by about 10%. When conducting vehicle conflict warning in the merging area, the monitoring of the service area should be strengthened to further clarify the speed limits and lane change behavioral specifications for different vehicle types near the service area.

(4) When small vehicles merge into the main road from the ramp, most of them chose to merge in the range of 0–50 m. However, most of the large vehicles chose to merge in the range of 50–150 m. In the next study

of vehicle conflict locations, the focus could be on enhancing the characterization of vehicle driving in these two areas.

(5) There were clear differences in the style of drivers in different types of merging areas. Compared to the service area, drivers in the interchange merge area have more aggressive driving styles.

Compared with existing studies, this study considers a variety of scenarios and describes more specifically the driving patterns of vehicles in the merging area of interchanges and service area ramps. The behavioral patterns of drivers in different merging zones can provide a theoretical basis for planning highway merging zones, such as speed limit setting, road structure, and traffic conflict warning. Further, corresponding management methods are proposed for different types of merging zones or different car models. It is of great significance to improve road traffic safety. In this study, the extracted feature data types were relatively single due to the limitation of equipment conditions. To improve the accuracy of the driving style clustering model. The next step will take into account the physiological and psychological characteristics of drivers. Combined with radar point cloud data of vehicles on real roads, the driver's behavioral patterns will be analyzed in depth.

Data Availability

Data were collected and obtained for the article authors. If required, data can be provided by contacting the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Hu, L. He, and R. Wang, "Review of safety evaluation of freeway interchange," *China Journal of Highway and Transport*, vol. 33, no. 7, pp. 17–28, 2020.

[2] D. J. Torbic, D. W. Harwood, D. K. Gilmore, K. R. Richard, and J. G. Bared, "Safety analysis of interchanges," *Transportation Research Record*, vol. 2092, no. 1, pp. 39–47, 2009.

[3] J. Wu, Q. Wang, L. Zhouyuan, and T. Yuan, "Review of risk conflict identification and early warning for interchange," *Journal of Shandong University*, vol. 52, no. 6, pp. 1–13, 2022.

[4] J. Reason, A. Manstead, S. Stradling, J. Baxter, and K. Campbell, "Errors and violations on the roads:a real distinction," *Ergonomics*, vol. 33, no. 10-11, pp. 1315–1332, 1990.

[5] Y. Jiao, X. Wang, D. Hurwitz, G. Hu, X. Xiaoyan, and X. Zhao, "Revision of the driver behavior questionnaire for Chinese drivers' aberrant driving behaviors using naturalistic driving data," *Accident Analysis and Prevention*, vol. 187, 2023.

[6] B. Hussain, T. Miwa, H. Sato, and T. Morikawa, "Subjective evaluations of self and others' driving behaviors: a comparative study involving data from drivers in Japan, China, and Vietnam," *Journal of Safety Research*, vol. 84, pp. 316–329, 2023.

[7] S. Helman and N. Reed, "Validation of the driver behaviour questionnaire using behavioural data from an instrumented vehicle and high-fidelity driving simulator," *Accident Analysis and Prevention*, vol. 75, pp. 245–251, 2015.

[8] J. Shi and M. Liu, "Impact of driving style on freeway lane changing behaviors:A driving simulator-based study," *Journal of Southeast University*, vol. 47, no. 5, pp. 1037–1041, 2017.

[9] F. P. Orfanou, E. I. Vlahogianni, and M. G. Karlaftis, "Associating driving behavior with hysteretic phenomena of freeway traffic flow," *IFAC Proceedings Volumes*, vol. 45, no. 24, pp. 209–214, 2012.

[10] Y. Yang, Y. Yin, Y. Wang, R. Meng, and Z. Yuan, "Modeling of freeway real-time traffic crash risk based on dynamic traffic flow considering temporal effect difference," *Journal of Transportation Engineering Part A Systems*, vol. 149, no. 7, Article ID 04023063, 2023.

[11] H. Zhao, H. Liu, M.-M. Qiu, and L. Cao, "Research on the driving style identification method considering the influence of traffic density," *Automotive Engineering*, vol. 42, no. 12, pp. 1718–1727, 2020.

[12] J. Aguilar, D. Chavez, and J. Cordero, "A dynamic recognition approach of emotional States for car drivers," *International Conference on Technologies and Innovation*, vol. 658, 2016.

[13] Y. Chen, D. Wenrui, L. Hongguang, W. Meng, and W. Xu, "Vehicle detection in UAV image based on video interframe motion estimation," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 46, no. 3, pp. 634–642, 2020.

[14] Y. Zhang, Z. Zuo, X. Xu et al., "Road damage detection using UAV images based on multi-level attention mechanism," *Automation in Construction*, vol. 144, 2022.

[15] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, "Driving style analysis using data mining techniques," *International Journal of Computers, Communications ad Control*, vol. 5, no. 5, pp. 654–663, 2010.

[16] A. B. Rodriguez Gonzalez, M. R. Wilby, J. J. Vinagre Diaz, and C. Sánchez Ávila, "Modeling and detecting aggressiveness from driving signals," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1419–1428, 2014.

[17] J. Wu, H. Xu, R. Yue, Z. Tian, Y. Tian, and Y. Tian, "An automatic skateboarder detection method with roadside Li-DAR data," *Journal of Transportation Safety and Security*, vol. 13, no. 3, pp. 298–317, 2021.

[18] X. Li, J. Hu, H. Jiang, and M. Wuqiang, "Establishing style-oriented driver models by imitating human driving behaviors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 1–9, 2015.

[19] J. Zhang, H. Lu, and J. Sun, "Improved driver clustering framework by considering the variability of driving behaviors across traffic operation conditions," *Journal of Transportation Engineering, Part A: Systems*, vol. 148, no. 7, 2022.

[20] Y. Yang, N. Tian, Y. Wang, and Z. Yuan, "A parallel FP-growth mining algorithm with load balancing constraints for traffic crash data," *International Journal of Computers, Communications and Control*, vol. 17, no. 4, p. 4806, 2022.

[21] J. Wu and H. Xu, "Driver behavior analysis on rural 2-lane, 2-way highways using SHRP 2 NDS data," *Traffic Injury Prevention*, vol. 19, no. 8, pp. 838–843, 2019.

[22] G. Li, L. Zhao, W. Tang, L. Wu, and J. Ren, "Modeling and analysis of mandatory lane-changing behavior considering heterogeneity in means and variances," *Physica A: Statistical Mechanics and Its Applications*, vol. 622, Article ID 128825, 2023.

[23] G. Li, Z. Yang, Y. Pan, and J. Ma, "Analysing and modelling of discretionary lane change duration considering driver heterogeneity," *Transportation Business: Transport Dynamics*, vol. 11, no. 1, pp. 343–360, 2023.

[24] Y. Yang, B. Yang, Z. Yuan, R. Meng, and Y. Wang, "Modelling and comparing two modes of sharing parking spots at residential area: real-time and fixed-time allocation," *IET Intelligent Transport Systems*, pp. 1–20, 2023.

[25] K. Steinhauser, F. Leist, K. Maier et al., "Effects of emotions on driving behavior," *Transportation Research Part F:Psychology and Behaviour*, vol. 59, p. 156, 2018.

[26] Y. Sun, Y. Bi, Y. Han, D. Xie, and R. Li, "Research on safe driving behavior of transportation vehicles based on vehicle network data mining," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 5, pp. 9–13, 2020.

[27] T. W. Victor, J. L. Harbluk, and J. A. Engström, "Sensitivity of eye-movement measures to in-vehicle task difficulty," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 167–190, 2005.

[28] H. Lin, G. Guo, J. Huang, X. Wu, K. Chen, and D. Zhou, "Review on the impact of drivers' familiarity on their safe traffic driving behavior," *China Journal of Highway and Transport*, vol. 35, no. 6, pp. 240–253, 2022.

[29] M. Danaf, S. H. Hamdar, M. Abou-Zeid, and I. Kaysi, "Comparative assessment of driving behavior at signalized intersections using driving simulators," *Journal of Transportation Safety and Security*, vol. 10, no. 1-2, pp. 124–158, 2018.

[30] E. Cho, S. Park, and C. Oh, "Analysis of intrinsic factors leading to aggressive driving behavior to derive safety policy implications for bus drivers," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2676, no. 11, pp. 664–675, 2022.

[31] L. Xia, P. Li, Z. Su, T. Chen, Z. Deng, and D. Sun, "Longitudinal driving behavior before, during, and after a left-turn movement at signalized intersections: a naturalistic driving study in China," *Sustainability*, vol. 14, no. 18, Article ID 11630, 2022.

[32] R. Niroumand, L. Hajibabai, A. Hajbabaie, and M. Tajalli, "Effects of autonomous driving behavior on intersection performance and safety in the presence of white phase for mixed-autonomy traffic stream," *Transportation Research Record*, vol. 2676, no. 8, pp. 112–130, 2022.

[33] R. Chauhan, A. Dhamaniya, and S. Arkatkar, "Driving behavior at signalized intersections operating under disordered traffic conditions," *Transportation Research Record*, vol. 2675, no. 12, pp. 1356–1378, 2021.

[34] T. Zhang and W. Hong, "RETRACTED: driving behavior based on big data analysis and EEG data analysis," *Journal of Physics: Conference Series*, vol. 2066, no. 1, Article ID 012096, 2021.

[35] H. Singh and A. Kathuria, "Profiling drivers to assess safe and eco-driving behavior-A systematic review of naturalistic driving studies," *Accident Analysis and Prevention*, vol. 161, Article ID 106349, 2021.

[36] L. Wu, Q. Shen, and G. Li, "Identifying risk factors for autos and trucks on highway-railroad grade crossings based on mixed logit model," *International Journal of Environmental Research and Public Health*, vol. 19, no. 22, Article ID 15075, 2022.