

Research Article

Vehicle Path Recognition Approach Based on Incomplete Automatic Vehicle Identification

Meimei He,¹ Xianyu Wu ,² Zhenxi Zhan,² and Jingxue Guo²

¹Zhongzi Huake Traffic Construction Technology Co Ltd, Beijing, China

²Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing, China

Correspondence should be addressed to Xianyu Wu; wuxy@bjtu.edu.cn

Received 15 March 2022; Accepted 6 April 2022; Published 9 May 2022

Academic Editor: Zhihong Yao

Copyright © 2022 Meimei He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Vehicle path recognition is one of the key methods used in urban traffic research, such as traffic flow characteristics analysis. Automatic vehicle identification (AVI) is often used for vehicle path recognition and is suitable for mixed traffic flow with connected automated vehicles (CAVs). However, there still remain issues in overcoming the difficulty of vehicle path identification caused by the discontinuity of AVI data and solving the problem of low precision of AVI application. To model the vehicle path, this paper selects the AVI system of Yicheng Town, Linfen City, Shanxi Province, as a test bed. The travel modes of private cars and taxis are discussed, and the quantified indicators of the model are determined. By combining the analytic hierarchy process (AHP) with the entropy weight method (EWM) to get the weights of the indicators, the path recognition model under incomplete AVI data is proposed. Finally, based on the path recognition model proposed in this paper, case studies are carried out for the private car and taxi path recognition, respectively. The validity of the path identification through practical studies and the effect of the number of missing nodes of AVI equipment on the accuracy of the model are discussed. The results show that the recognition of the travel path using the proposed model is consistent with the actual travel path. The accuracy of the proposed model is more than 60% when the number of missing nodes is less than 7 in total 31 nodes. Considering the decision models for private cars and taxis, respectively, the proposed model provides a method for vehicle path recognition based on incomplete AVI data.

1. Introduction

With the prevalent traffic information and data mining technology, it becomes possible to extract the complete path information of vehicles from the collected data by the automatic vehicle identification (AVI) system in most cities. The AVI data is the most direct reflection of the urban traffic travel law, and on the other hand, it also raises attention from transportation engineers and researchers. With AVI data, the complete path information of both human-driven vehicles (HDVs) and connected automated vehicles (CAVs) could be extracted uniformly, which is the groundwork for the further study of the characteristics of mixed traffic flow with CAVs. However, due to funding limitations, the system is usually incomplete. These AVI devices had to be installed

at some critical places instead of all the places that require the devices, which may result in some blind spots. The missing data then increased the difficulty of obtaining the vehicles' travel paths, whereas there is no current practical model to solve this issue. With this consideration, it is necessary to establish a vehicle path recognition method under a set of incomplete AVI data, which would help to exploit data and benefit both traffic police departments and engineers.

Most of the research on AVI data originated from the computer vision aspect, focusing on how to extract accurate license plate information from an image or video system [1, 2]. With the improvement of image recognition technology and computer computing efficiency, image processing and license plate recognition functions are gradually

added to the AVI system [3]. Ajanthan [4] built a license plate recognition system under low-resolution surveillance video, which is robust for environmental changes such as illumination. Lee [5] developed a license plate recognition method in the traffic monitoring scene and trained the license plate data set through the deep convolution network so that the license plate recognition accuracy could be higher than the existing license plate recognition method.

For the application of AVI data, prevalent studies mainly apply the data of the road network monitoring system to obtain traffic parameters such as road network OD matrix [6, 7], road traffic flow [8, 9], and road traffic time [10]. In China, some scholars used AVI data to estimate the average travel time of road segments [11], collect the traffic flow of road sections [12], and calculate the average driving speed under the signal progression [13, 14]. However, there is a lack of research on vehicle trajectory reconstruction based on AVI data. In recent years, some Chinese scholars used AVI data to recover the vehicle trajectory. By fusing the fix-point detector and signal timing data, Tang [15] invented a vehicle trajectory reconstruction method based on traffic wave theory and traffic simulation theory, which overcomes the interference of roadside vehicles. Xiang [16] invented a fusion of AVI devices and a fixed-point detector trunk trajectory reconstruction method, which can improve the accuracy of trajectory reconstruction. Yu [17] presented a vehicle identification data-based trajectory reconstruction method for signalized-link by constructing a phase-to-phase backtracking framework and using the shockwave theory to reconstruct vehicular trajectory segments involved in each backtracking step. Lin and Yang [18, 19] collected the AVI data and extracted vehicle travel trajectories. He also used the data to analyze the pollutant emission sources and emission intensity. Zhang [20] used AVI data at the signalized intersection to reconstruct the trajectory of the vehicles and extract the delay information.

The above studies are applications of AVI data in vehicle trajectory reconstruction, while these studies did not consider the problem of incomplete AVI data, which leads to a certain deviation between the traffic in the road network and the actual situation. Some shortest path algorithms are used to fill the path between the two missing nodes, but when there are more missing nodes, the vehicle does not exactly follow the shortest path, and it is impossible to depict the specific travel path of the vehicle in the road network. Rao [21] used a particle filter to estimate the probability of a vehicle's trajectory from all possible candidate trajectories based on AVI data. Li [22] used the improved Dijkstra algorithm to search for the shortest path of the first K bars between two points by distance. Mo [23] developed a Bayesian path reconstruction model to replenish the lost information resulting from the recognition error and insufficient coverage rate of the AVI system. Guo [24] extracted road travel time based on AVI data.

Some scholars use multiobjective decision-making methods to identify missing paths. Wang [25], Yang et al. [26], Yang et al. [27], and Yang et al. [28] proposed using multitarget decision-making methods to complete the vehicle's travel path, but they assumed that all vehicles use the

same indicators for multitarget path decision-making, which did not consider the road network of different models of the path selection factors and did not distinguish the weight value of each indicator. Most studies use a single method to determine the index weight in calculating the weight of the indicator. The idea of AHP-EWM combining the calculation of the index combination weight has been partially applied in other fields. But the AHP-EWM study combined with path recognition is relatively scarce.

Based on the background of the above research, this paper considers the different path selection indicators of the two models (taxi and private car) under the background of the incomplete AVI data and establishes a vehicle path recognition model to solve the difficult path recognition problem caused by the discontinuity of vehicle travel records.

2. Problem Description

Vehicle travel path recognition refers to the path selection behavior at the missing point. In the set of feasible paths between the origin and the destination, the path matching the actual vehicle travel path may include multiple decision points. The schematic diagram of the path selection behavior is shown in Figure 1. There are three travel paths to choose between OD: Path 1: link 3-link 8 (O-F-D); Path 2: link 1 - link 4 - link 6 - link 7 (O-A-E-C-D); and Path 3: link 1 - link 2 - link 5 - link 7 (O-A-B-C-D).

When there are more than two missing nodes between travel origin and destination, it is difficult to complete the path directly based on the neighboring intersections in the road network. In previous studies, the shortest path algorithm is mostly used to recognize the path between two missing nodes. However, vehicle travel in the actual road network is jointly influenced by a variety of factors, and different types of vehicles are correlated with different factors. So, the method of using the shortest path algorithm to fill in the missing path cannot be adapted to all vehicle types.

In small cities with incomplete public transit systems, the number of taxis in the road network is gradually increasing with the development of taxi services. The proportion of taxis in the traffic flow is gradually increasing. Private cars travel mainly for commuters, while taxis travel mainly to collect passengers for profit. Different travel purposes determine different travel paths. If the travel paths of taxis in the road network are recognized according to the path recognition model of private cars, the received travel paths often do not match the actual travel paths of taxis. If the taxis in the road network are ignored, and only the travel paths of private cars are recognized, it will cause errors in the traffic volume of the road network.

In summary, it is difficult for the existing methods to recognize all vehicles' paths in the road network. In this paper, we will determine the corresponding path decision indicators for the different travel factors of taxis and private cars. A multiobjective decision model considering different vehicle types is proposed. The missing paths of private cars and taxis in the AVI data are recognized separately using this model.

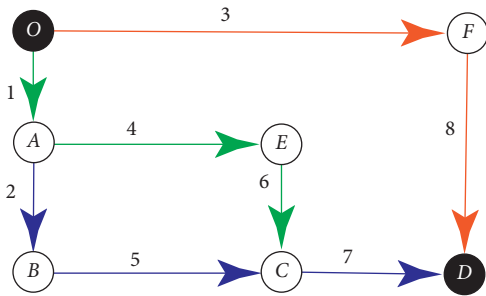


FIGURE 1: Path selection behavior diagram.

3. Methodology

3.1. *Model Assumptions.* The following assumptions are adopted to simplify the analysis:

- (1) The case of vehicles traveling back and forth between ODs is not considered. That is, there is no circular path in the path.
- (2) Vehicles travel only on primary roads, secondary roads, feeder roads, and some other roads with wider widths. Vehicles driving on abnormal segments (e.g., nonmotorized roads) are not considered.

3.2. *Indicator Selection.* To better describe the vehicle path recognition model, it is necessary to select the corresponding indicators for the path recognition model of private cars and taxis. Factors affecting vehicle routing [29] are generally divided into 3 aspects: subjective factors of travelers, such as gender, age, income, and familiarity with the road network; travel purposes, such as commuting and entertainment; environmental characteristics of roads, such as path length, road grade, the number of signalized intersections on the path, the number of turns on the path, the real-time traffic status on the path, the nature of land use, and the effective passage time of the path, etc.

All the data in this paper are based on the AVI data. There is no additional survey to the data. Therefore, the data of the traveler's characteristics cannot be known, which is temporarily not considered. Only factors of road environment characteristics are considered. Different vehicle types have different preferences to choose the travel path. 1000 path data are randomly selected from the extracted complete driving paths as the analysis sample. Combined with vehicle travel data, the Pearson correlation coefficient is used to analyze the path behavior of each indicator in private cars and taxis in SPSS.

The correlation coefficient results of private cars and taxis are shown in Figure 2. The ordinate Y-axis in Figure 2 is the correlation coefficient between each index factor and the probability of path selection. When the value is positive, it means that the indicator is positively correlated with the probability of path selection. The height of the ordinate in the diagram represents the strength of the correlation.

As can be seen from Figure 2, the correlation order of each indicator in the path selection of private cars is as follows:

Consistency between actual and ideal travel time > Traffic operation > Left turn times > Path length > Number of signalized intersections > Road grade.

That means, travelers are prone to choose the ideal path with the shortest path time, no congestion, and relatively smooth driving. At the same time, the correlation between the nature of land use and the probability of private car path selection is not very high, which is almost zero.

The correlation order of each indicator in the path selection of taxis is as shown in Figure 3: Consistency between actual and ideal travel time > Number of signalized intersections > Traffic operation > Land-use > Left turn times > Path length.

Different from private cars, taxis pay more attention to land use when choosing the path, which is closely related to the passenger-taking behavior of taxis. At the same time, the number of signalized intersections is also considered because taxis tend to have smaller delays. The significant correlation level of each indicator can meet the requirements ($p \leq 0.05$).

In summary, the decision-making indicator of the private car and taxi travel path is shown in Tables 1 and 2.

3.3. *Modelling.* The consistency between actual and ideal travel time, traffic operation, left turn times, path length, the number of signalized intersections, and road grade are selected as the decision indicators of the private car impedance model. In the other side, the decision indicators of the taxi path model include the consistency between actual and ideal travel time, the number of signalized intersections, traffic operation, land use, left turn times, and path length. The effective travel time, traffic operation, left turn times, path length, and the number of signal intersections are the common indicators of the two types of vehicles. The road grade is a special indicator for private cars, and the land use is a special indicator for taxis.

Path recognition for vehicles with missing nodes in the AVI data is a process of determining the optimal path from the set of reasonable paths obtained according to the path search algorithm. The vehicles always hope to obtain the shortest path length and the shortest travel time, which is a multiobjective decision-making problem.

Every indicator has a different degree of importance because of its different logic system. Through the no-dimension treatment of different indicators and the determination of the weight value, the multiobjective path decision-making problem is transformed into the problem of finding the optimal value of a single objective. The vehicle selects the overall most satisfactory path in the travel, and the variables (ρ_1, ρ_2) are set:

$\rho_1 = 1$ means that the traveling vehicle is a private car. Otherwise, it should show $\rho_1 = 0$. The variable $\rho_2 = 1$ indicates that the vehicle is a taxi. Otherwise, the value should be 0. The vehicle path recognition model is constructed as shown in formula (1), and formula (2) is the objective function of the model.

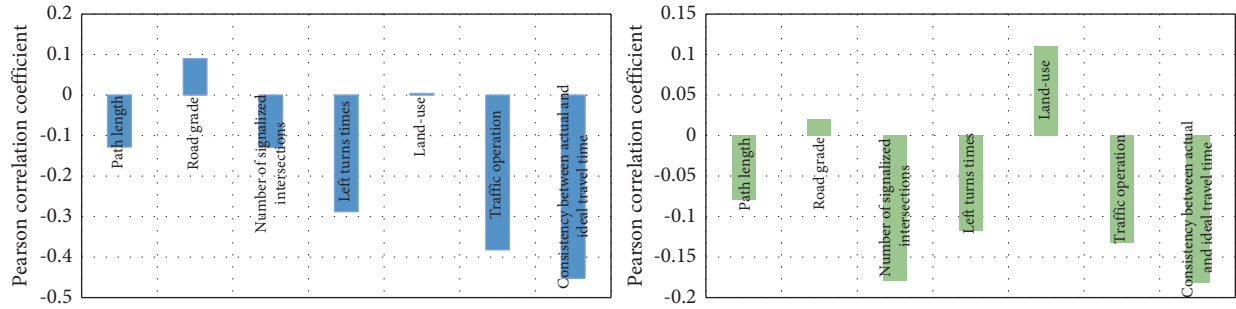


FIGURE 2: Correlation analysis of private car and taxi.

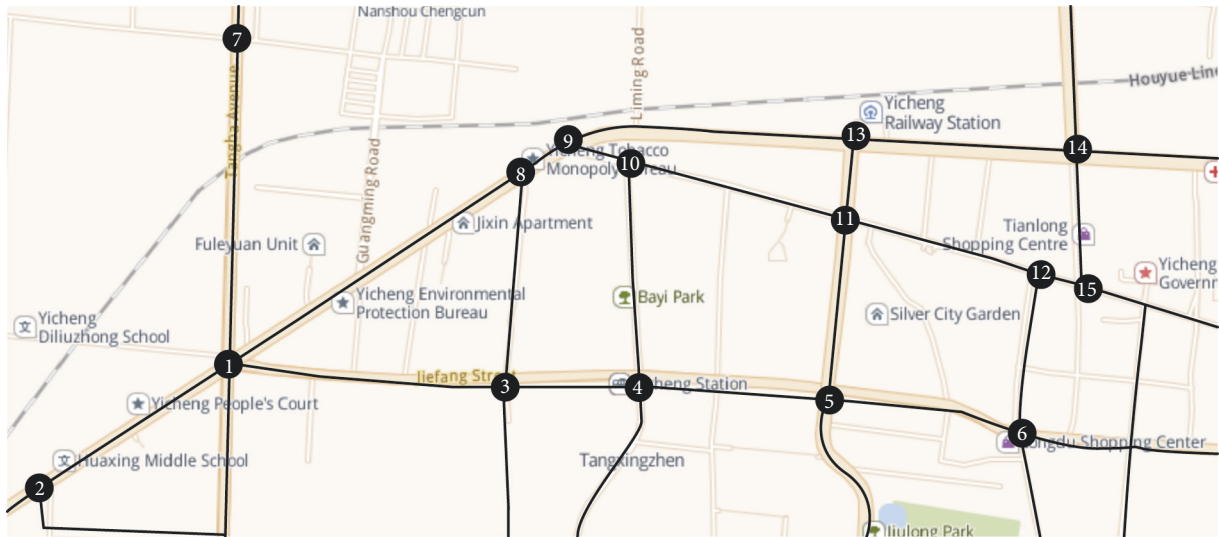


FIGURE 3: Road network map of Yicheng County.

TABLE 1: Correlation analysis of decision indicator and route selection probability of private cars.

Indicator	<i>p</i> value	Calculation method
Consistency between actual and ideal travel time	.000	$\tau_k^{rs} = T_k' - T_k/T_k$
Traffic operation	.002	$C_k^{rs} = \sum_{a(i,j) \in E} l_{a(i,j)} c_{a(i,j)} / L_k^{rs}$
Left turn times	.008	$\sum \psi_{rps} = P_{rp} \otimes P_{ps}$
Path length	.010	$L_k^{rs} = \sum_{a(i,j) \in E} l_{a(i,j)} \xi_{a(i,j),k}^{rs}$
Number of signalized intersections	.011	$S_k^{rs} = \sum_{a(i,j) \in E} l_{a(i,j)} s_{a(i,j)} / L_k^{rs}$
Road grade	.019	$D_k^{rs} = \sum_{a(i,j) \in E} l_{a(i,j)} d_{a(i,j)} / L_k^{rs}$

TABLE 2: Correlation analysis of decision indicator and route selection probability of taxis.

Indicator	<i>p</i> value	Calculation method
Consistency between actual and ideal travel time	.000	$\tau_k^{rs} = T_k' - T_k/T_k$
Number of signalized intersections	.001	$S_k^{rs} = \sum_{a(i,j) \in E} l_{a(i,j)} s_{a(i,j)} / L_k^{rs}$
Traffic operation	.002	$C_k = \sum_{a(i,j) \in E} l_{a(i,j)} c_{a(i,j)} / L_k^{rs}$
Land-use	.003	$N_k = \sum_{a(i,j) \in E} l_{a(i,j)} n_{a(i,j)} / L_k^{rs}$
Left turns times	.011	$\psi_{rps} = P_{rp} \otimes P_{ps} = 1$
Path length	.021	$L_k^{rs} = \sum_{a(i,j) \in E} l_{a(i,j)} \xi_{a(i,j),k}^{rs}$

$$c_{a,k} = \omega_1 L_k^{rs} + \omega_2 C_k^{rs} + \omega_3 S_k^{rs} + \omega_4 \sum \psi_{rps} + \omega_5 \tau_k^{rs} + \rho_1 \omega_6 D_k^{rs} + \rho_2 \omega_7 N_k^{rs}, \quad (1)$$

$$\min c_{a,k}, \quad (2)$$

where $c_{a,k}$ is the satisfaction value of the path k . ω_j is the combined weight value of the j indicator. L_k^{rs} is the path length of the path k between r and s . C_k^{rs} is the road operation condition of the path k between r and s . S_k^{rs} is the number of signalized intersections of the path k between r and s . $\sum \psi_{rps}$ is the number of left turns at the path k . τ_k^{rs} is the accordance rate of the theoretical transit time of path k between r and s with the actual transit time. D_k^{rs} is the road grade of the path k between r and s . N_k^{rs} is the land-use indicator of the path k between r and s . ρ_1 is the private car representative indicator. ρ_2 is the taxi representative indicator.

The constraint of the model is $\sum \omega_j = 1$, ensuring that the sum of weights is always 1.

4. Model Indicator Weight Calculation Based on AHP-EWM Method

AHP method is a method of determining weights that combines quantitative and qualitative analysis, and it is the principle that the decision maker's thinking and decision-making process can always be consistent in this paper. However, since expert scoring is used in weighting indicators in the decision-making process of the AHP method, it leads to the subjective orientation appearing in the decision-making process. The single use of the AHP method will have an impact on the evaluation results, leading to more subjective factors in the path selection. Therefore, to avoid subjective decision-making, the EWM method is introduced to correct the weight of the indicators, then the path recognition process is more consistent with the actual situation.

Assuming that there are n evaluation indicators for the k alternative paths of a certain type of vehicle. y_{ij} is the evaluation indicator j for the scheme i , where $i = 1, 2, \dots, k$, $j = 1, 2, \dots, n$. The calculation steps of the EWM method to correct the weight value are as follows.

Step 1. : Standardized processing. The purpose is to eliminate the interaction of dimensions and orders of magnitude between the indicators. And it is necessary to standardize the indicator y_{ij} in the original data. The common methods to eliminate the dimension are the extreme-range method and the Z-score method. The Z-score method will result in a negative entropy value due to the accumulation of errors, which is not following the entropy principle.

In this paper, the extreme-range method is considered to eliminate the dimensional problem between the indicators. The correlation analysis between the indicators and the probability of path selection shows that the two indicator values of road grade and land use are positively correlated benefit indicators, and all other indicators are negatively correlated cost indicators.

For the benefit indicator, the extreme-range formula is as follows:

$$y'_{ij} = \frac{y_{ij} - y_{\min}}{y_{\max} - y_{\min}}, \quad (3)$$

where y_{\max} and y_{\min} are the maximum and minimum values of the indicator j .

For cost indicators, the extreme-range formula is as follows:

$$y'_{ij} = \frac{y_{\max} - y_{ij}}{y_{\max} - y_{\min}}. \quad (4)$$

However, the extreme-range method will cause the maximum and minimum boundary values to be zero after the elimination of dimensions, which will affect the subsequent calculation of entropy. To avoid the problems, the shift operation is adopted after the standardization of the indicators so that:

$$y''_{ij} = y'_{ij} + 0.001. \quad (5)$$

Obtain the standardized decision matrix after the dimensionless processing of the indicators $Y'' = \{y''_{ij}\}_{k \times n}$, which is as follows:

$$Y'' = \begin{bmatrix} y''_{11} & y''_{12} & \cdots & y''_{1n} \\ y''_{21} & y''_{22} & \cdots & y''_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ y''_{k1} & y''_{k2} & \cdots & y''_{kn} \end{bmatrix}. \quad (6)$$

Step 2. : Calculate the proportion of each indicator value in the standardized matrix z_{ij} , which is as follows:

$$z_{ij} = \frac{y''_{ij}}{\sum_{i=1}^k y''_{ij}}. \quad (7)$$

Step 3. : According to the definition of information entropy, calculate the information entropy value of each indicator j , namely:

$$e_j = -\frac{1}{\ln k} \sum_{i=1}^k z_{ij} \ln z_{ij} \quad j = 1, 2, \dots, n. \quad (8)$$

Step 4. : Determine the weight of the indicator, which is as follows:

$$\omega_{E,j} = \frac{1 - e_j}{n - \sum_{j=1}^n e_j} \quad j = 1, 2, \dots, n. \quad (9)$$

To make the analysis results reflect the subjective factors and objective reality of the travelers, the weight values of each index obtained by the two methods are combined and calculated by using the Lagrange multiplier method based on the principle of minimum information entropy. Based on the AHP-EWM method, the weight vector ω is as follows:

$$\omega_j = \frac{\sqrt{\omega_{A,i}\omega_{E,j}}}{\sum_{j=1}^n \sqrt{\omega_{A,i}\omega_{E,j}}} \quad j = 1, 2, \dots, n. \quad (10)$$

$$Y = \begin{bmatrix} 1.55 & 2.76 & 1 & 2.13 & 4 & 6.37 \\ 4.37 & 4.56 & 0 & 2.07 & 3 & 5.26 \\ 8.50 & 2.95 & 2 & 2.34 & 4 & 5.85 \\ 5.12 & 3.73 & 1 & 2.27 & 4 & 5.24 \end{bmatrix}. \quad (11)$$

5. Case Study

5.1. Experimental Design. To verify the accuracy of the vehicle path recognition model, take the incomplete path set (remove some nodes from the full path) as the testbed and make a comparison of real vehicle path data and experimental results. Figure 3 is a map of the actual road network in a town. The node represents the major intersection in the road network, each intersection has AVI equipment, but due to the unstable system conditions, some equipment cannot do the real-time collection to the vehicle information.

We randomly select two complete travel paths of private cars and taxis between node 3 and node 14. The diagram of the actual travel paths of private cars and taxis is shown in Figure 4. The paths chosen by the two types of vehicles are different. To obtain the incomplete paths, we erased all links between node 3 and node 14, retained node 3 and node 14, and recomplete the path according to the vehicle path recognition model.

The actual travel records of private cars are shown in Table 3.

The actual travel records of taxis are shown in Table 4.

5.2. Calculation Process

5.2.1. Generate Reasonable Path. According to the K-shortest algorithm and the deletion algorithm, the reasonable path set of private cars and taxis between node 3 and node 14 is obtained. The direction of private cars at node 14 is from West to East. It can be inferred that the upstream intersection is node 13 in the West. Search for the first five shortest paths with the fixed upstream intersection of the end node and delete the paths with loops and detours. Finally, the first four alternative paths $R_k^{car} = \{R_1^{car}, R_2^{car}, R_3^{car}, R_4^{car}\}$ for private cars are selected. The detailed alternative paths for private cars are shown in Table 5.

Similarly, the first three alternative paths $R_k^{taxi} = \{R_1^{taxi}, R_2^{taxi}, R_3^{taxi}\}$ for taxis are selected. The detailed alternative paths for taxis are shown in Table 6.

5.2.2. Calculation Process of the Path Recognition Model.

The path recognition process of a private car is taken as an example to verify the calculation process. The attribute value of each indicator, according to the quantitative calculation formula of the attribute value of each path, is calculated to obtain the indicator matrix $Y = \{y_{ij}\}_{4 \times 6}$ (Formula (3)). The rows of the matrix represent the path number (Table 3), and the columns from left to right each represent the value of an indicator, such as the consistency of passing time, road operation, number of left turns, path length, number of signalized intersections, and road category.

5.2.3. Indicator Weight Calculated by AHP Method. By constructing the indicator comparison matrix, the weight value vector of each indicator is obtained.

$$\omega_A = (0.3894 \quad 0.2125 \quad 0.1035 \quad 0.0335 \quad 0.0659 \quad 0.1952). \quad (12)$$

The maximum eigenvalue is as follows:

$$\lambda_{\max} = (6.58 \quad 6.3276 \quad 6.5926 \quad 6.3738 \quad 6.6484 \quad 6.2146). \quad (13)$$

The consistency indicator $CI = 0.074 < 0.1$, which passes the consistency test. Therefore, the indicator weight meets the requirements.

5.2.4. Calculate Indicator Weight by EWM Method. After range standardization and shifting 0.001 units to the right, the standardized matrix $Y'' = \{y_{ij}''\}_{4 \times 6}$ is calculated.

$$Y'' = \begin{bmatrix} 1.001 & 1.001 & 0.501 & 0.781 & 0.001 & 1.001 \\ 0.591 & 0.001 & 1.001 & 1.001 & 1.001 & 0.021 \\ 0.001 & 0.891 & 0.001 & 0.001 & 0.001 & 0.531 \\ 0.491 & 0.461 & 0.501 & 0.261 & 0.001 & 0.001 \end{bmatrix}. \quad (14)$$

The weight value vector of each indicator is obtained.

$$\omega_E = (0.1274 \quad 0.1472 \quad 0.1273 \quad 0.1595 \quad 0.2434 \quad 0.1951). \quad (15)$$

5.2.5. Calculate Combination Weight. According to formula (10), the weight values of each indicator obtained by the two methods are combined and calculated to obtain the weight vector ω based on the improved weight calculation by the AHP-EWM method.

$$\omega^{car} = (0.245 \quad 0.1945 \quad 0.1263 \quad 0.0804 \quad 0.1393 \quad 0.2146). \quad (16)$$

5.2.6. Calculate the Most Satisfactory Path. Substitute the weight calculation result into the model formula (1) to calculate the value of $c_{a,k}$.

$$c_{a,k}^{car} = [3.1374 \quad 3.6713 \quad 4.9079 \quad 3.9704], \quad (17)$$

$$\min c_{a,k}^{car} = 3.1374.$$

The result means that the first path is the actual shortest path calculated by the model, as the path where the private car is finally recognized.

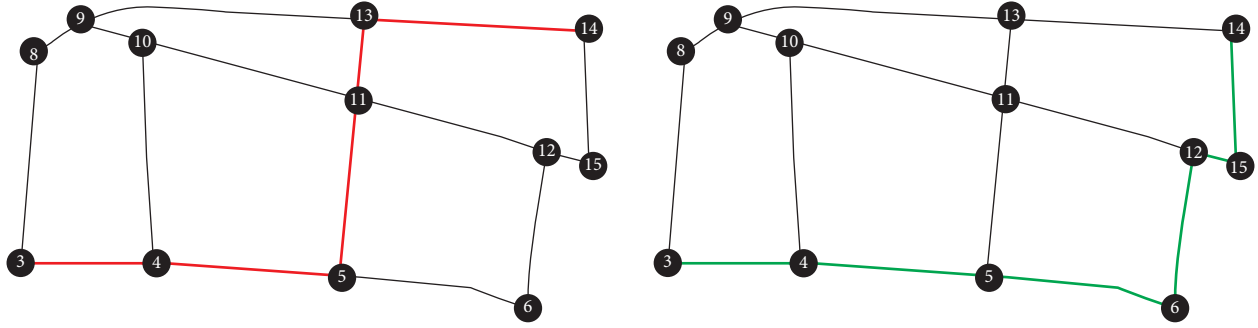


FIGURE 4: Diagram of the actual travel path of private cars and taxis.

TABLE 3: Actual travel records of private cars.

Number	Spot location	Direction	Departure time
1	3	West to east	16 : 11 : 56
2	4	West to east	16 : 12 : 33
3	5	South to north	16 : 13 : 11
4	11	South to north	16 : 15 : 56
5	13	West to east	16 : 17 : 32
6	14	West to east	16 : 18 : 34

TABLE 4: Actual travel records of taxis.

Number	Spot location	Direction	Departure time
1	3	West to east	11 : 15 : 11
2	4	West to east	11 : 15 : 59
3	5	West to east	11 : 17 : 12
4	6	South to north	11 : 19 : 06
5	12	West to east	11 : 22 : 32
6	15	South to north	11 : 23 : 04
7	14	South to north	11 : 24 : 07

TABLE 5: Alternative paths for private cars.

Alternative paths number	R_1^{car}	R_2^{car}	R_3^{car}	R_4^{car}
Path sequence	{3, 4, 5, 11, 13, 14}	{3, 8, 9, 13, 14}	{3, 4, 10, 11, 13, 14}	{3, 8, 9, 10, 11, 13, 14}

TABLE 6: Alternative paths for taxis.

Alternative paths number	R_1^{taxi}	R_2^{taxi}	R_3^{taxi}
Path sequence	{3, 4, 5, 6, 12, 15, 14}	{3, 4, 5, 11, 12, 15, 14}	{3, 8, 9, 10, 11, 12, 15, 14}

Similarly, the most satisfactory path for taxis is calculated.

$$c_{a,k}^{taxi} = [3.912 \ 4.112 \ 4.832], \quad (18)$$

$$\min c_{a,k}^{taxi} = 3.912.$$

The result means that the first alternative path is the path where the taxi is finally recognized.

5.3. Results and Discussion. The calculation results of the model are consistent with the actual results, and the most satisfactory path is also consistent with the actual path.

For private car: the path selection order sorted by model results is $R_1^{car} > R_2^{car} > R_4^{car} > R_3^{car}$. However, if we sort the path selection order by its length, R_2^{car} is the shortest path.

Considering the traffic condition of the link R_2^{car} , sections 3-7 have lower road categories and poor traffic conditions. Congestion is likely to occur due to the increased travel activities during the evening peak. The result shows that it is consistent with the actual situation analysis and can objectively reflect the actual situation.

For taxi: the path selection order sorted by model results is $R_1^{taxi} > R_2^{taxi} > R_3^{taxi}$. The road conditions of sections 5-11 are better than sections 6-12. However, both paths of model results and actual results contain sections 6-12 instead of sections 5-11. Based on actual investigations, sections 6-12 are in a commercial area with large shopping malls and stores. Combining the travel time periods, the commercial area is the preferred destination for taxis, no matter whether the taxi is carrying passengers or not.

Through the above analysis, it can be concluded that the actual path chosen by travelers is not necessarily the shortest distance path, which is also compatible with the K-Shortest Paths (KSP) problem. For the same OD, the actual path results of private cars and taxis are different, as well as the path results obtained through the path recognition model are also different and consistent with the actual situation. It is proved that the model can accurately identify the vehicle travel path to a certain extent, which shows that the model established in this paper is effective for identifying the missing paths in the AVI data.

6. Influence of the Number of Missing Nodes on Model Accuracy

6.1. Experiment Design. To increase the test sample capacity, the road network is expanded to the central area of Yicheng County. The model assumes that vehicles driving in extremely narrow alleys are not considered, so the road network is simplified into a network composed of main trunk roads, secondary trunk roads and slip roads, as shown in Figure 5. The black solid nodes in the road network map are intersections with AVI device, and the white nodes are virtual. There is little AVI data at some intersections, and the phenomenon of missing vehicle detection is common. When the distribution density of the AVI devices in the road network is equal to that of the nodes in the road network, it is easy to obtain the travel records of the vehicles and analyze the traffic characteristics by mining the complete AVI data. However, due to the imperfect equipment and low coverage, the travel path cannot be obtained directly from the incomplete AVI data.

For the three paths with different distances, 100 private car paths and 50 taxi paths are randomly selected as the test data set. The actual paths of the vehicle are shown in Figures 6–8. To compare the experimental results with the real situation and analyze the error, three groups of experiments are designed according to the number of missing nodes. Remove some nodes in the complete path, and complete the path between the start and end nodes of the missing path according to the path recognition model.

The short-distance path represents the path shorter than 1 km, the medium distance path represents the path within 1 to 2 km, and the long-distance path represents the path longer than 2 km.

According to the number of missing nodes, the experiment is divided into three groups, as shown in Table 7.

The first group: 20% of the nodes are missing.

The second group: 50% of the nodes are missing.

The third group: all nodes are missing except for the origin and destination.

The method for selecting missing nodes is to generate random numbers in percentage from the sequence numbers of other nodes on the path except the start node and delete the corresponding node records.

The procedure is as follows.

Step1 : Use Dijkstra and deletion algorithm to obtain the alternative path set of the missing path in each case

combined with the driving direction of the vehicle and determine the attribute indicator values on the path.

Step2 : Calculate the weight value of each attribute indicator according to the analytic hierarchy process, use MATLAB to calculate the entropy weight value of each attribute indicator, and then calculate the combined weight and the model result.

Step3 : Compare the quantity ratio of the model calculation results that are consistent with the actual path, and calculate the accuracy of the model.

Theoretically, the model calculation results should be consistent with the actual travel path. However, due to the different traffic operations in different periods, some paths may have high similarities with the theoretical path (path calculated by the proposed model). The length of the common substring $Public(R_k, R_{actual})$ represents the number of common sections of actual paths and theoretical paths, where R_k represents the k^{th} alternative path in the alternative path set and R_{actual} represents the theoretical travel path as well as the actual path. Then the accuracy of the model could be expressed as follows:

$$\gamma_j = \frac{\sum_{i=1}^m Public(R_k, R_{actual})}{m} \times 100\%. \quad (19)$$

6.2. Results Analysis. Through the analysis of the above experimental results, we can get the path decision-making results and accuracy of the three groups of experiments.

6.2.1. 20% of the Total Number of Nodes in the Path Were Missing. We assume that path (a) missed the data of node 3, whose front node is a six-legged intersection with complex flow directions. So, node 3, as the first intersection after turning, plays a key role in path selection. The results show that 95 percent of private car experimental paths are consistent with the actual paths; 100 percent of taxi experimental paths selections are consistent with the actual paths. Node 4 is lost in the long-distance path (c). The results show that all vehicles in the test path set of private cars and taxis choose the path consistent with the actual path. Since node 4 is between node 3 and node 5, when the number of missing nodes is low, the vehicle driving direction is clear, and there will not be too many path selection behaviors on short-distance sections, which is in line with the actual travel situation.

In conclusion, the model has good robustness when the number of missing nodes in the path is less than or equal to 20 percent of the total number of nodes on the path. The experimental results are shown in Table 8.

6.2.2. 50% of the Total Number of Nodes in the Path Were Missing. The short-distance path (a) missed nodes 3 and 4. Node 1 is a six-legged intersection, at which it is hard to know the vehicle's movement direction. So the vehicle needs to make a path decision after node 1. The results show that there are 13 vehicle paths in the private car test path set

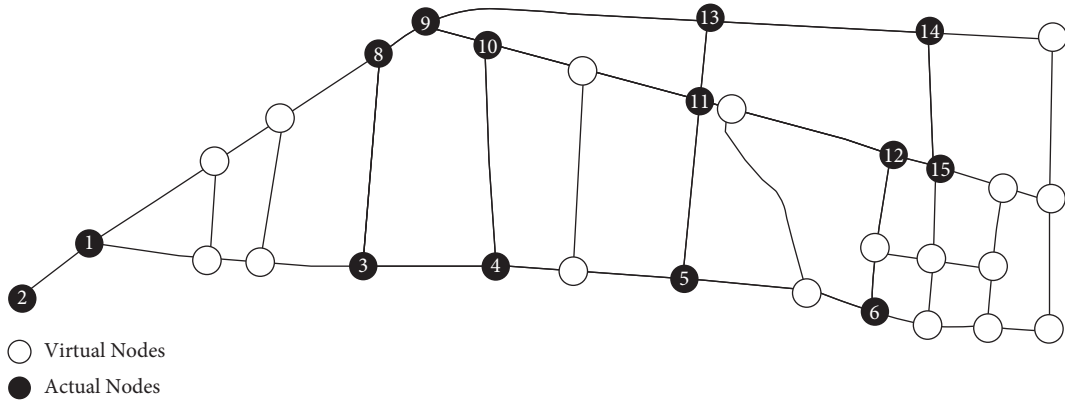


FIGURE 5: Simplified road network map of Yicheng County.

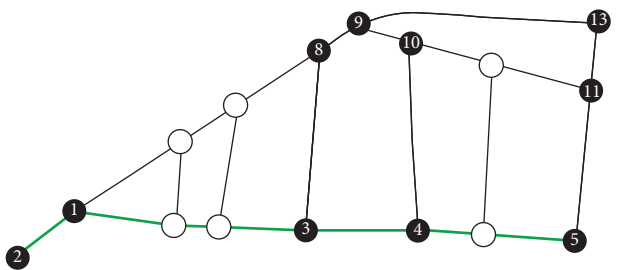


FIGURE 6: Short distance path a.

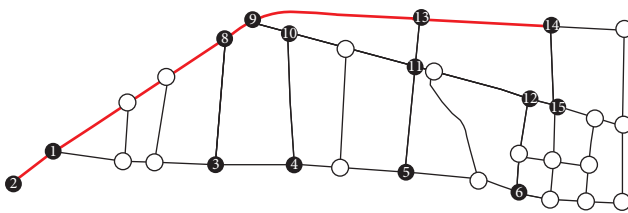


FIGURE 7: Medium distance path b.

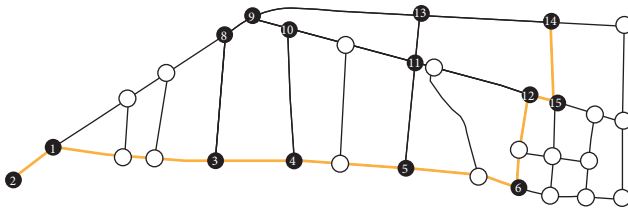


FIGURE 8: Long-distance path c.

inconsistent with the actual path. Besides the actual path, the most selected path is {2-1-8-9-13-11-5}. Although the distance of this path is longer, other attributes of the path have certain advantages, such as high road grade and short travel time. In addition, through the actual survey, traffic congestion occurs on the link {3-4-5} during the peak period, which is also one of the reasons for vehicles to select other paths. 16 percent of the taxi test paths chose other paths.

The medium distance path (b) missed node 9 and node 13. Node 9 is the key node of this path, where the vehicle may have decision behavior. The results show that 90 percent of the private car test paths are the same as the actual paths, as

shown in Table 9. Because this path has good performance and fewer decision nodes, the alternative paths have poorer performance than the actual paths. 80 percent of the taxi test paths are consistent with the actual paths.

The long-distance path (c) missed nodes 3,4,5, and 6 as the key node of the path, the vehicle at node 1 has a greater possibility of decision making. According to the movement direction of the vehicle at node 12, the position of the previous node can be known, which can narrow down the search for alternative paths.

6.2.3. All Nodes in the Path are Missing Except for the Origin and Destination. Three key nodes are missing between the origin and destination of the short-distance path (a). Since the vehicle has node 1 as the only direction of movement after leaving node 2, this experimental condition is similar to the second set of experiments in which path (a) missed 2 nodes. The calculation results are also similar, as shown in Table 10.

Four nodes were missed between the origin and destination of the medium distance path (b). Similar to the path (a), the downstream intersection of node 2 is determined as node 1 based on the uniqueness of the direction after leaving node 2. Multiple path decisions are possible at node 1. The top 5 paths are selected based on the K-shortest path problem, and the front node position can be known according to the vehicle’s moving direction at node 14, to reduce the path search range. The results are shown in Table 10.

Seven nodes were missed between the origin and destination of the long-distance path (c). Similarly, the path search range of the K-shortest path problem can be narrowed down, and the actual number of missing nodes is 5. The results are shown in Table 10. The accuracy of the results is decreased due to the accumulation of errors when the vehicle makes path decisions at nodes.

Meanwhile, the shortest path algorithm (Shortest Path, SP) is selected to compare with the path decision model in this paper, and the selected metric is the accuracy rate.

When the number of missing nodes is low, the accuracy of the results calculated by the algorithm in this paper and the shortest path algorithm is close, which indicates that

TABLE 7: Experimental design table.

Path number	Group 1	Group 2	Group 3
Short distance(a)	Missing node 3	Missing node 3, 4	Keeping node 2, 5
Medium distance(b)	Missing node 1	Missing node 9, 13	Keeping node 2, 14
Long distance(c)	Missing node 4	Missing node 3, 4, 5, 6	Keeping node 2, 14

TABLE 8: The first group of experimental results.

Path no.	Private car			Taxi		
	Sample size	Consistent number	Accuracy (%)	Sample size	Consistent number	Accuracy (%)
(A)	100	95	95	50	50	100
(B)		93	93		49	98
(C)		100	100		50	100

TABLE 9: The second group of experimental results.

Path no.	Private car			Taxi		
	Sample size	Consistent number	Accuracy (%)	Sample size	Consistent number	Accuracy (%)
(A)	100	87	87	50	42	84
(B)		90	93		40	80
(C)		69	77		43	86

TABLE 10: The third group of experimental results.

Path no.	Private car			Taxi		
	Sample size	Consistent number	Accuracy (%)	Sample size	Consistent number	Accuracy (%)
Short distance(a)	100	81	81	50	40	80
Medium distance(b)		73	73		37	74
Long distance(c)		55	69		35	70

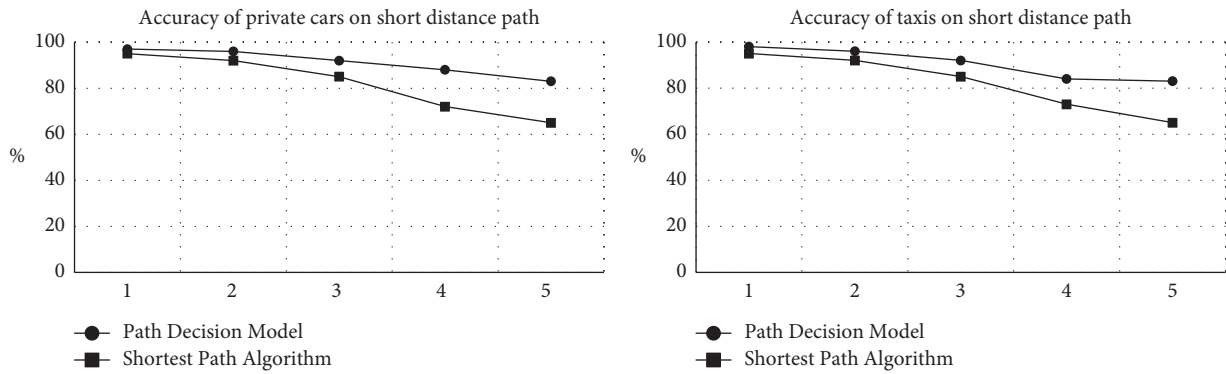


FIGURE 9: Accuracy of taxis and private cars on short-distance path.

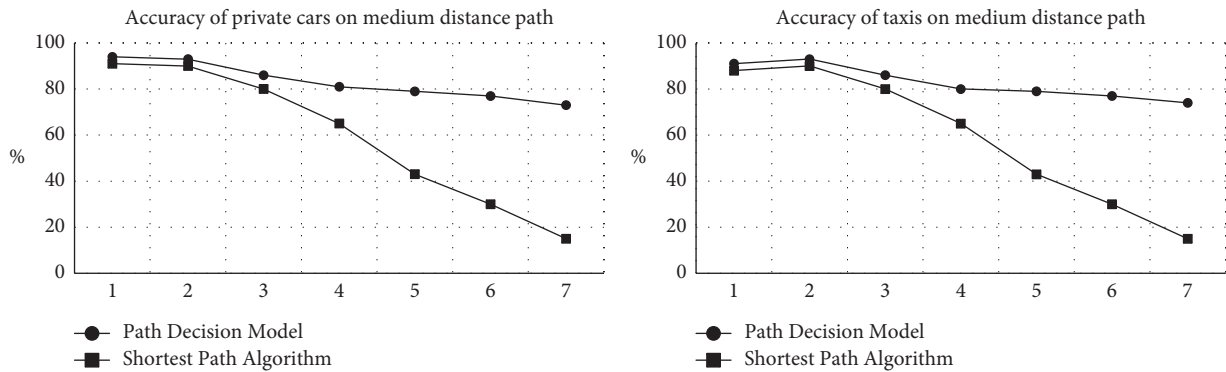


FIGURE 10: Accuracy of taxis and private cars on medium distance path.

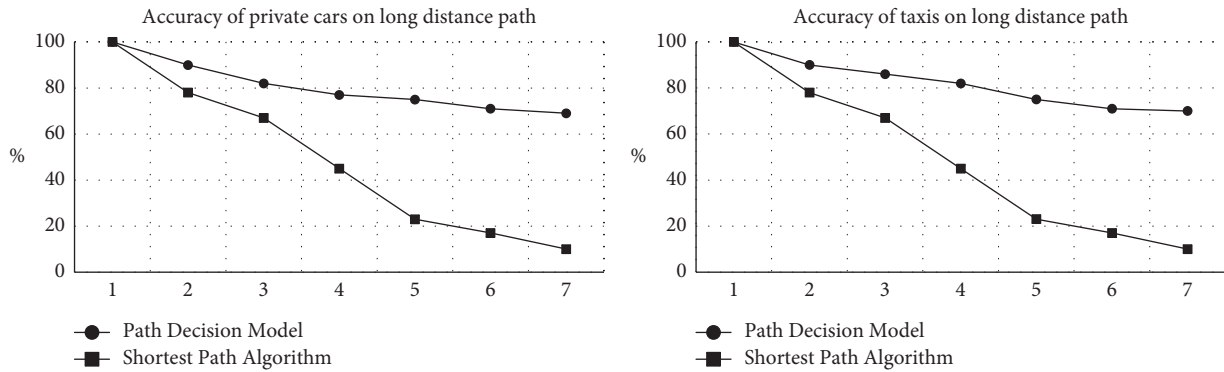


FIGURE 11: Accuracy of taxis and private cars on the long-distance path.

vehicles generally choose the shortest path to travel when the travel distance is short. When the number of missing nodes increases, the accuracy of both methods decreases, but the decision results of the model are significantly better than those of the shortest path algorithm.

The number of missing nodes in short-distance paths is generally less than or equal to 5. In Figure 9, it can be seen that the sensitivity of private cars and taxis to missing nodes is close in short-distance travel. The accuracy of the model is also close to the results of the shortest path algorithm.

For medium-distance paths, the performance of private cars and taxis is close, as shown in Figure 10. The accuracy of the decision results of the model in this paper has a relatively stable accuracy rate. When the number of missing nodes for medium distance and long-distance paths is over 4, the accuracy of the shortest path algorithm drops sharply, and the accuracy of the results of the model in this paper still maintains around 80%.

For the long-distance path decision, the accuracy of the proposed model decreases significantly, as shown in Figure 11. When the number of missing nodes exceeds 3, the accuracy of the shortest path algorithm starts to drop sharply, and when the number of missing nodes exceeds 4, the accuracy of the shortest path algorithm falls below 50%, but the accuracy of the proposed model is still around 80%.

In all, the accuracy of the model in this paper is better than the shortest path algorithm overall. In the short-distance missing paths, the accuracy of the model calculation results can still maintain a good level as the number of missing nodes increases. In the medium-distance missing paths, the accuracy of the model calculation results decreases slightly as the number of missing nodes increases. In the long-distance missing paths, the accuracy decreases as the number of missing nodes increases.

7. Conclusion

This paper established a vehicle path recognition model based on the road network of small cities with low coverage of AVI devices and data deficiency, considering the difference in travel path selection indicators of private cars and taxis. The method of decision indicators is defined, and the AHP-EWM combination method is used to get the

indicator weights to carry out the path recognition. Firstly, the quantitative calculation methods of decision indicators are defined. The AHP-EWM combination method is used to get the indicator weights to carry out the path recognition. Through the validation of the validity and accuracy of the model, it is proved that the travel path recognized by the model is basically consistent with the actual travel paths.

However, with the increase of the number of missing nodes (when the number of missing nodes is more than 7), the accuracy of the model may further decrease.

Considering the actual deployment of the AVI system in road networks, this paper draws the following conclusions.

- (1) In small cities with low coverage of AVI devices, when the missing records in the AVI data are serious, the vehicle travel paths of private cars and taxis can be identified separately according to the path recognition model in this paper. The results can effectively characterize the actual situation of vehicle travel paths, which could provide data resources for analysis of the characteristics of mixed traffic flow with CAVs.
- (2) Through the analysis of the number of missing nodes in the AVI system, the following conclusions can be drawn. Within a certain area, it is not necessary to deploy the AVI devices in all sections. The model of this paper can effectively identify the travel path of vehicles, which could save system construction and equipment maintenance costs.
- (3) Considering the different travel factors of taxis and private cars, the travel paths of vehicles are recognized separately and consistent with the travel behavior of vehicles in the actual road network.

Data Availability

Data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This study was supported by the project National Key R&D Program of China (No. 2019YFF0301401), and Beijing Natural Science Foundation (J210001).

References

- [1] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate-recognition algorithm for intelligent transportation system applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 3, pp. 377–392, 2006.
- [2] C. A. Rahman, W. Badawy, and A. Radmanesh, "A real time vehicle's license plate recognition system," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 163–166, Miami, FL, USA, July 2003.
- [3] H. A. Hegt, R. Haye, and N. A. Khan, "A high performance license plate recognition system," in *Proceedings of the IEEE International Conference on Systems*, San Diego, USA, October 1998.
- [4] T. Ajanthan, P. Kamalaruban, and R. Rodrigo, "Automatic number plate recognition in low quality videos," in *Proceedings of the IEEE International Conference on Industrial & Information Systems*, Peradeniya, Sri Lanka, December 2013.
- [5] Y. Lee, J. Jun, Y. Hong, and M. Jeon, "Practical license plate recognition in unconstrained surveillance systems with adversarial super-resolution," in *Proceedings of the 14th International Conference on Computer Vision Theory and Applications*, Prague, Czech Republic, 2019.
- [6] Y. Asakura, E. Hato, and M. Kashiwadani, "Origin-destination matrices estimation model using automatic vehicle identification data and its application to the Han-Shin expressway network," *Transportation*, vol. 27, no. 4, pp. 419–438, 2000.
- [7] Y. Yang, Z. Yuan, J. Li, and Y. Wang, "Multi-mode public transit OD prediction and scheduling model," *Advances in Transportation Studies*, vol. 3, pp. 133–146, 2018.
- [8] S. Sánchez-Cambrotero, E. Castillo, J. M. Menéndez, and P. Jiménez, "Dealing with error recovery in traffic flow prediction using bayesian networks based on license plate scanning data," *Journal of Transportation Engineering*, vol. 137, no. 9, pp. 615–629, 2011.
- [9] Y. Yang, Z. Yuan, and R. Meng, "Exploring traffic crash occurrence mechanism towards cross-area freeways via an improved data mining approach [J]," *Journal of Transportation Engineering Part A Systems*, 2022.
- [10] F. Dion and H. Rakha, "Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates," *Transportation Research Part B: Methodological*, vol. 40, no. 9, pp. 745–766, 2006.
- [11] X. Li and Z. Zhai, "Travel time calculation method based on the electric police data," in *Proceedings of the 2014 9th China Intelligent Transportation Conference Paper Collection*, China Intelligent Transportation Association, Prague, Czech Republic, 2014.
- [12] L. Li, W. Wang, and Y. Wang, *An Adaptive Traffic Signal Control System Based on Electric Police Data and its Working Method*, 2018.
- [13] T. Xu and S. Fei, *A Method of Dynamic Green Wave Control of Traffic Signals Based on Electro-Police Data*, p. 31, 2019.
- [14] Y. Yang, K. He, and Y. Wang, "Identification of dynamic traffic crash risk for cross-area freeways based on statistical and machine learning method," *Physica A: Statistical Mechanics and Its Applications*, vol. 595, p. 127083, 2022.
- [15] K. Tang, T. Xu, A. Pan, and S. Li, "Signal timing and detector data-based reconstruction of vehicle trajectories on urban arterials," *Journal of Tongji University*, vol. 44, no. 10, pp. 1545–1552, 2016.
- [16] J. Xiang and K. Tang, *The Method of Refactoring the Trajectory of the Main Road Vehicle Based on the Fusion of Electric Police and Fixed-Point Detector Data*, p. 24, 2018.
- [17] Z. Yu, Q. Liao, and Z. He, "Vehicle trajectory reconstruction in signalized-link using vehicle identification data," *Journal of Transportation Systems Engineering and Information Technology*, vol. 19, no. 4, pp. 87–93, 2019.
- [18] Y. Lin, H. Ding, Y. Liu, and X. Lin, "A study of the emission trajectory of a single vehicle based on vehicle identity detection data," *Environmental Science in China*, vol. 39, no. 12, pp. 4929–4940, 2019.
- [19] Y. Yang, Z. Yuan, J. Chen, and M. Guo, "Assessment of osculating value method based on entropy weight to transportation energy conservation and emission reduction," *Environmental Engineering & Management Journal (EEMJ)*, vol. 16, no. 10, 2017.
- [20] H. Zhang, X. Liu, and Y. Xu, "Delay extraction based on vehicle's trajectory reconstruction at signalized intersection," *Journal of Transportation Systems Engineering and Information Technology*, vol. 20, no. 02, pp. 237–243, 2020.
- [21] W. Rao, Y. Wu, and J. Xia, "Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data," *Transportation Research Part C: Emerging Technologies*, vol. 95, no. OCT, pp. 29–46, 2018.
- [22] W. Li, *A Study on the Method of Vehicle Travel Path Extraction Based on License Plate Identification Data*, Guilin University of Electronic Science and Technology, Guilin, China, 2019.
- [23] B. Mo, R. Li, and J. Dai, *Estimating dynamic origin-destination demand: a hybrid framework using license plate recognition data*, *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 7, pp. 734–752, 2020.
- [24] Y. Guo, *Research on Dynamic Identification and Perimeter Control Strategies of Urban Traffic Congested Regions*, Shandong University, Jinan, China, 2020.
- [25] L. Wang, "Loss track reduction in vehicle travel track survey analysis," *Computer Application Research*, vol. 31, no. 01, pp. 162–165, 2014.
- [26] S. Yang and H. Yu, "Vehicle travel trajectory reconstruction based on monitoring data. China intelligent transportation systems association," in *Proceedings of the Papers of the First China Intelligent Transportation Annual Conference*, China Intelligent Transportation Systems Association, 2016.
- [27] B. Yang, *Motor Vehicle Track Extraction and Key Road Excavation Analysis Based on Traffic Jam Data*, Beijing Jiaotong University, Beijing, China, 2020.
- [28] Y. Yang, Z. Yuan, X. Fu, Y. Wang, and D. Sun, "Optimization model of taxi fleet size based on GPS tracking data," *Sustainability*, vol. 11, no. 3, p. 731, 2019.
- [29] Y. Qiu, *Analysis of the Choice Factors of Private Car Travel Path Based on RFID*, Wuhan University of Technology, Wuhan, China, 2016.