

Article

Method Based on Floating Car Data and Gradient-Boosted Decision Tree Classification for the Detection of Auxiliary Through Lanes at Intersections

Xiaolong Li ^{1,2}, Yuzhen Wu ^{1,*}, Yongbin Tan ^{1,2}, Penggen Cheng ^{1,*}, Jing Wu ¹
and Yuqian Wang ¹ 

¹ Faculty of Geomatics, East China University of Technology, Nanchang 330013, China; lixiaolong@ecit.cn (X.L.); ybtan@ecit.cn (Y.T.); wuj@ecit.cn (J.W.); neo@ecit.cn (Y.W.)

² Key Laboratory of Watershed Ecology and Geographical Environment Monitoring, NASG, Nanchang 330013, China

* Correspondence: lixlgis@163.com (Y.W.); pgcheng@ecit.cn (P.C.)

Received: 25 May 2018; Accepted: 19 July 2018; Published: 5 August 2018



Abstract: The rapid detection of information on continuously changing intersection auxiliary through lane is a major task of lane-level navigation data updates. However, existing lane number detection methods possess long update cycles and high computational costs. Therefore, this study proposes a novel method based on floating car data (FCD) for the detection of auxiliary through lane changes at road intersections. First, roads near intersections are divided into three sections and the spatial distribution characteristics of the FCD of each section are analyzed. Second, the FCD is preprocessed to obtain a standardized FCD dataset by removing redundant data through an improved amplitude-limiting average filtering method. Third, a basic classifier for the number of lanes is constructed. Fourth, the final number of lanes of the road section is determined by combining the basic classifier and the gradient-boosted decision tree model. Finally, the presence of an auxiliary through lane at the intersection is determined in accordance with the change in the number of intersection lanes. The method was tested using data for a road in Wuchang District, Wuhan City. Experimental results show that this method can rapidly obtain auxiliary through lane information from the FCD and is superior to other classification methods.

Keywords: floating car data; gradient-boosted decision tree; intersection lane; lane number; auxiliary through lane

1. Introduction

The intersection is a basic element of the urban road network. Setting up an auxiliary through lane effectively relieves traffic congestion by increasing the number of lanes at intersections [1–5]. However, the number of lanes at intersections continually changes due to alterations in urban construction, traffic control, and lane functions. The timely detection of information on intersection auxiliary through lanes is therefore necessary to facilitate the rapid updating of navigation data. Traditional methods for lane information extraction, such as professional measurements and the interpretation and recognition of road features from images and videos, are characterized by high data acquisition costs and long update cycles, which result in update lag and obsolete road data [6].

A high-speed differential global positioning system (DGPS) can be used to extract lane information rapidly from current ubiquitous and massive space-time trajectory data containing abundant road information. Although this system can extract information on lane level [7,8], lane centerline [9], lane boundary line [10], and lane position and width [11], its implementation requires sophisticated data acquisition equipment and complicated processes. Moreover, its sampling area is limited.

By contrast, the use of floating car data (FCD) is characterized by large data volume, strong real-time capability, wide coverage, and low cost. The timely detection of road information using the trajectory data of floating cars is a hot research topic [12–16]. Tang et al. [17] used low-frequency FCD to identify the intersections of urban road networks automatically and to extract the detailed structure of the intersection. These methods show that FCD can enable rapid access to information on urban road intersections and lanes.

Road construction generally widens intersections on the side of the stop line to relieve traffic pressure based on the space-for-time principle. Existing research on auxiliary through lanes have mainly focused on the analysis of lane selection models and capacity. Video recording, intersection monitoring, and manual survey methods are commonly used in these studies [1,2]. These methods have the disadvantages of high data collection cost and limited data coverage as well as the inability to obtain timely access to changes in the number of auxiliary through lanes at intersections due to traffic control, urban construction, and lane function changes. Although lane information can be obtained through low-cost FCD, related research has mainly focused on detecting the number of lanes on the road segment and the structure of the intersection [18]. Thus, a new approach based on FCD for exploring auxiliary through lane information at intersections must be explored.

In this study, a road is divided into different regions (intersection and non-intersection) in accordance with the Urban Road Intersection Planning Specification [19], and the spatial distribution characteristics of the FCD covered by the regions are analyzed. The regions are segmented in intervals using a vector map of the city road network, and an improved amplitude-limiting average filtering method is used to eliminate gross errors and obtain the characteristic parameters of the target data. Subsequently, a classifier is constructed using sections with a known number of lanes and an FCD coverage state as the training samples, and the number of lanes is confirmed using a gradient-boosted decision tree (GBDT). Finally, the different regions of the road are compared and analyzed, and road intersection information, especially the number of lanes at the intersection, is then effectively acquired. This method compensates for the inadequacies of traditional road detection technology in the extraction of intersection lane information using FCD to improve the accuracy rate of lane detection. Furthermore, it promptly and efficiently detects the number of auxiliary through lanes at an intersection and enables the updating of road network navigation data.

2. Related Works

This research is mainly aimed at detecting auxiliary through lanes at intersections based on floating car data (FCD). So far, research on the timely detection of road information using FCD have focused on the calculation of number of road lanes and the detection of the geometric structure of intersections; there is a lack of research on methods and techniques for the detection of auxiliary through lanes at intersections. The existing related research work can be divided into two major parts according to the research content: road intersection structure and road lane information detection methods.

Research on road intersections in the field of traffic control have mainly focused on lane model selection and lane capacity analysis, and field surveys have been generally used to obtain experimental data. Ma et al. [1] used a combination of field investigation and statistical analysis to study the traffic characteristics of the auxiliary through lane at the intersection. Guo [2] adopted the research method of lane selection model and traffic capacity analysis to simulate and analyze the canalization and traffic capacity of the auxiliary through lane. Tarawneh et al. [4] used field data to analyze the length and utilization rate of auxiliary through lanes and then studied the effect of auxiliary through lanes on utilization of downstream right-turn volume. The above research methods have limited data coverage and cannot describe the overall situation of auxiliary through lanes at full time.

Detection of road lane information has become a research hotspot in the realm of geo-information, with the increase in lane-level navigation requirements. Existing road information acquisition technologies mainly include the use of professional measurement, images and videos, high-quality Global Positioning System (GPS) data, and FCD, as shown in Table 1.

Table 1. Contrast analysis of lane number detection techniques.

Techniques	Advantage	Shortcomings
Professional measurement	Rich details and reliable accuracy	The high cost, long update cycle, and complicated processes
Images and videos	Wide data coverage and mature image interpretation technology	The data quality is easily affected by weather, light, pedestrians, vehicles, and other factors
High-quality GPS	Can extract refined lane information, complex intersection structure	Professional equipment, high data acquisition costs, long update cycles, and complex processes
FCD	Large amount of data, good real-time capability, wide coverage, low cost	Low data accuracy

Extracting road information using dedicated high-quality GPS, Rogers [7] combined spatial-temporal GPS trajectory data with a hierarchical agglomerative clustering algorithm to extract road centerlines and lane boundary lines. Relevant scholars have done a lot of research in the use of space-time GPS trajectory data on this basis to detect lane information of urban roads and the geometry structure of complex intersections. Chen et al. [8] used a Gaussian mixture model (GMM) to model the distribution of GPS trajectory data of a shuttle vehicle across multiple traffic lanes and automatically computed the number and locations of driving lanes on a road. Edelkamp et al. [9] used hierarchical clustering algorithms to cluster DGPS trajectories, treating the number of clusters as the number of lanes and using the cluster center as the center of the lane. Wagstaff et al. [10] proposed to use the K-means clustering algorithm to cluster high-precision DGPS trajectories, analyze the number of lanes from the results, and obtain lane boundaries. Liu [20] used the high-precision trajectory data obtained by the measurement vehicle to construct a detailed structure of urban intersections and generate urban traffic networks. Although the use of high-quality GPS data to extract lane information has the advantages of high quality and reliable accuracy, its data coverage in time and space is limited and cannot describe the changes in the whole urban road information.

Li and Huang [13] proposed a GPS global map-matching method that combines the restrictions of urban road network traffic with the geometric connectivity of roads to extract road information based on FCD. Uduwaragoda et al. [21] proposed a method that statistically mines GPS trajectory data obtained from vehicles to generate a lane-level digital map of roads that is independent of lane width and lane parallelism and can handle lane splits and merges. Tang et al. [22] analyzed the characteristics of FCD and used the density clustering method based on Delaunay triangulation to optimize the data and built a naive Bayes classifier by detecting the coverage width of the floating car data and its distribution in the cross section of a road. The naive Bayes classification method was used to determine the number of lanes in the target road section. Subsequently, Tang et al. [23] further used the constrained Gaussian mixture model to simulate the distribution of FCD on the road surface, compared the advantages and disadvantages of the model under different Gaussian component combinations, and selected the number of Gaussian components corresponding to the optimal model for the number of lanes. This improved the overall extraction accuracy of the number of lanes in the experiment to 82.3%. Wang et al. [24] used high-sampling-frequency FCD to extract lane information from urban road network data and focused on a method for detecting intersections in complex road networks from low-precision FCD. Tang et al. [17] identified the turning point pairs in the low-frequency FCD trajectory data, identified intersections by spatial clustering of distances and angles, and used the range of the intersection and turning point pairs to extract urban road network intersections and their detailed structure. The above methods all used FCD to achieve rapid acquisition of information on urban road intersections and road lanes. However, there is not yet any detailed study on the identification of an auxiliary through lane at intersections. The next section of this paper proposes a technical process using FCD to detect auxiliary through lanes at intersections.

3. Detection Method

A specific data processing and calculation process for the detection of the intersection auxiliary through lane from FCD is established. First, the spatial distribution characteristics of the FCD on the lanes at the intersection are analyzed based on the four characteristic parameters of the distribution of FCD for GBDT classification. These characteristics are: (1) the road coverage width of the FCD, the distribution status of the FCD in a road cross section, the direction angle of the floating car front, and the speed of the floating car; (2) the mapping relationship between the FCD and the road network; (3) the time cycles of the FCD; and (4) the structure of road intersections. Second, the FCD is segmented, and the improved amplitude-limiting average filtering method is used to eliminate gross errors. Concurrently, the Eigenvalues required for GBDT classification, such as standardized datasets, are acquired. Finally, the auxiliary through lane at the intersection is identified through a GBDT-based classification method. The technical process of the detection method is shown in Figure 1.

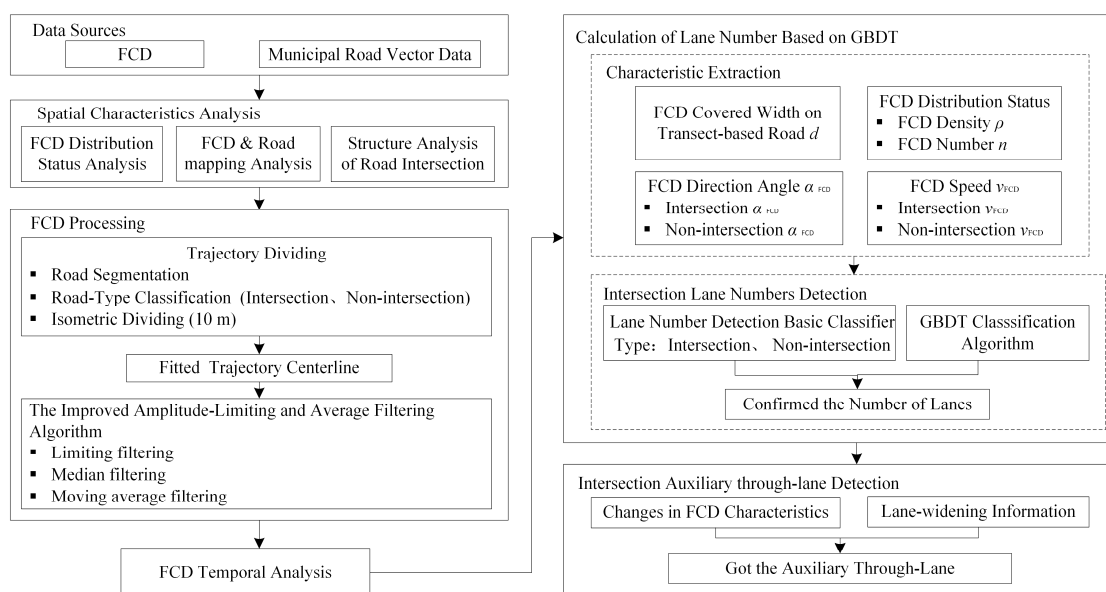


Figure 1. The technical process of detection method. FCD: floating car data; GBDT: gradient-boosted decision tree.

3.1. Analysis of the Spatial Distribution Characteristics of FCD

At present, auxiliary through lanes are established at the upstream and downstream of intersections [1]. However, floating car trajectories have different characteristics in different road sections due to road traffic regulations. According to existing intersection design and construction specifications [19], roads are divided into three sections: intersection widening section, gradient section, and road middle section. The spatial distribution characteristics of the floating car trajectory in different sections of the road are mainly reflected by the mapping relationship between the FCD and road network, the influence of road constraints, the situation of trajectory changes, and the distribution of the FCD at the road intersection, as shown in Figure 2.

The GPS coordinates of the floating car are continually updated at an observation frequency of approximately 40 s; thus, the data completely encompasses all the lanes of the road middle sections and intersections in the city, as shown in Figure 2a. The distribution of the floating car trajectory on the road surface can reflect the situation of road lanes; thus, the change in the lanes can be observed from the road gradient section and compared with that of the middle section of the road. Figure 2b shows that the number of lanes at the intersection has increased. The trajectory data are unclear on the boundary lines of each lane of the road. Owing to the driving habit of selecting the middle lane and the normal distribution of GPS

errors, the density of the trajectory data at the centerline of the road is the highest on the road cross section and decreases toward both sides of the road, as shown in Figure 2c.

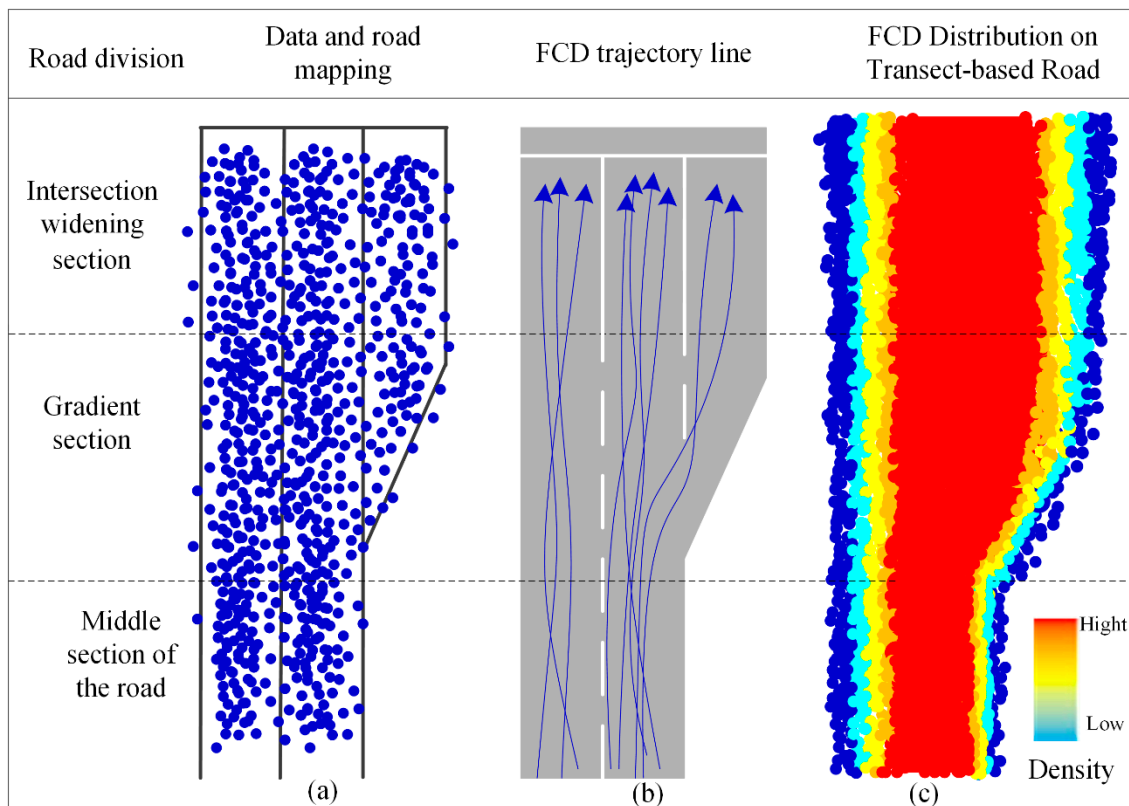


Figure 2. FCD spatial distribution diagram (a) data and road mapping; (b) FCD trajectory line; (c) FCD distribution on transect-based road.

3.2. FCD Preprocessing

The data preprocessing methods in References [18,19] do not consider the different FCD spatial distribution characteristics caused by different driving behaviors at intersections and non-intersection road sections and only divide the FCD equidistantly along the driving direction. The Delaunay triangulation method is used to remove drift coordinates, and the FCD coverage width is then calculated. This processing method can eliminate the abnormal distribution coordinate points of the FCD on the cross section of the middle section of the road. However, the presence of auxiliary through lanes at some intersections, such as in the road cross section near the intersection, increases FCD coverage width and decreases FCD distribution density compared with that in the middle section of the road. The Delaunay triangulation method is unsuitable for the FCD covered in the area near the road intersection. Thus, the authors used the improved amplitude-limiting average filtering method [25,26] to eliminate gross errors and obtain relevant attributes, such as trace coverage width and density distribution, from the FCD queue, thereby reducing data retrieval time and costs. The advantages of this method include ease of operation and efficiency of use.

The authors preprocessed the original FCD after analyzing the improved amplitude-limiting average filtering method and the FCD spatial distribution characteristics at the intersection with the auxiliary through lane. Data preprocessing is mainly divided into the following three steps: (1) Referring to the vector map, two types of road areas are divided in accordance with the distance between the road node and the distance from the node. The intersection area is defined as the area within 100 m from the intersection node, and the remaining areas are considered as non-intersection areas [19]. Next, one-way equidistant basic FCD research units are obtained by dividing the FCD

covered by the road segments in each area by 10 m and combining with the direction of the GPS trajectory; (2) Principal component analysis [27] is used to obtain the main direction of the FCD segment, the centerline of the FCD covered by each road segment is fitted, and the distance from each FCD point to the centerline is then detected; (3) The improved amplitude-limiting average filtering method is used to eliminate drift points based on the maximum width of the road design [19] and FCD coverage width and density. The experience threshold of amplitude-limiting filtering of the distance d_i from the FCD point to the trajectory centerline is set to 15 m. Subsequently, the average distance \bar{d}_i between the trajectory points and the trajectory centerline in the FCD segment, the average velocity \bar{v}_i of the FCD trajectory, and the average direction angle $\bar{\alpha}_i$ are used as constraint conditions, and the improved amplitude-limiting average filtering method is used to eliminate the drift points contained in the original FCD.

3.3. FCD Temporal Analysis

As the incidence sampling rate of the low-frequency FCD used in the experiment is 40 s, in the inversion of road information, the coverage density of FCD on the road surface is sparse in a single day. It is difficult to meet the road detection requirements, thus the temporal fusion of the FCD is needed. The principle of this is to add the cleaned FCD on a day-by-day basis, and carry out superposition analysis, and calculate the width of FCD covered on the road surface after daily superposition until the coverage width does not change. This will allow the number of days to be used as the time for extracting lane information using FCD.

3.4. Auxiliary Through Lane Detection Based on GBDT

GBDT is a widely used nonlinear classification model [28–31]. Based on the principle of boosting algorithms, a new decision tree in the gradient direction is established to reduce the residual error in each iteration, and classification accuracy is improved through iteration. During actual situations, the error of locating a taxi in the city causes drift in the coordinates of the collected trajectory point. The distribution of the trajectory points on the road surface cannot directly describe the specific number of lanes, but the law that each trajectory point falls on the lane follows geometric distribution. Thus, a reasonable model can be established using trajectory point distribution characteristics to obtain the number of lanes.

3.4.1. Characteristic Selection

Four characteristic parameters of FCD distribution on the road segment must be obtained before the number of lanes can be detected. These parameters are the coverage width D of the FCD on the road cross section, the distribution density P , the direction angle of FCD A , and the velocity V .

Figure 3 uses α as the angle between the front direction of each FCD trajectory point and the main direction of the trajectory. According to the Urban Road Intersection Planning Specification [19] and the inverse trigonometric function solution, the segment is the intersection-type when $\alpha \leq 17^\circ$ and the non-intersection type when $\alpha \leq 14^\circ$. According to the general rule of vehicle driving and the influence of traffic signals, the driving speed of each FCD trajectory point v is >0 when the segment is the non-intersection-type and ≥ 0 when the segment is the intersection-type. Regarding the j th FCD segment, d_{up} is the distance from the centerline of the FCD point above the trajectory centerline; d_{down} is the distance from the centerline of the FCD point below the trajectory centerline; $D_i = \max|d_{up}| + \max|d_{down}|$ is the coverage width of the FCD on the road cross section; P is the trajectory point density in each unit width, i.e. the trajectory centerline of the segment equidistantly divides 100 units by the distance d' , the ratio of the number n of trajectory points in each segmentation unit to the total number N in the segment, where $d' = D_{max}/100$.

Type	Non-intersection	Intersection	
	Middle Section	Gradient section	Widening Section
	$\alpha \leq \alpha_{\max}=14^\circ$ $v > 0 \text{ m/s}$	$\alpha \leq \alpha_{\max}=17^\circ$ $v \geq v_{\min} = 0 \text{ m/s}$	
	$D = d_{\text{up}} + d_{\text{down}}$ $P = n_d / N$	$D = d_{\text{up}} + d_{\text{down}}$ $d_{i-1} \leq d_i \leq d_{i+1}$ $P = n_d / N$	

Figure 3. Description of characteristic parameters.

3.4.2. Basic Classifier Construction

GBDT is a supervised learning algorithm. When GBDT is used to classify FCD samples, the number of classifications must be given. The road segments with a known number of lanes (such as class labels) and FCD spatial characteristics $\{D, P, A, V\}$ are used as training samples to establish a classifier containing the corresponding relationship between the number of lanes and the spatial characteristics of FCD (Table 2) for convenience and efficiency. According to the FCD spatial characteristics of the segment and the basic classifier, the possible number of lanes of the sample to be tested (the segment number of the unknown lane) is estimated and then the GBDT model is used to confirm the final number of lanes. The GPS error obeys a normal distribution [19] and follows the $3\sigma^2$ principle. The widths of the FCD under σ^2 , $2\sigma^2$, and $3\sigma^2$ are selected as the Eigenvalues of the FCD coverage width on the road cross section. The geometric structure and traffic characteristics of the road intersection are different from those of the middle of the road; thus, the basic classifier of the number of lanes based on the probability density of FCD is divided into two categories—intersection-type and non-intersection-type—to ensure the representativeness of the sample data.

Table 2. Basic classifier based on floating car data (FCD) probability density for lane number detection in the experiential area.

Intersection/ Non-Intersection	FCD Coverage Width (m)			Density %	Direction Angle	Speed	Number of Training Samples
	σ^2	$2\sigma^2$	$3\sigma^2$	p	A	V	N
One-way lane	D_1^1	D_1^2	D_1^3	P_1	A_1	V_1	N_1
One-way two lanes	D_2^1	D_2^2	D_2^3	P_2	A_2	V_2	N_2
One-way three lanes	D_3^1	D_3^2	D_3^3	P_3	A_3	V_3	N_3
One-way four lanes	D_4^1	D_4^2	D_4^3	P_4	A_4	V_4	N_4
One-way five lanes	D_5^1	D_5^2	D_5^3	P_5	A_5	V_5	N_5

3.4.3. Lane Acquisition Based on GBDT Classification

The FCD segment with a known number of lanes is used as the training sample, and the GBDT model is trained in accordance with the FCD spatial characteristics of the training sample to divide the sample effectively and allow its measurement. Finally, the number of lanes of the sample to be measured is determined.

The training sample X contains n FCD trajectory points: $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. The road surface with k lanes is covered with n points, $x_i = \{d_i, \rho_i, \alpha_i, v_i\}$ is the i th trajectory point characteristic, and y_i is the type of lane where the i th point is located (for example, Lane 1). The loss function is defined as $L[y, F(x)]$, $F(x)$ is the lane type identified by GBT model [28,31], m is the number of iterations (such as m trees), and k is the number of categories.

The 0th iteration (initialization) of the GBT model, $F^{(0)}(x)$, is a fixed value with the following formula:

$$F^{(0)}(x_i) = \operatorname{argmin}_{\gamma} \sum_{i=1}^m L(y_i, \gamma) \quad (1)$$

where x is the FCD spatial characteristic, γ is the lane type initialized by this iteration, and $h^{(m-1)}(x)$ is the m th decision tree divided from the m th iteration. The best residual fitted value for the division result is $F^{(m)}(x)$, the fitted value for the division result of $m-1$ th decision tree is $F^{(m-1)}(x)$, residual γ_{ik} is:

$$\gamma_{ik} = - \left[\frac{\partial L(y_i, F^{(m-1)}(x_i))}{\partial F^{(m-1)}(x_i)} \right] \quad (2)$$

A new training set $\{(x_i, \gamma_{ik})\} (i = 1, 2, \dots, n)$ is obtained from Formula (2). When the $m + 1$ th decision tree $h^{(m)}(x)$ is trained, the corresponding leaf node region is $R_{ik}, k = 1, 2, K$, and the best fitting value of the leaf region $c_{ik}^{(m)}$ is calculated.

$$c_{ik}^{(m)} = \operatorname{argmin}_{\gamma} \sum_{i=1}^m L[(y_i, F^{(m-1)}(x_i) + c_{ik}^{(m)})] \quad (3)$$

where $x_i \in R_{ik}$, The m th updated model obtained from decision trees $h^{(m)}(x)$ and $c_{ik}^{(m)}$ is:

$$F^{(m)}(x) = F^{(m-1)}(x) + c_{ik}^{(m)} h^{(m)}(x) \quad (4)$$

The final GBT model $F^{(m)}(x)$ is obtained after m (the maximum number of) iterations. The FCD of the known number of lanes is selected as the training sample, then the corresponding GBDT models are obtained for different classification numbers (such as the number of lanes, k). The number of lanes of the road segment to be measured is obtained by estimating the candidate number of lanes belonging to the road segment by comparing the FCD characteristics of the segment and the basic classifier $\{D, P, A, V\}$ (for example, the segment may have two or three lanes). The FCD in the segment is divided into different numbers of lanes using the GBDT models of different classification combinations and the FCD in each segment can provide multiple lane number division results. Based on the statistics of the different classification results and the model entropy [8,30], the authors selected the number of classifications for which the distribution of FCD on each lane is relatively uniform and the model entropy is as small as the number of lanes in the segment.

3.5. Intersection Auxiliary Through Lane Detection

Auxiliary through lane acquisition involves two processes: lane number acquisition and auxiliary through lane detection. Regarding auxiliary through lane detection, the variations in the number of lanes across different road segments are analyzed. Lane numbers in the same types of segments and different types of segments also are analyzed comparatively.

3.5.1. Analysis of Changes in FCD Characteristics

Owing to the differences in the units of measurement of the four characteristic parameters of the FCD, minimum-maximum normalization must be used to nondimensionalize each characteristic parameter for the analysis of FCD space distribution characteristics at intersections and non-intersections in the same value range (between 0 and 1).

A road is divided into n segments where the first i segments are denoted by $\{X_1, X_2, \dots, X_i\}$, and the characteristic variables of the segment are normalized, then: $X_i = (X_i - X_{\min}) / (X_{\max} - X_{\min})$.

Considering the continuous segmentation of an entire road, the change trend of the four characteristic parameters of the FCD is analyzed. Taken from the intersection widening section to the middle section of the road, the density of the trajectory points of each lane decreases, then the trajectory coverage width increases, as shown in Figure 4a,b. The angle A of the direction of the FCD in the gradient segment shows a trend of decreasing first and then increasing, as shown in Figure 4c. Regarding the speed characteristic V of the FCD, the speed from the middle section to the intersection division is gradually reduced. Following normalization, the speed gradually approaches 0 at the intersection portion, as shown in Figure 4d.

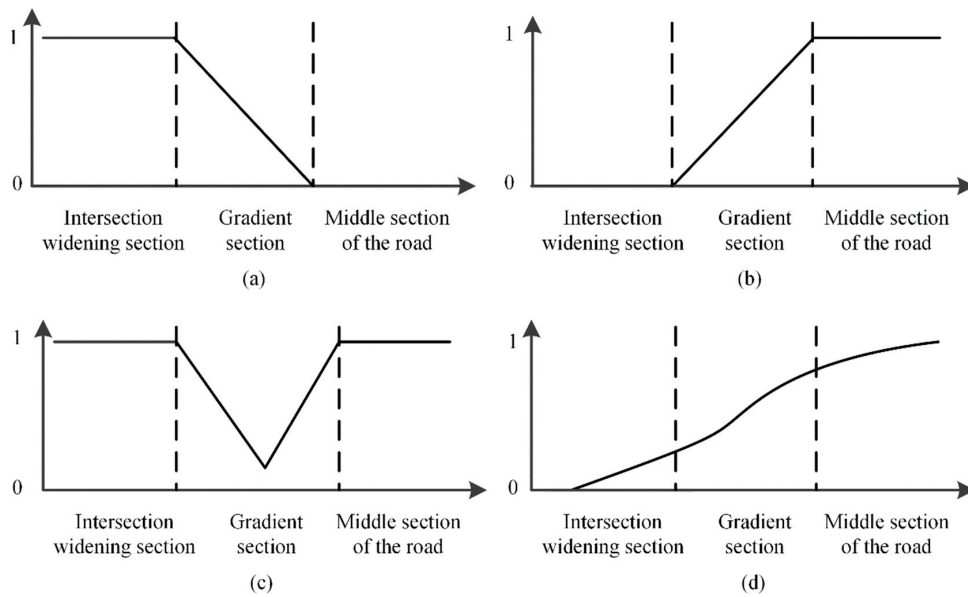


Figure 4. FCD Eigenvector normalized value (a) coverage width (b) density (c) direction (d) speed.

3.5.2. Analysis of Lane-Widening Information

The number of road lanes is obtained through GBDT classification, and the numbers of lanes in different areas (intersections/non-intersections) on the same road are compared.

Generally, the distribution of the FCD on each lane of a one-way, two-lane section is approximately 50% and that of a one-way, three-lane section is approximately 33%. Variations in the lane traffic volume result in the variation in the number of floating cars distributed on each lane of a road with different numbers of lanes, but the FCD distribution continues to float around a fixed value. Figure 5 demonstrates that on the same road, the number of lanes in the middle section of the road is three lanes, and the FCD distribution on each lane is approximately 33%. Where the intersection widens, the number of lanes obtained is four, and the FCD distribution is approximately 25%. Regarding the gradient section, the number of lanes is between three and four but is closer to four. Lane widening at the intersection can be detected based on the change in the number of lanes when another lane is added to the intersection.

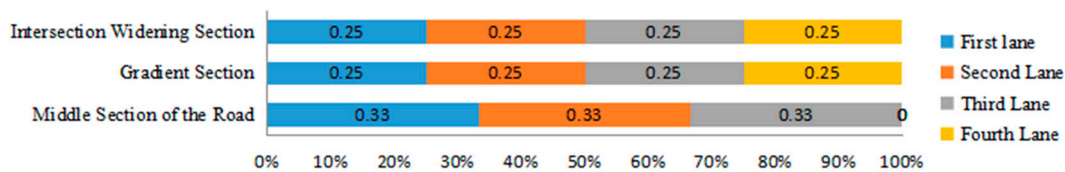


Figure 5. Percentage of FCD per lane.

4. Experimental Test

4.1. Study Area and Data Sources

Twenty-one plane intersections of four typical roads in Wuchang District, Wuhan City, China were used as training samples. The trajectory data were taxi GPS trajectory data generated by taxis in Wuhan City over the period of 3 July 2013 to 17 July 2013 (a total of 15 days). The trajectory data included vehicle identification, observation time, GPS latitude and longitude, and other information. The trajectory point sampling frequency was 40 s and collected additional data once when taxiing on or off. The road network data were the 2012 Wuhan standard road vector data. The coordinate system of the data source was WGS-84.

The authors used SPSS software for data timing fusion and illegal value rejection, ArcGIS software for data projection transformation, and Matlab software for data-filtering calculations and FCD coverage width detection during the period. The Python with SciKit-Learn package interface was used for GBDT classification.

4.2. Data Processing

Given that the general positioning accuracy of vehicular GPS equipment is not high, the effectiveness of trajectory data preprocessing directly affects the detected number of intersection lanes. The specific steps for preprocessing experimental data were as follows: (1) During data fusion, the data for a certain period (7 days/10 days, denoted as 7 d/10 d) were cumulatively superimposed, and illegal points in the data were eliminated; (2) Regarding road-network matching, the coordinates of the trajectory sampling point were converted from the WGS-84 coordinate system into the same coordinate system as the road network, so that they fall on the correct driving lane; (3) For trajectory segmentation and data standardization, the FCD in the segment was converted to a rectangular coordinate through Mercator transformation to establish mathematical models and derive quantitative statistics and calculations; (4) During data cleaning, the cleaned data were summarized and homogenized to obtain the standardized data set of the overall FCD in the segment $\{D, P, A, V\}$. The data processing results are shown in Figure 6.

The experimental results showed the following:

The trajectory data at the intersections and non-intersections for different traffic characteristics and different driving behaviors (mainly considering roadside parking and normal parking under the control of traffic lights) were cleaned using the improved amplitude-limiting average filtering method. When the data-filtering rates were 46.51–61.37% and 58.24–68.72% for the intersection area and non-intersection areas, respectively, FCD could satisfactorily detect the number of road lanes.

4.3. Analysis of the Sampling Period of Floating Cars

The data used here had low density given the sampling frequency of FCD was 40 s. Furthermore, FCD experienced difficulty meeting road detection requirements on one of the experiment days, so FCD had to be accumulated daily until the detected road width remained unchanged. A comparison of the detection results of the coverage widths of the trajectory data at different periods of distinct trajectory segments showed that the trajectory data coverage width of the middle segment of the road remained unchanged after seven days. This result is consistent with those of previous studies [10,11]. However, the width of trajectory data coverage at intersections did not change after 10 days, as shown in Figure 7. Therefore, when using the current method to detect the lane number at a road intersection, a data collection period of 10 days is reasonable.

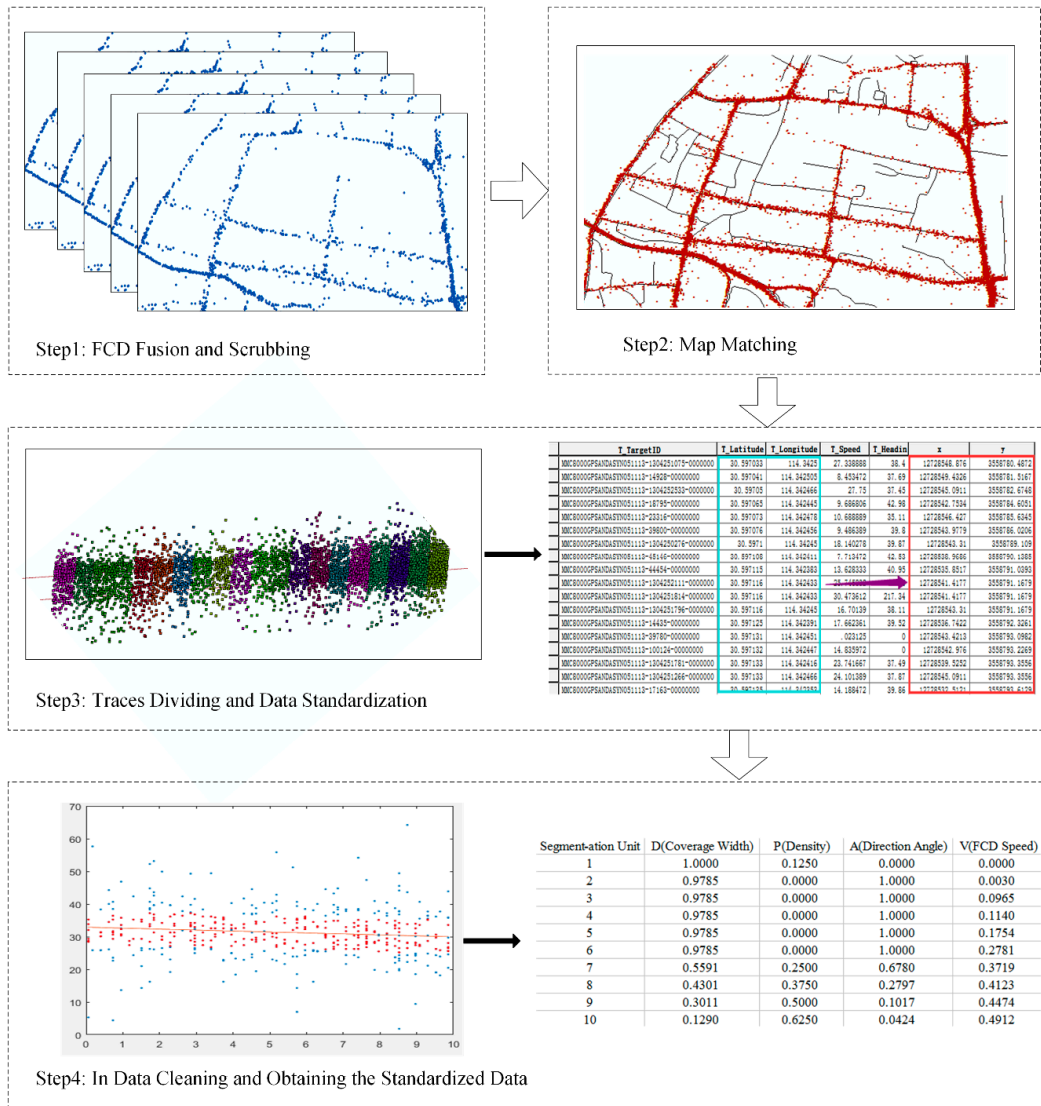
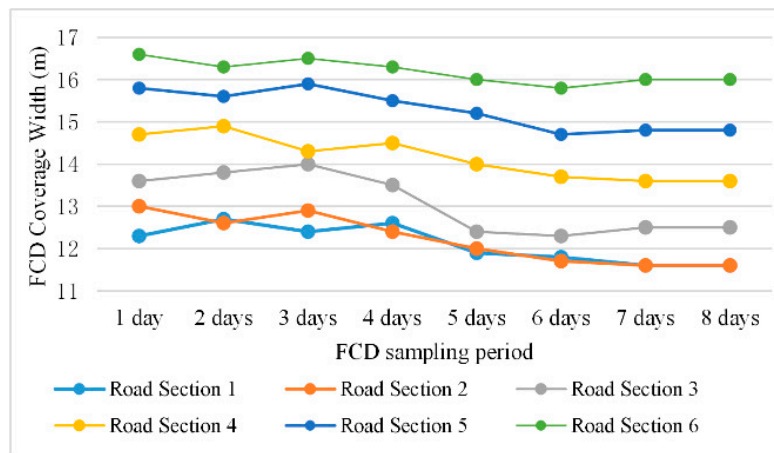


Figure 6. Experimental data processing flow.



(a)

Figure 7. Cont.

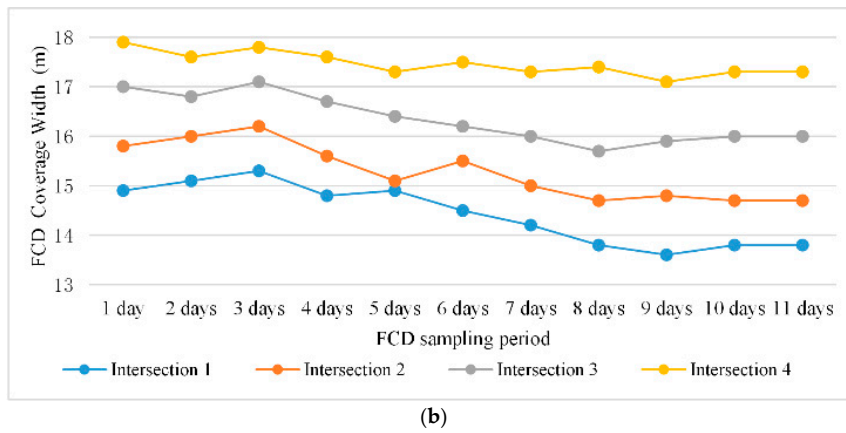


Figure 7. Results for different collection periods for FCD in the (a) middle segment of the road and (b) at the road intersection.

4.4. Auxiliary Through Lane Detection and Discussion

4.4.1. FCD Characteristic Variables Transformation Analysis

A road was divided into n segments where the first 20 segments were denoted by $\{X_1, X_2, \dots, X_n\}$ and the segment characteristic variables $X_i = \{D_i, P_i, A_i, V_i\}$ were normalized. The normalized calculation was $x'_i = (x_i - x_{imin}) / (x_{imax} - x_{imin})$, where $x_i \in X_i$. The authors found that the density of trajectory points in each lane decreased with the increment in the coverage width of the trajectory points at the intersection, as shown in Figure 8.

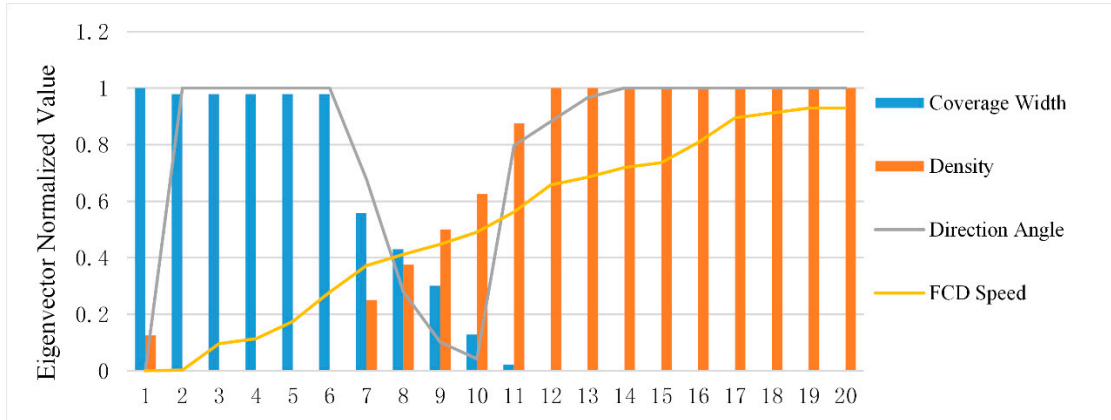


Figure 8. FCD Eigenvector normalized value.

Sections 1–6 in the actual road constituted the intersection widening segment, Sections 7–10 constituted the intersection gradient unit, and the remaining sections constituted the middle portion. Regarding Sections 1–20, the normalized values of FCD coverage width tended to decrease, whereas the normalized values of the FCD coverage density and average velocity tended to increase (Figure 4). These results were consistent with the actual situation.

4.4.2. Lane Number Detection in the FCD Experimental Area

The FCD coverage width and its distribution (such as normal distribution) on the cross section of the road of Wuchang District were detected. The experimental detection results and the number of lanes corresponding to the experimental road were combined to construct a basic classifier for the number of lanes in the FCD experimental area, as shown in Table 3.

Table 3. Results of the basic classifier based on lane number detection in the FCD experimental area.

Intersection/ Non-Intersection	Lane Number	FCD Coverage Width /m			Number of Training Samples
		σ^2	$2\sigma^2$	$3\sigma^2$	<i>N</i>
One-way lane	1	5.42/4.89	6.73/5.94	8.32/7.23	45/67
One-way two lanes	2	8.30/7.64	12.13/11.24	12.97/11.93	76/95
One-way three lanes	3	10.98/9.66	18.72/16.63	19.08/17.12	102/165
One-way four lanes	4	13.45/13.24	23.63/21.78	25.49/23.97	106/189
One-way five lanes	5	16.46/15.57	26.22/24.29	28.51/24.83	52/82

The authors randomly selected a segment X_4 of a certain sample X' for measurement. Its FCD distribution width was 15.2 m. The segment was a road intersection. The basic classifier of the number of lanes in the experimental area showed that the selected segment might have belonged to a one-way, three-lane section or a one-way, four-lane section. The parameters of the GBDT model when $k = 3$ and $k = 4$ were calculated.

The calculation results showed that the model entropy was $F(3) = 0.0073$ when $k = 3$ and that the model entropy was $F(4) = 0.051$ when $k = 4$. Thus, the model is optimal when $k = 3$. Figure 8 shows that when four-lane segmentation was adopted for segment X_4 , the proportions of the FCD on the fourth lane were considerably lower than 25%, whereas the three-lane segmentation was closer to the actual situation. Therefore, segment X_4 was a one-way segment with three lanes. Subsequently, the remaining segments were calculated individually following the same calculation procedure, and the test results are shown in Table 4.

Table 4. Ultimate results for the objective road (intersection section).

Intersection Section	Segmentation Unit	FCD Coverage Width (m)	FCD Density	Direction Angle	Average Speed	Candidate Value	Classification Results
X'	X_1	17.85	0.25	0	0	4/5	4
	X_2	17.75	0.25	290.8	3.1	4/5	4

	X_7	15.8	0.29	279.3	15.5	3/4	4
	X_8	15.2	0.30	287.4	18.7	3/4	3
	X_9	14.6	0.29	280.2	8.2	3/4	3
	X_{10}	13.8	0.3	279.5	8.7	3/4	3

Regarding Table 5, the non-intersection section was the middle section of the road. Given that the scale of the intersection area was not unique in existing intersection design specifications, the inaccurate first segment should be removed. Thus, segment X_1 was used only for reference. The coverage width of the FCD was 13.2 m from segment X_2 to segment X_{16} , and the coverage width of the trajectory point in each segment was relatively close. The classification results obtained by the current method indicated a one-way, three-lane section. Considering the target intersection X' , the trajectory point coverage widths and candidate results of segments X_1 , X_2 , X_3 , X_4 , and X_5 were analyzed, and the number of lanes at each intersection segment were detected. The classification results showed that intersection segments X_1 , X_2 , and X_3 were one-way and had four lanes, whereas segments X_4 and X_5 were one-way and had three lanes. An analysis of the classification results for intersection X' showed that the lane detection results for each intersection section were inconsistent. A comparison of the distribution characteristics of the FCD in each segment (Figure 5) demonstrated that the data for the FCD distribution width of the intersection were discrete and the variation in the trajectory point coverage width gradually decreased. Combined with the result of lane classification, this result suggested that the road at the intersection had been widened, that is, a lane had been added.

Table 5. Ultimate results for the objective road (non-intersection section).

Non-Intersection Section	Segmentation Unit	FCD Coverage Width (m)	FCD Density	Direction Angle	Average Speed	Candidate Value	Classification Results
X	X ₁	13.3	0.32	288.4	9.5	3/4	3
	X _{x2}	13.2	0.33	289.4	10.6	3/4	3

	X ₁₆	13.2	0.33	290.8	13.7	3	3

Furthermore, the actual image corresponding to section x and intersection X' verified that a lane had been added at the intersection but not at the middle road section. Therefore, the results obtained by dividing floating car trajectory at non-intersection X and intersection X' of the road (Figure 9) showed that the width of FCD coverage increased and the distribution density decreased. The actual image of the road was the street view (Figure 10).

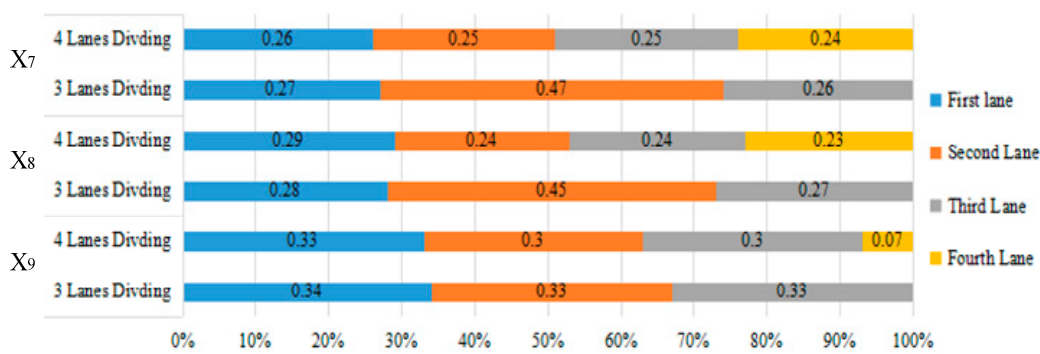


Figure 9. Different division strategies corresponding to the number of lanes.

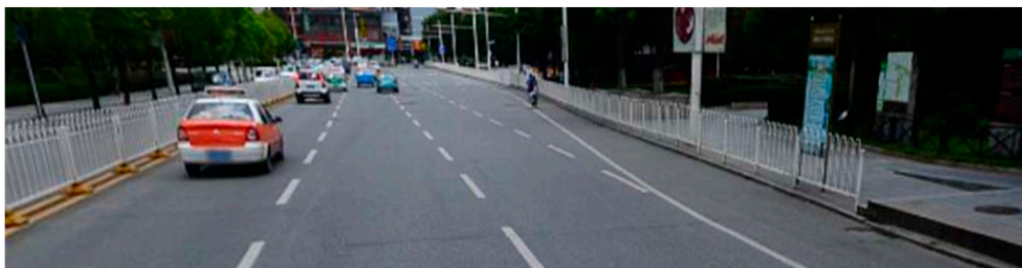


Figure 10. Corresponding image of road segment X and intersection segment X'.

4.5. Quantitative Analysis of Intersection Lane Detection

The accuracy of the FCD-based method for detecting the number of intersection lanes was studied intensively, beginning with the unique geometric structure of the intersection. The number of lanes at intersections and non-intersections were extracted using three methods. Actual images were used to calculate the accuracy of each method. The results are shown in Table 6.

The overall accuracy of the proposed method for detecting the number of lanes at intersections was 81.86%, and its overall accuracy for detecting the number of lanes at non-intersections was 83.89%. The detection accuracy of the current model was superior to those of the two other methods. The analysis results showed that (1) the road was divided into intersection areas and non-intersection areas. The GBDT classification model was constructed based on the spatial characteristics of the FCD in different research areas, thereby improving the FCD simulation results of the road sections to be measured and the accuracy of detecting the number of intersection lanes. (2) The number of lanes exhibited a large deviation when a single variable (*d*) assessment was used in the intersection gradient section. This deviation was reduced by adding the attributes *v* and *α* of the GPS trajectory data to the

FCD as classification characteristics. Consequently, the overall accuracy for detecting the number of road lanes was improved.

Table 6. Quantitative evaluation of lane number detection.

Method	Type	Number of Lanes Corresponding to Different Prediction Accuracy (%)					Overall Accuracy (%)
		1	2	3	4	5	
Current method	Intersection	82.63	81.57	80.21	81.42	83.47	81.86
	Non-intersection	85.54	83.41	82.71	82.56	85.27	83.89
Naïve Bayesian Classification [22]	Intersection	76.63	75.37	72.44	73.26	74.12	74.36
	Non-intersection	79.89	78.43	76.80	74.21	75.42	76.95
Constrained Gaussian Mixture Model [23]	Intersection	76.63	74.57	72.21	73.87	75.35	74.53
	Non-intersection	84.31	82.71	79.56	81.49	82.54	82.12

5. Conclusions

This study analyzed the shortcomings of existing domestic and international detection methods for intersection auxiliary through lanes. Moreover, it applied FCD to establish a new method for detecting the change in the number of lanes at road intersections. The trajectory data were preprocessed through the improved amplitude-limiting average filtering method among other methods. The GBDT classification method was used to detect the number of lanes in each road area and to subsequently infer whether an auxiliary through lane existed at the intersection. This method was tested using information for a road in Wuchang District, Wuhan City. The experimental results showed that the proposed method can rapidly obtain lane widening information from low-precision FCD, and the validity of the method was verified.

The proposed method presents several advantages. First, this method has a shorter data update cycle time and lower cost compared to traditional methods for lane information detection. Second, the ability of the method to detect the number of lanes at intersections has been intensively studied, beginning with the unique geometric structure of the intersection. The method can detect the change in lane numbers and enables the rapid updating of navigation data. Third, the proposed method has improved accuracy relative to that of existing associated methods.

However, the proposed method presents two limitations. First, its calculation accuracy is not particularly high. Future works will attempt to improve the calculation accuracy of the method by integrating deep learning algorithms or other data sources. Second, the data processing and analysis process of the method involves numerous steps; the software operation is cumbersome, and the calculation speed is slow. Future works will attempt to reduce the number of steps and increase the speed of calculation of the proposed method through the integration of automation tools.

Author Contributions: Conceptualization, X.L.; Methodology, Y.W.; Software, Y.T.; Validation, J.W. and Y.W.; Formal Analysis, Y.W.; Resources, X.L.; Writing-Original Draft Preparation, Y.W.; Writing-Review & Editing, X.L.; Supervision, P.C.; Funding Acquisition, P.C.

Funding: This research was funded by National Key Research and Development Program of China (No. 2016YFB0502600), the Key laboratory of watershed ecology and geographical environment monitoring, NASG (No. WE2015007), National Natural Science Foundations of China (No. 41501437, 41601416), Jiangxi Province Science Foundation for Youths (No. 20161BAB213092), Science Foundation of the Educational Committee of Jiangxi Province (No. GJJ160571).

Acknowledgments: We would like to thank Huayi Wu and Longgang Xiang of Wuhan University for the floating car data provided for the experiment. At the same time, We would like to thank Luliang Tang of Wuhan University for his preliminary research work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ma, Y.; Gao, Y.; Leng, X.; Li, Z. Traffic operation characteristics of auxiliary through lane at signalized intersection. *J. Harbin Inst. Technol.* **2015**, *47*, 42–45. [[CrossRef](#)]
2. Guo, W. Optimization Method of Signal Control Adapting to Stretching-Segment Length. Ph.D. Thesis, Jilin University, Changchun, China, 2011.
3. Yang, W.; Ai, T. A Method for Road Network Updating Based on Vehicle Trajectory Big Data. *J. Comput. Res. Dev.* **2016**, *53*, 2681–2693. [[CrossRef](#)]
4. Tarawneh, M.S.; Tarawneh, T.M. Effect on utilization of auxiliary through lanes of downstream right-turn volume. *J. Transp. Eng.* **2002**, *128*, 458–464. [[CrossRef](#)]
5. Bugg, Z.; Roupail, N.M.; Schroeder, B. Lane choice model for signalized intersections with an auxiliary through lane. *J. Transp. Eng.* **2012**, *139*, 371–378. [[CrossRef](#)]
6. Moriyama, Y.; Mitsuhashi, M.; Hirai, S.; Oguchi, T. The effect on lane utilization and traffic capacity of adding an auxiliary lane. *Procedia-Soc. Behav. Sci.* **2011**, *16*, 37–47. [[CrossRef](#)]
7. Fang, L.; Yang, B. Automated extracting structural roads from mobile laser scanning point clouds. *Acta Geod. Cartogr. Sin.* **2013**, *42*, 261–267.
8. Chen, Y.; Krumm, J. Probabilistic modeling of traffic lanes from GPS traces. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 81–88. [[CrossRef](#)]
9. Edelkamp, S.; Schrödl, S. Route planning and map inference with global positioning traces. In *Computer Science in Perspective*; Klein, R., Six, H.W., Wegner, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 128–151.
10. Wagstaff, K.; Cardie, C.; Rogers, S.; Schrödl, S. Constrained k-means clustering with background knowledge. In Proceedings of the Eighteenth International Conference on Machine Learning, San Francisco, CA, USA, 28 June–1 July 2001; pp. 577–584. [[CrossRef](#)]
11. Knoop, V.L.; Bakker, P.F.D.; Tiberius, C.C.J.M.; Arem, B.V. Lane Determination with GPS Precise Point Positioning. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2503–2513. [[CrossRef](#)]
12. Zheng, Y.; Zhou, X. *Computing with Spatial Trajectories*, 1st ed.; Springer: New York, NY, USA, 2011.
13. Li, Q.; Huang, L. A Map Matching Algorithm for GPS Tracking Data. *Acta Geod. Cartogr. Sin.* **2010**, *39*, 207–212.
14. Chen, Y. Research on Urban Road Network Overpass Recognition Technology Based on GPS Data. Master's Thesis, Jilin University, Changchun, China, 2011.
15. Zheng, N.; Lu, F.; Li, Q.; Duan, Y. The Adaption of A* Algorithm for Least-time Paths in Time-dependent Transportation Networks with Turn Delays. *Acta Geod. Cartogr. Sin.* **2010**, *39*, 534–539.
16. Tang, L.; Huang, F.; Zhang, X.; Xu, H. Road Network Change Detection Based on Floating Car Data. *J. Netw.* **2012**, *7*, 1063–1070. [[CrossRef](#)]
17. Tang, L.; Niu, L.; Yang, X.; Zhang, X.; Li, Q.; Xiao, S. Urban Intersection Recognition and Construction Based on Big Trace Data. *Acta Geod. Cartogr. Sin.* **2017**, *46*, 770–779. [[CrossRef](#)]
18. Yang, X.; Tang, L.; Stewart, K.; Dong, Z.; Zhang, X.; Li, Q. Automatic Change Detection in Lane-level Road Networks Using GPS Trajectory. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 601–621. [[CrossRef](#)]
19. Yang, P.; Li, K.; Zhao, J.; Chen, X.; Liu, Y.; Zhu, Z.; Wang, X.; Zheng, L.; Lin, Q.; Teng, S.; et al. *GB50647-2011, Code for Planning of Intersections on Urban Roads*; Chinese Standard; China Planning Press: Beijing, China, 2011.
20. Liu, J.; Cai, B.; Wang, Y.; Wang, J. Generating enhanced intersection maps for lane level vehicle positioning based applications. *Procedia-Soc. Behav. Sci.* **2013**, *96*, 2395–2403. [[CrossRef](#)]
21. Uduwaragoda, E.R.I.A.C.; Perera, A.S.; Dias, S.A.D. Generating lane level road data from vehicle trajectories using Kernel Density Estimation. In Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), Hague, The Netherlands, 6–9 October 2014; pp. 384–391. [[CrossRef](#)]
22. Tang, L.; Yang, X.; Kan, Z.; Wang, X.; Li, Q.; Shaw, S. Traffic Lane Number Detection on Based on the Naive Bayesian Classification. *China J. Highw. Transp.* **2016**, *29*, 116–123.
23. Tang, L.; Yang, X.; Jin, C.; Liu, Z.; Li, Q. Traffic Lane Number Extraction Based on the Constrained Gaussian Mixture Model. *Geom. Inf. Sci. Wuhan Univ.* **2017**, *42*, 341–347. [[CrossRef](#)]
24. Wang, J.; Rui, X.; Song, X.; Tan, X.; Wang, C.; Raghavan, V. A novel approach for generating routable road maps from vehicle GPS traces. *Int. J. Geogr. Inf. Syst.* **2015**, *29*, 69–91. [[CrossRef](#)]

25. Wen, C.; Gao, L.; Fang, J.; Fang, J.; Ju, Y.; Li, Y. The High-Precision Weighing System Based on the Improved Amplitude-Limiting and Average Filtering Algorithm. *Chin. J. Sens. Actuators* **2014**, *27*, 649–653. [[CrossRef](#)]
26. Kweon, S.J.; Shin, S.H.; Yoo, H.J. High-order temporal moving-average filter using a multi-transconductance amplifier. *Electron. Lett.* **2012**, *48*, 961–962. [[CrossRef](#)]
27. Yang, B.; Dong, Z. A shape-based segmentation method for mobile laser scanning point clouds. *ISPRS J. Photogr. Remote Sens.* **2013**, *81*, 19–30. [[CrossRef](#)]
28. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
29. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
30. Kong, X.; Xia, F.; Wang, J.; Rahim, A.; Das, S.K. Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data. *IEEE Trans. Ind. Inform.* **2017**, *13*, 1201–1212. [[CrossRef](#)]
31. Wang, S.; Liu, T. Gradient Boosting Decision Tree Method for Residential Load Classification Considering Typical Power Consumption Modes. *Proc. CSU-EPSC* **2017**, *29*, 27–33. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).