A Vehicle Matching Algorithm by Maximizing Travel Time Probability Based on Automatic License Plate Recognition Data

Chunguang He, Dianhai Wang, Zhengyi Cai, *Member, IEEE*, Jiaqi Zeng, and Fengjie Fu

***Abstract*— Vehicle re-identification aims to match and identify the same vehicle crossing multiple surveillance cameras and obtain traffic information such as travel time. The Automatic License Plate Recognition (ALPR) data are widely employed in urban surveillance. However, vehicle re-identification based on ALPR data is challenging due to license plate recognition errors and unrecognized issues. This paper proposes a vehicle matching algorithm designed to maximize the travel time probability using ALPR data, while accounting for recognition errors and unrecognized issues. The proposed algorithm consists of several modules, including the estimation of travel time distribution, computation of travel time probability, calculation of travel time confidence intervals and matching time window size, restricted fuzzy matching, and vehicle matching optimization. To evaluate the effectiveness of the proposed algorithm across varying lighting and weather conditions, ALPR data was collected from a survey road in four scenarios: sunny day, sunny night, rainy day, and rainy night. The results indicate that when compared to a sunny day scenario, severe lighting and adverse weather conditions lead to decreased matching accuracy and increased matching accuracy errors for all methods evaluated. However, our proposed model outperforms benchmark algorithms in both scenarios, demonstrating its superior performance.**

***Index Terms*— Vehicle reidentification, vehicle matching algorithm, automatic license plate recognition (ALPR) data, travel time distribution, travel time probability.**

I. Introduction

THI Automatic License Plate Recognition (ALPR) data contains plentiful information. In recent years, researchers have widely applied ALPR data in traffic information mining by matching the upstream and downstream ALPR data, such as traffic demand estimation [1], [2], travel time estimation [3], [4], speed estimation [5], link dynamic vehicle count [6], queue length estimation [7], [8], [9], and route reconstruction [10]. The ALPR data is obtained using optical character recognition (OCR) within the captured image. In some situations, the ALPR data may suffer the recognized errors or unrecognized due to the recognition environment, leading to the upstream and downstream license plates unmatched in practice. Therefore, how to match license plate recognition errors and unrecognized vehicles is the key to data application.

To address the issue of vehicles going unmatched due to recognition errors and other issues, we propose a vehicle-matching algorithm that aims to maximize the probability of travel time. This algorithm will help in re-identifying vehicles in cases of recognition errors and other unrecognized issues in the ALPR data. The main steps of the algorithm are: travel time extract and preprocessing module; travel time distribution constructing module; travel time probability, travel time confidence interval, and matching window size calculating module; restricted fuzzy matching module. The ALPR data of two consecutive intersections of a road in Hangzhou, China covering peak and off-peak hours are used to test the proposed algorithm. We investigated the high-definition video data of the experimental road section at the same time and used it for manual identification to verify the proposed matching algorithm.

Different from the approach presented by Zhan, et al. [7], we estimate travel time based on the ALPR data and apply the optimization model to calculate the vehicle matching result that maximizes the travel time probability. The proposed method does not require additional data and traffic flow rate assumptions. There are three main contributions:

1) For the first time, the entire sample data of the road section was obtained, and the actual recognition accuracy of the ALPR equipment was calculated and analyzed.
2) The dynamic travel time distribution of the road section was estimated based on the data-driven approach, and the travel time probability was automatically calculated from the travel time distribution.
3) A matching algorithm by maximizing the travel time probability was proposed to match the unmatched upstream and downstream vehicles, which fully employs the travel time information based on the ALPR data.

Manuscript received 13 August 2022; revised 10 March 2023, 16 September 2023, and 16 December 2023; accepted 18 January 2024. Date of publication 7 February 2024; date of current version 1 August 2024. This work was supported in part by the "Pioneer" and "Leading Goose" Research and Development Program of Zhejiang under Grant 2023C01240 and Grant 2023C03155; in part by the National Natural Science Foundation of China under Grant 52131202, Grant 71901193, and Grant 52072340; and in part by the Zhejiang Province Basic Commonweal Project under Grant LGF22F030008. The Associate Editor for this article was A. Hegyi. *(Corresponding author: Zhengyi Cai.)*

Chunguang He is with the School of Transportation and Logistics Engineering, Xinjiang Agricultural University, Urumqi 830052, China, and also with the College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China (e-mail: hechunguang@zju.edu.cn).

Dianhai Wang and Jiaqi Zeng are with the College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China (e-mail: wangdianhai@zju.edu.cn; zengjiaqi@zju.edu.cn).

Zhengyi Cai is with the Intelligent Transportation System Research Center, Hangzhou City University, Hangzhou 310058, China, and also with the College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China (e-mail: caizhengyi@zju.edu.cn).

Fengjie Fu is with the Department of Traffic Management Engineering, Zhejiang Police College, Hangzhou 310058, China (e-mail: fufengjie@zjjcxy.cn).

Digital Object Identifier 10.1109/TITS.2024.3358625

The rest of the paper is organized as follows: Section II reviews the literature on data and methods for vehicle reidentification. Section III define the problem, expounds on the matching principle by maximizing the travel time probability, and proposes the vehicle matching algorithm framework. Section IV introduces the approach of data survey and analyzes the survey data quality. The proposed vehicle matching algorithm is verified based on the survey data, and the performance of different matching algorithms are compared and analyzed. Section V summarizes the paper.

II. Literature Review

In the past few decades, many kinds of technologies are used for vehicle reidentification, such as conventional inductive loop detectors, automatic vehicle identification (AVI), and video image processing techniques.

The inductive loop detectors are the traditional and most commonly used traffic collection devices. Some studies apply the upstream and downstream loop data to vehicle matching, extra for travel time, and other traffic information. The approaches of vehicle matching based on loop data don't need additional equipment on the highway and expressway. However, the matching accuracy of loop data is low. For example, Coifman and Cassidy [11] proposed a vehicle reidentification method based on loop data using upstream and downstream dual-loop detectors on the freeway. Coifman and Krishnamurthy [12] improved the vehicle matching approach by applying the distinct information of the long vehicles.

Furthermore, the traffic loop detectors can extract the vehicle signature for vehicle reidentification. Oh, et al. [14] presented a vehicle reidentification model using the inductive vehicle signature based on the vehicle matching algorithm proposed by [13]. Then, the road segment speeds are extracted and the level of service in real-time is evaluated by re-identifying vehicles passing detectors on expressway. Then Oh, et al. [15] proposed a vehicle matching algorithm using the heterogeneous detection in the urban signalized section. Jeng et al. [16] applied the inductive loop vehicle signature information for vehicles matching on the upstream and downstream freeway.

The wireless magnetic sensors are also convenient for vehicle reidentification by extracting the vehicle magnetic signatures [17], [18]. Kwong, et al. [19], [20] proposed a vehicle matching algorithm based on the statistical model of vehicle signatures. The wireless magnetic sensors are utilized to collect vehicle signatures and the exact time. The algorithm achieves 75% accuracy, and the link vehicle count and the travel time distribution are stably estimated.

The AVI data have unique identification information and are convenient for vehicle re-identification. Dion and Rakha [21] presented an adaptive filtering algorithm to forecast the average travel time of the freeways and signalize links using low levels of market penetration AVI data. Li and Rose predicted the travel time range of the motorways employing the tollway AVI data based on vehicle identification [22]. However, AVI technology needs to install many new detectors in the road network [23], and the low penetration rate limits the application.

Vehicle reidentification based on computer vision methods, especially deep feature based methods become more and more popular [24]. The challenges of computer vision-based methods are the images of the same vehicle under different cameras have a various viewpoint, pose, illumination, and profile. Besides, different vehicles with the same type and color are very difficult to distinguish [25]. Some studies adopted more tiny features for vehicle reidentification, such as windows, lights, and pasted marks on the windshield [26]. Spatiotemporal information can further help vehicles re-identify [27], [28]. However, the deep network will be more complex and bloated and most vision-based methods seldom consider the spatiotemporal information of the vehicle. Moreover, fewer datasets have spatiotemporal information, which limits the development of methods [29].

The ALPR data have the characteristics of wide equipment in the road network, generous sampling rate, and high recognition accuracy. Given the excellent characteristics of ALPR data, many studies used the ALPR data for traffic information mining and modeling. Wang, et al. [30] propose a deep learning framework to model the dynamic characteristics of lane-level traffic flow as complex networks and uses spatial temporal graphs to predict network-scale traffic volumes, which outperforms various advanced deep learning models on two ALPR datasets.Yao, et al. [31] use ALPR data from Hangzhou, China to extract nine features reflecting commuting behavior. An, et al. [32] propose a method to estimate a lane-based traffic arrival pattern using ALPR data collected at upstream and downstream intersections. Zeng and Tang [33] propose a data-driven approach that combines ALPR data with taxi GPS trajectory data to estimate traffic flow in large road networks.

Different traffic parameters have different accuracy and rate requirements for license plate matching algorithms. For example, travel time and speed estimation don't require high sampling and matching rates. While traffic demand estimation, queue length estimation, and route reconstruction require higher matching rates. However, researchers have seldom conducted more in-depth studies on issues of vehicle matching in the case of recognition errors or unrecognized. Fu studied the recognition accuracy of ALPR and assumed the ALPR sensors achieved 95% of the recognition accuracy [34]. The recognition accuracy is affected by many factors and the actual recognition accuracy changes in different situations. Zhan et al. [7] applied the ALPR data to analyze the matched and unmatched information and proposed an interpolation approach based on the Gaussian process. The proposed methodology estimated link vehicle count and queue length by reconstructing the equivalent cumulative arrival-departure curve. Mo, et al. [2] proposed a Bayesian path reconstruction model to restore the missing information caused by the recognition error by utilizing match information of ALPR data.

III. Methodological Framework

A. Problem Description

The ALPR systems are widely deployed in major urban centers in China. These systems are primarily installed to monitor

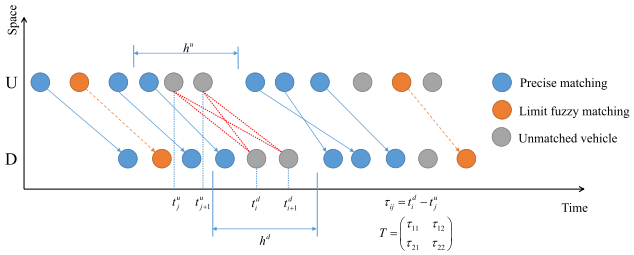


Fig. 1. Illustration of Time-space diagram of vehicle matching based on upstream and downstream ALPR data. The blue dots indicate that the vehicle is accurately recognized both upstream and downstream and can achieve precise matching. Orange dots indicate that some characters of the vehicle license plate are recognized incorrectly in the upstream or downstream and can be matched using the fuzzy matching method. Gray dots indicate unmatched vehicles.

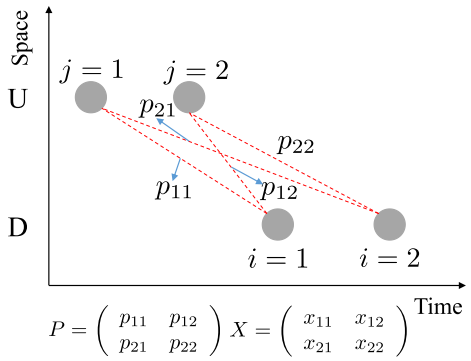


Fig. 2. Illustration of the travel time probability matrix and matching variables matrix during the matching process for two unmatched vehicles.

and record traffic violations, including red light violations. When the vehicle passes the stop line, the systems will take a picture and record the license plate number, timestamp, lane location, etc. If the vehicle license plates from the upstream intersection and downstream stop line can be identified correctly, they can be matched. Although mis-recognition or non-recognition often occurs, the timestamp and lane position can still be accurately recorded.

The typical license plate recognition data can be divided into three scenarios according to weather mis-recognition or non-recognition occurs in the upstream and downstream intersections: first, the license plates are correctly identified at upstream and downstream intersections and can be matched by the precise matching algorithm, as indicated by the blue dots in Fig. 1. Second, license plate recognition errors only occur on individual characters and can be matched by restricted fuzzy matching, as shown by the orange dots in Fig. 1. Third, the license plate cannot be matched by the characters when most of the characters on the license plate are incorrectly recognized or unrecognized, as shown by the gray circles in Fig. 1. For example, if two identical unmatched vehicles are detected at upstream and downstream intersections, respectively. Then, there may be two matching results. The travel time matrix of the two matching results can be obtained by calculating the timestamps differences between the downstream vehicles and the upstream vehicles. We need to find out which one is the correct match result by using the possible travel time matrix information. In Fig. 1, where h^u and h^d are the matching time windows of the upstream and downstream, t_j^u is the timestamp of the jth vehicle detected upstream, t_i^d is the timestamp of ith vehicle detected downstream, τ_{ij} means the possible travel time of two vehicles, T represents the travel time matrix.

B. Modeling Principles

Supposing there are n vehicles go through the downstream intersection from upstream, and these vehicles need to be matched, then there will be $n!$ matching results. The maximum probability matching result is obtained by calculating the probability of matching travel time as shown in Fig. 2, p_{ij} represents the travel time probability of the jth upstream vehicle matching with the ith downstream vehicle, P represents the travel time probability matrix. x_{ij} represents the matching variable, and X represents the matching variables matrix. $x_{ij} = 1$ if the jth upstream vehicle matches with the ith downstream vehicle. $x_{ij} = 0$ if they are don't match each other.

When there are n upstream vehicles and m downstream vehicles that need to match, the objective function is shown as follows to maximize the travel time probability:

$$\max z = \prod_{i}^{m}\prod_{j}^{n} p_{ij}^{x_{ij}} \tag{1}$$

where z is the travel time probability product of matching vehicles, p_{ij} is the travel time probability, and x_{ij} represents the matching variable.

The objective function is nonlinear, so we take the logarithm of the objective function, then $\log z = \sum_{i}^{m}\sum_{j}^{n} x_{ij} \log p_{ij}$. Let $\log p_{ij} = a_{ij}$, and $\log z = z'$, thus the objective function turns to $\max z' = \sum_{i}^{m}\sum_{j}^{n} x_{ij} a_{ij}$.

When $m > n > 0$, the downstream vehicles are more than upstream, we add $m \times (m-n)$ columns virtual vehicles and make the corresponding travel time probability equal to a small positive number greater than zero. Hence, its logarithm is a massive negative value $\log p_{ij}^v = -M$, where p_{ij}^v is the travel time probability of a virtual vehicle, and M is a large positive value. Therefore, the logarithm of new $m \times m$ travel time probability matrix as in

$$A'_{m\times m} = \begin{pmatrix} a_{11} & \cdots & a_{1n} & -M & \cdots & -M \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & -M & \cdots & -M \end{pmatrix}_{m\times m} \tag{2}$$

When $n > m > 0$, the downstream vehicles are less than upstream vehicles, we add $(n-m)\times n$ rows of virtual vehicles, thus the logarithm of the new $n \times n$ travel time probability matrix as in

$$A'_{n\times n} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \\ -M & \cdots & -M \\ \vdots & \ddots & \vdots \\ -M & \cdots & -M \end{pmatrix}_{n\times n} \tag{3}$$

Let $q = \max(m, n)$, then the objective function and constraints as in

$$\max\ z' = \sum_{i}^{q}\sum_{j}^{q} x_{ij}a_{ij}$$
$$s.t. \begin{cases} \sum_{i=1}^{q} x_{ij} = 1 \\ \sum_{j=1}^{q} x_{ij} = 1 \\ x_{ij} = 0 \qquad or\ 1. \end{cases} \tag{4}$$

Then the model becomes a linear integer optimization, which can be easily solved. The constraints demonstrate that an upstream vehicle only can match one downstream vehicle. The matching matrix is as

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & \ddots & \vdots \\ x_{q1} & \cdots & x_{qq} \end{pmatrix}_{q\times q} \tag{5}$$

Then the upstream or downstream vehicles that matched with virtual vehicles will be reset to unmatched data.

C. Matching Algorithm Framework

The proposed vehicle matching algorithm aims to address unmatched vehicles caused by recognition errors or unrecognized issues by maximizing travel time probability based on ALPR data. The algorithm automatically calculates the travel time probability under the different matching situations and implements the best matching by taking advantage of plentiful spatiotemporal information of ALPR data.

To achieve the above goals, our proposed framework includes travel time distribution estimation, travel time probability computing, the travel time confidence interval and the matching time window size calculating, restricted fuzzy matching, and the optimization model by maximizing the travel time probability, etc. The framework of the vehicle matching algorithm is shown in Fig. 3.

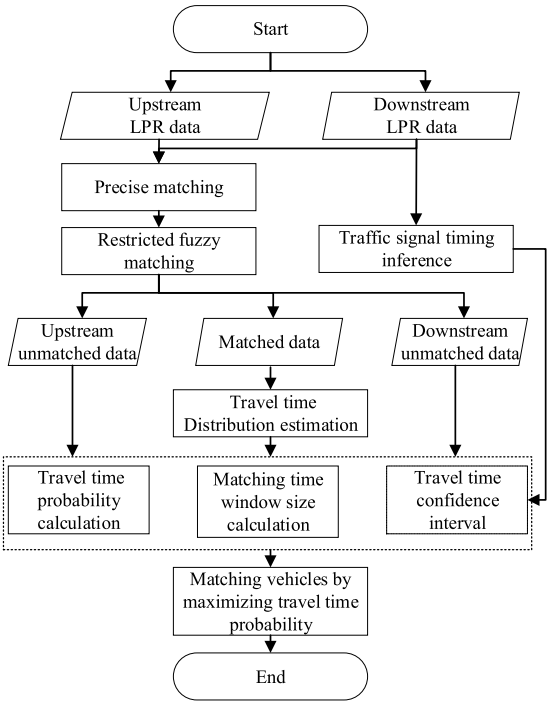


Fig. 3. The framework of the vehicle matching algorithm. The framework first uses the precise and restricted fuzzy matching method to obtain the matched license plates, then establishes the travel time distribution model for unmatched vehicles, and sets the matching time window size and confidence interval; finally, a matching optimization model is established to maximize travel time probability and the matching results are solved.

D. Modules of Matching Algorithm

1) Travel Time Distribution Estimation: The travel time distribution estimation requires extracting the travel time of the road segment and ALPR data preprocessing. Specifically, it includes precise matching of upstream and downstream license plates of the target road section, travel time extraction, and travel time outlier processing. The method of travel time removing outliers is mature, and a statistical-based approach [35], [36], [37] is adopted.

We construct the travel time distribution using kernel density estimation [38]. Assuming n travel time sample point data $\{\tau_1, \tau_2, \ldots, \tau_n\}$ that belong to independent and identical distribution F, and its probability density function (PDF) is f. Therefore, the estimated PDF $\hat{f}_h$ is shown as in

$$\hat{f}_h(\tau) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\tau - \tau_i}{h}\right) \tag{6}$$

where $K(\cdot)$ is the kernel function, $h > 0$ is a smoothing parameter announced bandwidth; τ is the travel time variable. We choose the Gaussian kernel as the kernel function, scilicet $K(\tau) = \varphi(\tau)$, where $\varphi(\tau)$ is the standard normal density function, as in

$$K(\tau) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\tau^2\right) \tag{7}$$

The two elements of kernel density estimation are the selection of kernel functions and the calculation of bandwidth. The method of bandwidth computing [39] for kernel density estimation is as in

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\hat{\sigma}n^{\frac{-1}{5}} \tag{8}$$

where $\hat{\sigma}$ is the standard deviation of the travel time sample, n is the number of samples for travel time data.

Afterward, the travel time cumulative distribution function (CDF) $F(\tau)$ is calculated as in

$$F(\tau) = \int_0^{\tau} \hat{f}_h(\tau)d\tau = \int_0^{\tau} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\tau - \tau_i}{h}\right)d\tau \tag{9}$$

2) The Travel Time Probability Calculating: The travel time probability is calculated by the estimated travel time CDF. The probability calculation for travel time is as in

$$p_{ij} = F\left(\tau_{ij}\right) - F\left(\tau_{ij} - \Delta t\right) \tag{10}$$

where τ_{ij} is the travel time of the downstream ith vehicle matching with the upstream jth vehicle, and p_{ij} is the

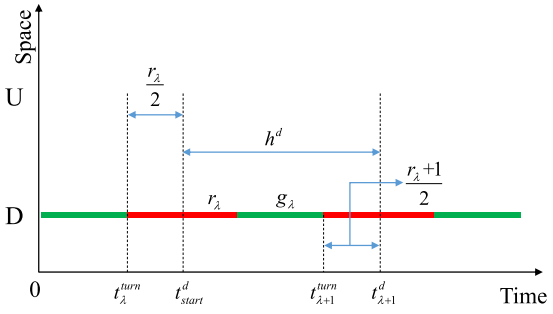


Fig. 4. Illustration of the downstream matching time window size calculation considering the traffic signal control scheme. The red line indicates the red light, and the green line means the green light.

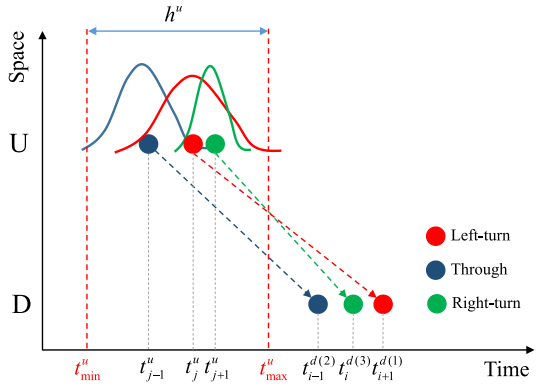


Fig. 5. Illustration of the upstream matching time window size calculation considering the differences in travel time of different turning traffic flows.

probability of travel time. $F\left(\tau_{ij}\right)$ is the travel time CDF, Δt is the timestamp accuracy for ALPR data, and $\Delta t = 1$s.

3) Travel Time Confidence Intervals Calculating: For the travel time cumulative distribution function F, with a probability $p\ (0 \le p \le 1)$, the inverse cumulative distribution function (ICDF) F^{-1} returns the travel time τ threshold as in

$$F^{-1}(p) = inf\{\tau \in R : F(\tau) \ge p\} \tag{11}$$

Suppose that confidence level is β, and $\beta = 1-\alpha$, then the travel time CI (confidence interval) is calculated as

$$CI = \left[F^{-1}\left(\frac{\alpha}{2}\right), F^{-1}\left(1-\frac{\alpha}{2}\right)\right] \tag{12}$$

where F^{-1} is the travel time ICDF, α is the significance level, $F^{-1}\left(\frac{\alpha}{2}\right)$ is lower limit of the travel time CI, and $F^{-1}\left(1-\frac{\alpha}{2}\right)$ is upper limit of the travel time CI.

4) The Matching Time Window Size Calculating: The matching time windows include the downstream and the upstream matching time window.

At signal-controlled intersections, right turn movement vehicles are generally not controlled by signals and right turn movement vehicles also pass during red light. But the downstream left turn and through movement vehicles are passing during the green light. Therefore, we consider two points in cutting the downstream matching time window: first, the cut matching time window contains at least one cycle of green light time; second, the vehicles passing during the green light are placed in the middle of the matching time window. As shown in Fig. 4, the downstream matching time window h^d is determined by the start time t_{start}^d and end time t_{end}^d, and h^d is calculated as in

$$t_{start}^d = t_\lambda^{turn} + \frac{r_\lambda}{2} \tag{13}$$

$$t_{end}^d = t_{\lambda+1}^{turn} + \frac{r_{\lambda+1}}{2} \tag{14}$$

$$h^d = \left[t_{start}^d, t_{end}^d\right] \tag{15}$$

where t_λ^{turn} is the moment the green light turns to the red light at the λth traffic light cycle, r_λ is the λth red light duration. Therefore, the downstream matching time window demands the traffic light timing scheme inference. In detail, please refer to the traffic light timing inference module.

To determine the upstream matching time window h^u, we extract the timestamps and lane positions of downstream vehicles within the matching time window at first. If n unmatched vehicles are found downstream, we record the corresponding timestamp vector as $T^d = \left\{t_i^{d(k)} : i = 1, 2, \ldots, n\right\}$, where $t_i^{d(k)}$ represents the timestamp of the ith downstream unmatched vehicle in the kth lane.

Second, besides the timestamps and the lane location of unmatched vehicles in downstream, we use the estimated travel time distribution to calculate h^u. When the downstream vehicle travels on different lanes, as shown in Fig. 5, the travel time has lane heterogeneity, and the travel time of different lanes doesn't belong to the same distribution.

In the upstream matching time window, it is assumed that m unmatched vehicles are found, and the corresponding timestamp vector is $T^u = \left\{t_j^u : j = 1, 2, \ldots, m\right\}$, where t_j^u is the timestamp of the jth upstream unmatched vehicle.

The earliest and latest upstream time points are calculated as follows by employing the above spatiotemporal information, as in

$$t_{min}^u = min\left(t_i^{d(k)} - F_{d(k)}^{-1}\left(1-\frac{\alpha}{2}\right)\right) \tag{16}$$

$$t_{max}^u = max\left(t_i^{d(k)} - F_{d(k)}^{-1}\left(\frac{\alpha}{2}\right)\right) \tag{17}$$

where $t_i^{d(k)}$ is the timestamp of the ith downstream unmatched vehicle in the kth lane, $F_{d(k)}^{-1}$ is the travel time ICDF of the kth lane.

Therefore, the matching time window corresponding to the upstream at the confidence level β, as in

$$h^u = \left[t_{min}^u, t_{max}^u\right] \tag{18}$$

5) Restricted Fuzzy Matching Module: In China, a general vehicle license plate consists of 7 letters of Chinese characters, letters and numbers, such as the actual license plate number in Fig. 6. After the license plate is photographed by the camera, the license plate information can be obtained through optical character recognition (OCR) technology and automatically stored in a dedicated ALPR database. However, the letters and numbers in these license plates may be identified incorrectly, such as the letter 'B' and the number '8' may be

Fig. 6. Camera view of license plate recognition system and license plate recognition results. In this case, the letter 'G' is incorrectly recognized as the number '8'.

confused with each other. We classify this individual character recognition error as the second type ALPR data recognition result. In Fig. 6, the letter 'G' is incorrectly recognized as the number '8'.

Aiming at this kind of limited character recognition error, we design a restricted fuzzy matching algorithm, which can match the license plates due to individual character recognition errors. The restricted fuzzy matching improves matching accuracy based on the fuzzy matching of license plate characters [40] by adding spatial and travel time constraints.

There are three matching conditions in the restricted fuzzy matching algorithm:

a) The license plate data for matching has upstream and downstream spatial constraints;
b) There is a travel time constraint between the time difference between upstream and downstream of the ALPR data;
c) The number of the same characters is greater than a particular value after comparing the license plate characters of upstream and downstream vehicles. The number of characters on an ordinary vehicle license plate is 7, and the number of characters on the license plate of a new energy vehicle is 8. Through experiments, we determined that when the same number of ordinary license plates is greater than or equal to 5, and the same number of new energy license plates is greater than or equal to 6, the upstream and downstream license plates are matched.

The specific steps for limiting fuzzy matching are as follows:

a) Sorting the downstream ALPR data in ascending order in the order of detection time;
b) Dividing the ALPR data into matched and unmatched;
c) Extracting a set of unmatched license plate data according to the downstream matching time interval, which is represented by a matrix as D_i^{um};
d) In the upstream unmatched ALPR data, find the data within the travel time confidence interval, denoted by U_i^{um};
e) Comparing the characters of each license plate number in U_i^{um} and D_i^{um} one by one in the order of the license plate characters, record the number of each pair of license plates with the same characters in the same order, filter out the license plates that meet the restricted conditions, record the matching data, and update the unmatched data set;
f) Extracting the next set of upstream data U_{i+1}^{um} and downstream data D_{i+1}^{um} that meet the conditions and continue to perform matching next window time.

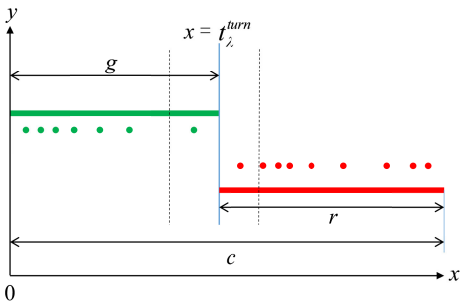


Fig. 7. Illustration of signal timing inference based on the ALPR data. Each point represents the moment when a car passes the stop line.

6) Traffic Signal Timing Inference Module: The traffic signal timing inference can be transformed into a binary classification problem by solving support vector machines (SVM) [41]. In fact, the ALPR data are not designed for the traffic signal control system, and the two systems don't connect directly. Especially when the traffic signal scheme is inductive control, the signal cycle and the green light time are real-time changes.

A typical signal control scheme has elements such as cycles, green lights, and red lights. ALPR data contains location and time information, which can divide the vehicle into green light phase and red light phase at the moment of passage. As shown in Fig. 7, vehicles are divided into green dots and red dots. Using the SVM method, the moment of the green to red light transition can be inferred (the split line in Fig. 7). Similarly, the moment of each red to green light transition can be estimated. In this way, we can solve the cycle duration of traffic signal timing c, the green light duration g, the red light duration r, and the phase change moment t_λ^{turn}.

IV. Case Studies and Experimental Results

A. Data Survey

To verify the effectiveness of the proposed vehicle matching algorithm under different lighting and weather conditions, we obtained the ALPR data of the survey road in four scenarios: sunny day (7:20- 10:30 am, December 30, 2020), clear night (19:00- 20:00, December 30, 2020), rainy day (8:00- 9:00 am, July 28, 2023) and rainy night (19:00- 20:00, July 28, 2023). we also collected the ground truth data of motor vehicles traveling on the actual road section by six high-definition (HD) cameras in the same period.

The survey road section is southbound of Minhe Road, with a length of about 327 meters. The upstream intersection is Zhenning Road-Minhe Road, the downstream intersection is Wenming Road-Minhe Road, and the middle entrance is the east gate of a residential district.

The survey locations are shown in Fig. 8. We deployed three cameras, two cameras, and one camera at the upstream intersection, the downstream intersection, and the entrance of the residential region in the middle, respectively. We record the license plate and time stamp of each vehicle passing the stop line on the upstream and downstream of the road. The time of the cameras are calibrated to guarantee the uniformity. We manually identified the vehicle video, then recorded the

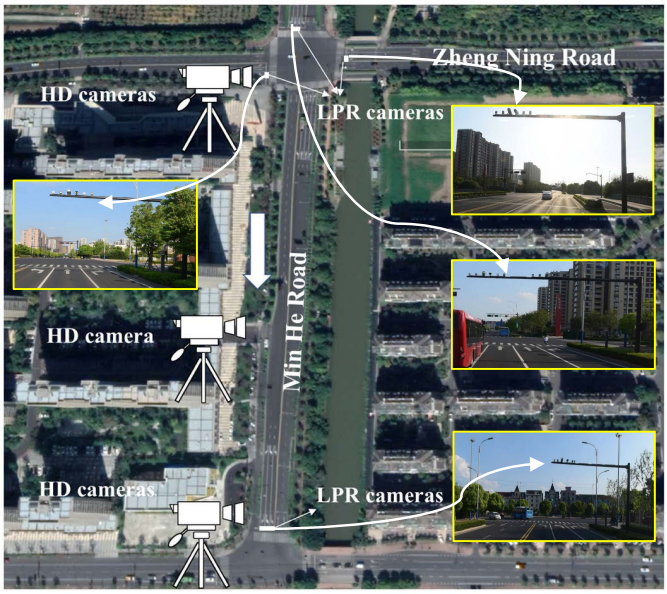


Fig. 8. Illustration of ALPR cameras in the case study and the HD cameras used in the manual survey.

TABLE I
AUTOMATIC LICENSE PLATE RECOGNITION RESULTS

No.	Section	Traffic volume of ALPR	Number of correctly identified vehicles	Recognition accuracy
①	**Upstream left turn movement**	209	161	77.03%
②	**Upstream through movement**	264	257	**97.35%**
③	**Upstream right turn movement**	121	73	**60.33%**
④	**Downstream left turn movement**	231	185	80.09%
⑤	**Downstream right turn movement**	321	286	89.10%
⑥	**Upstream summary**	594	491	82.66%
⑦	**Downstream summary**	552	471	85.33%

timestamp and license plate number when the vehicle passes the stop line.

B. Data Quality Analysis

We manually identified the surveyed video data and count the actual number of vehicles in each section. The actual traffic flow of the survey period in the sunny day(7:20- 10:30 am, December 30, 2020) is shown in Fig. 9. The quality analysis of ALPR data is shown in Table I. In the same period, we counted the total number of ALPR vehicles, the correct number of the ALPR data, and the recognition accuracy.

We summarized the traffic volume data based on different upstream and downstream flow directions in Table I. The Traffic volume of the ALPR is obtained by statistically aggregating the ALPR data. The ALPR data Recognition accuracy is obtained by calculating the ratio of the correct recognition number of vehicles to traffic volume of the ALPR.

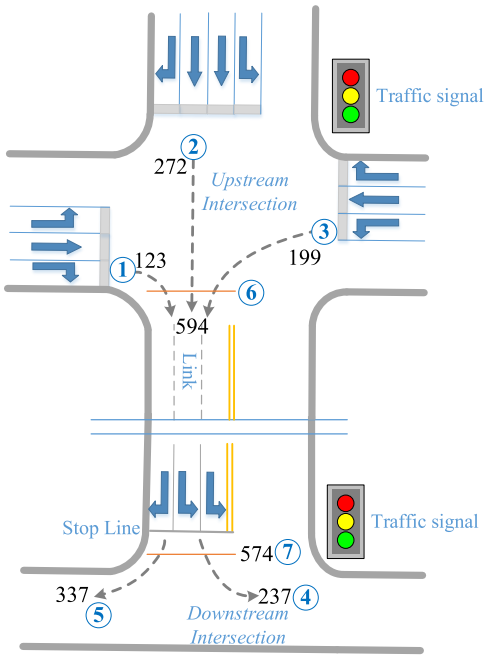


Fig. 9. The channelization and traffic flow statistics in the morning peak hours of sunny day.

TABLE II
LICENSE PLATE RECOGNITION ACCURACY UNDER DIFFERENT WEATHER AND LIGHTING CONDITIONS

Weather and light conditions	Recognition accuracy
Sunny day	85.33%
Rainy day	81.30%
Sunny nights	80.60%
Rainy nights	76.72%

In Table I, the bold numbers representing extreme values of recognition accuracy, the recognition accuracy of the upstream through movement is the highest: 97.35%. The recognition accuracy rate of upstream right turn movement is the lowest, which is 60.33%. the Recognition accuracy of total movements is 83.94%.

The captured images were under different lighting and weather conditions as shown in Fig. 10. The image shows that at night and in rainy conditions, the vehicle image, especially the license plate area, can still be captured clearly. we examined the accuracy of ALPR data under different lighting and weather conditions. We then conducted manual verifications to ensure that the recognized license plate numbers matched precisely with those in the images, allowing us to determine recognition accuracy. The results are summarized in Table II. Notably, the table illustrates that recognition accuracy is affected by both rainy weather and low-light conditions at night. However, even under these challenging conditions, the minimum accuracy level remains above 75%.

C. Analysis of Travel Time Distribution Characteristics

Travel time distribution builds a foundation for the vehicle matching algorithm. The manual survey gets the ground truth data of travel time distribution.

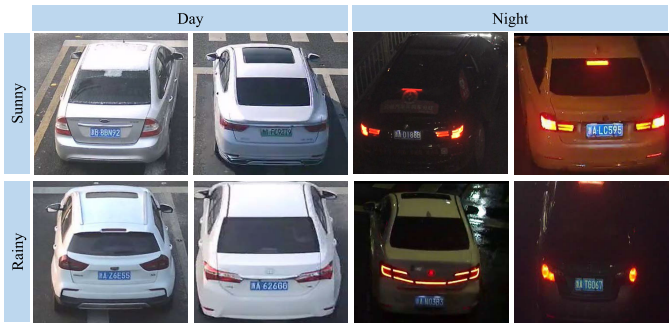


Fig. 10. The captured images under different light conditions and weather conditions. Please note that the last digit of the license plate is hidden to protect privacy.

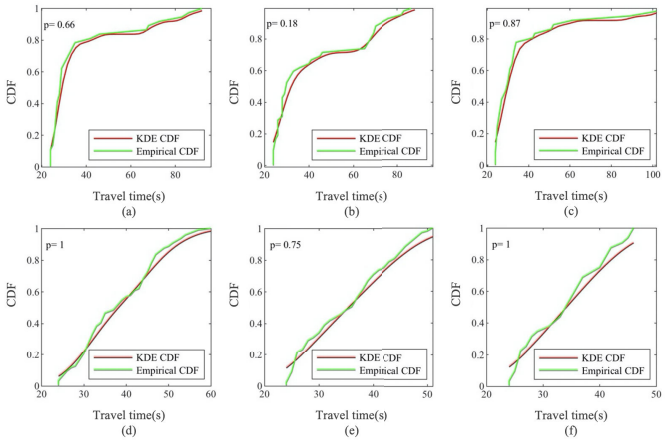


Fig. 11. The CDF of travel time distribution for different movements in different hours using kernel density estimation. The subgraphs (a) to (c) represent the travel time distribution of left-turn movement in the three hours from 7:30 to 10:30, and the subgraphs (d) to (f) represent the travel time distribution of right-turn movement.

With the travel time of the ALPR data sampled from the ground truth data, we further test whether the ALPR data belongs to the same distribution as the actual data. We use the 7:30 to 10:30 interval of survey data and divide each hour's data into an individual part. The original travel time data contains outliers, such as part of vehicles exceeding the speed limit or staying on the road too long. These vehicles lead to the travel time being too small or too large. We estimate travel time distribution by data after removing outliers. As Fig. 11 shows, the three sub-graphs in the upper row are the travel time CDF of the downstream left turn movement in three periods. The lower row sub-graphs represent the travel time CDF of downstream right turn movement. In Fig. 11, the green line stands for the actual travel time CDF, and the red line denotes the estimated travel time CDF. The p-value on the upper left corner of sub-graphs shows the result of the two-sample Kolmogorov-Smirnov test. When the p-value is more than the significance level, the estimated and the actual travel time distributions belong to the same distribution. The test results show that when the significance level is 0.05, the distribution of the travel time sampled by ALPR data is consistent with the actual travel time distribution.

We further study travel time characteristics and conclude that the travel time has time-varying and lane heterogeneity in the urban signalized link. We test travel time features by two-sample Kolmogorov-Smirnov test, and report H-values and p-values. We set the significance level of the test as 0.05. When the H-value equals 0, the two groups of data compared are from the same distribution. When the H-value equals 1, we reject the null hypothesis, and the two groups of travel time data belong to different distributions. The test results of the dynamic spatiotemporal characteristics of travel time distribution is shown in Table III. Where 'TT' represents the travel time data, 'H1' stands for the time interval for 7:30-8:30, 'H2' stands for the time interval of 8:30-9:30, 'H3' stands for the time interval of 9:30-10:30, 'L' denotes the left turn movement, 'R' means the right turn movement, and p-values marked with $*$ indicate significant differences.

TABLE III
TEST OF THE SPATIOTEMPORAL DYNAMICS OF TRAVEL TIME DISTRIBUTION

Characteristic	Contrast	H-value	p-value
Time-varying	$TT_{H1,L}$: $TT_{H2,L}$	1	0.04*
	$TT_{H2,L}$: $TT_{H3,L}$	0	0.47
	$TT_{H1,R}$: $TT_{H2,R}$	0	0.61
	$TT_{H2,R}$: $TT_{H3,R}$	0	0.84
Lane heterogeneity	$TT_{H1,L}$: $TT_{H1,R}$	1	0.00*
	$TT_{H2,L}$: $TT_{H2,R}$	1	0.02*
	$TT_{H3,L}$: $TT_{H3,R}$	1	0.02*

TABLE IV
COMPARISON OF MATCHING ALGORITHM RESULTS OF SUNNY DAY SCENARIO

Algorithms	Matching accuracy	Matching error rate	Unmatched rate
Precise matching [3, 4]	57.71%	0%	42.29%
Fuzzy matching [40]	80.43%	1.80%	17.77%
Box-Grained Reranking Matching [42]	76.24%	18.65%	5.11%
BOE [43]	75.91%	17.38%	6.71%
Ours	95.16%	2.48%	2.36%

In Table III, the time-varying feature test shows that the peak and off-peak travel time data are different distributions at the downstream left turn movement. However, the data on right turn movement didn't show the same time-varying feature. Besides, the periods of H2 and H3 are all during the off-peak time, and the travel time data of left or right turn movement during the off-peak period are the same distribution.

D. Benchmark Algorithms

1) Precise matching algorithm: The precise matching algorithm requires the license plates of upstream and downstream to be consistent. This algorithm is the most commonly used method for license plate matching and is applied in many traffic applications, such as travel time extraction, speed estimation [3], [4], etc.

TABLE V
COMPARISON OF MATCHING ALGORITHM RESULTS OF RAINY NIGHT SCENARIO

Algorithms	Matching accuracy	Matching error rate	Unmatched rate
Precise matching [3, 4]	46.62%	0%	53.38%
Fuzzy matching [40]	76.91%	2.50%	20.59%
Box-Grained Reranking Matching [42]	65.19%	28.83%	5.98%
BOE [43]	63.86%	26.92%	9.22%
Ours	91.41%	4.89%	3.70%

2) Fuzzy matching algorithm: In some ALPR data, recognition errors only occur in a few characters, but the other characters are correct. The fuzzy matching [40] addressed this part of the wrong license plate.
3) Box-Grained Reranking Matching [42]: This model deals with Multi-Camera Multi-Target tracking problems only by the appearance features of vehicles. The vehicle detection and ReID feature extraction modules are employed to locate all vehicles and extract the appearance features for all cameras and then the K-reciprocal nearest neighbors algorithm is used to match vehicles. The pretrain HRNet48 model in [42] is used as backbone for ReID training on our dataset.
4) BOE model [43]: this is another computer vision solution for vehicle matching based on the appearance features of vehicles in AI City Challenge 2022.

E. Result Analysis

In the proposed vehicle matching algorithm, we need to set parameters in advance: the matching time window size and the significance level of travel time distribution. In Table IV, the time window is set to 1 cycle of traffic control signal, and the significance level is set to 0.05. The results of sunny day is shown in TABLE IV, the proposed vehicle matching algorithm by maximizing travel time probability achieves the best cumulative matching accuracy: 95.16%. Our proposed method outperforms the compared methods in terms of matching accuracy and unmatched rate while maintaining a competitive matching error rate. The reidentification method [53-54] uses vehicle features only to match the upstream and downstream vehicles based on the similarity of the vehicle appearance features and achieve about 75% matching accuracy. The main reason for the low matching rate is that some vehicles in the traffic flow have the same or similar appearance.

To compare the results of this method and the benchmark methods under poor lighting and weather conditions, we selected the rainy night scenario (19:00- 20:00, July 28, 2023) for analysis. The results in TABLE V show that compared with the sunny day scenario, under severe lighting and weather conditions, the matching accuracy of all methods decreases, and the matching accuracy error increases. Our proposed method is superior to all comparative methods in terms of all indicators, and the matching accuracy remains above 90%. The fuzzy matching method has the smallest decrease, which is due to the increased probability of individual characters of the license plate being misrecognized under poor lighting conditions, but has less impact on fuzzy matching. The visual method based on appearance features has a larger decrease. This is due to the greater impact of rainy days and reflective conditions on appearance, which can be seen in Fig. 10.

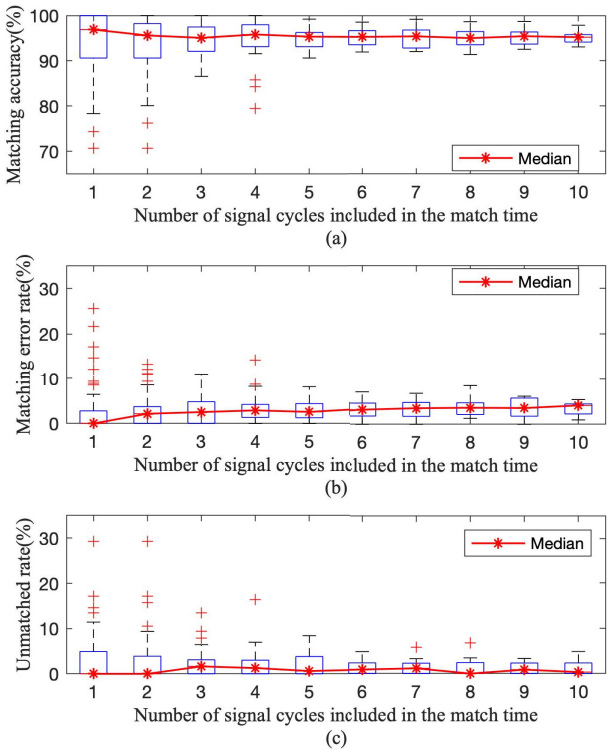


Fig. 12. Boxplots of evaluations index of proposed method in sunny day under different matching windows size.

F. Parameters Sensitivity Analysis

We performed parametric sensitivity analyses on time window size and significance level in the sunny day scenario. First, we fix the significance level equal to 0.01. Second, we calculate the effects of the different time window sizes on the three evaluations index: matching accuracy, matching error rate, and unmatched rate. We evaluate the above three indicators when a matching time window contains 1 to 10 cycles. As shown in Fig. 12(a), when a longer matching time window is used, the median and the variance of matching accuracy show a downward trend. In Fig. 12(b), as the matching time window size increases, the median of the matching error rate increases, but the variance of the matching error rate decreases.

In Fig. 13, we fix the matching time window size that includes one signal cycle. The upstream matching window sizes are calculated by (16) and (17), which depend on travel time distribution. As shown in Fig.11, the 90% to 99% values of the travel time cumulative distributions are close, leading to the upstream matching time window intervals are change little. From the results, we can conclude that the effects of different significance levels on the three evaluation indicators are insignificant.

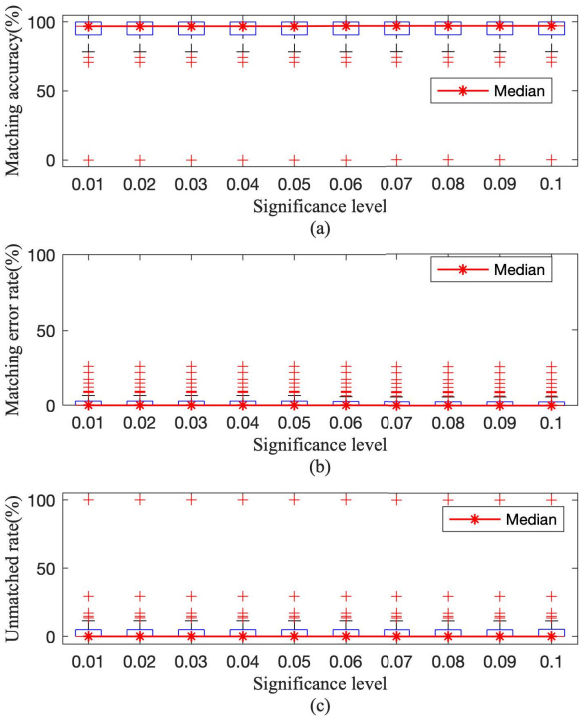


Fig. 13. Evaluations index under different significance levels setting. The results show that the matching accuracy, matching error rate, and non-matching rate of this method do not change significantly when setting different travel time confidence levels, indicating robustness.

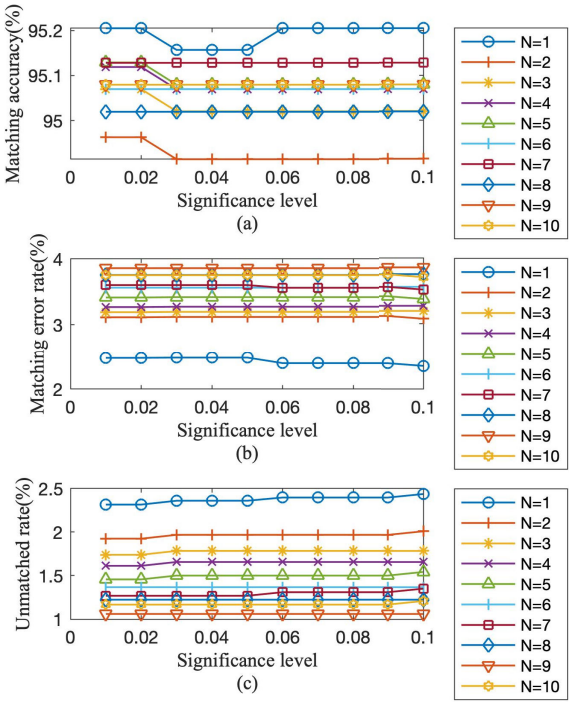


Fig. 14. Evaluations index under different combinations of significance level and matching window size.

We show the changes in the evaluation indicators for different parameter combinations in Fig. 14. We selected the matching time windows to be 1 to 10 signal cycles and set the significance levels to be 0.01 to 0.1, respectively. The results show that the matching accuracy of the algorithm is higher than 94.91% under different parameter combinations. The accuracy is relatively stable and at a high level. But the significance level has no significant impact on the evaluation index. In Fig. 14, as the bigger matching time window size is adopted, the matching error rate increases, but the matching accuracy and the unmatched rate decrease.

V. Conclusion

Through investigation analysis and experimental verification, we can draw these conclusions:

1) The travel time obtained from the ALPR data belongs to the same distribution as the ground truth data, and the travel time reflects time-varying and lane heterogeneity.
2) The proposed vehicle matching algorithm improves the matching accuracy based on the maximum travel time probability. Compared with the precise license plate matching, the cumulative matching accuracy is increased by 37.45%.
3) The results of parameter sensitivity analysis on the time window size and significance level show that under different parameter combinations, the matching accuracy of the algorithm is relatively stable and is at a high level, not lower than 94.91%. We conclude that the proposed vehicle matching algorithm is suitable for different matching time windows and different saliency levels. The proposed vehicle matching algorithm provides a good foundation for subsequent traffic information services. Different significance levels have no significant effect on the three evaluation indicators. It will match more upstream and downstream vehicles with the bigger matching time window, but the matching error rate will increase.

In summary, the new method proposed in this paper may provide advantages for future applications. The ALPR data with a high sampling rate and widespread coverage of road networks, these characteristics are proper for serving origin and destination demand estimation, link vehicle count estimation, travel time estimation, and queue length estimation. Aiming at the recognition errors and non-recognition issues of ALPR data, the vehicle matching algorithm based on the maximum travel time probability improves the cumulative matching accuracy of upstream and downstream vehicles. The new model does have limitations. This method is only suitable for road sections where license plate recognition can be achieved on both upstream and downstream sections, which limits the scope of this method. In the future, we will further estimate the link dynamic vehicle count with the vehicle matching algorithm to optimize the network-wide traffic signal light control.

References

[1] D. Ma, X. Luo, W. Li, S. Jin, W. Guo, and D. Wang, "Traffic demand estimation for lane groups at signal-controlled intersections using travel times from video-imaging detectors," *IET Intell. Transp. Syst.*, vol. 11, no. 4, pp. 222–229, May 2017, doi: 10.1049/iet-its.2016.0233.

[2] B. Mo, R. Li, and J. Dai, "Estimating dynamic origin–destination demand: A hybrid framework using license plate recognition data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 7, pp. 734–752, 2020.

[3] F. Fu, W. Qian, and H. Dong, "Estimation of route travel time distribution with information fusion from automatic number plate recognition data," *J. Transp. Eng., A, Syst.*, vol. 145, no. 7, Jul. 2019, Art. no. 04019029, doi: 10.1061/jtepbs.0000242.

[4] X. Luo, D. Wang, D. Ma, and S. Jin, "Grouped travel time estimation in signalized arterials using point-to-point detectors," *Transp. Res. B, Methodol.*, vol. 130, pp. 130–151, Dec. 2019, doi: https://doi.org/10.1016/j.trb.2019.10.007.

[5] B. Mo, R. Li, and X. Zhan, "Speed profile estimation using license plate recognition data," *Transp. Res. C, Emerg. Technol.*, vol. 82, pp. 358–378, Sep. 2017, doi: 10.1016/j.trc.2017.07.006.

[6] C. He, D. Wang, M. Chen, G. Qian, and Z. Cai, "Link dynamic vehicle count estimation based on travel time distribution using license plate recognition data," *Transportmetrica A, Transp. Sci.*, vol. 19, no. 2, pp. 1–22, Mar. 2023.

[7] X. Zhan, R. Li, and S. V. Ukkusuri, "Lane-based real-time queue length estimation using license plate recognition data," *Transp. Res. C, Emerg. Technol.*, vol. 57, pp. 85–102, Aug. 2015.

[8] D. Ma, X. Luo, S. Jin, W. Guo, and D. Wang, "Estimating maximum queue length for traffic lane groups using travel times from video-imaging data," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 3, pp. 123–134, Fall. 2018, doi: 10.1109/MITS.2018.2842047.

[9] X. Luo, D. Ma, S. Jin, Y. Gong, and D. Wang, "Queue length estimation for signalized intersections using license plate recognition data," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 3, pp. 209–220, Fall. 2019, doi: 10.1109/MITS.2019.2919541.

[10] R. Li, Z. Liu, and R. Zhang, "Studying the benefits of carpooling in an urban area using automatic vehicle identification data," *Transp. Res. C, Emerg. Technol.*, vol. 93, pp. 367–380, Aug. 2018.

[11] B. Coifman and M. Cassidy, "Vehicle reidentification and travel time measurement on congested freeways," *Transp. Res. A, Policy Pract.*, vol. 36, no. 10, pp. 899–917, Dec. 2002.

[12] B. Coifman and S. Krishnamurthy, "Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure," *Transp. Res. C, Emerg. Technol.*, vol. 15, no. 3, pp. 135–153, Jun. 2007.

[13] C. Sun, S. G. Ritchie, K. Tsai, and R. Jayakrishnan, "Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways," *Transp. Res. C, Emerg. Technol.*, vol. 7, no. 4, pp. 167–185, Aug. 1999.

[14] C. Oh, A. Tok, and S. G. Ritchie, "Real-time freeway level of service using inductive-signature-based vehicle reidentification system," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 138–146, Jun. 2005.

[15] C. Oh, S. G. Ritchie, and S.-T. Jeng, "Anonymous vehicle reidentification using heterogeneous detection systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 460–469, Sep. 2007.

[16] S.-T. Jeng, Y. C. A. Tok, and S. G. Ritchie, "Freeway corridor performance measurement based on vehicle reidentification," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 639–646, Sep. 2010.

[17] J. Wang, N. Indra-Payoong, A. Sumalee, and S. Panwai, "Vehicle reidentification with self-adaptive time windows for real-time travel time estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 540–552, Apr. 2014.

[18] R. O. Sanchez, C. Flores, R. Horowitz, R. Rajagopal, and P. Varaiya, "Vehicle re-identification using wireless magnetic sensors: Algorithm revision, modifications and performance analysis," in *Proc. IEEE Int. Conf. Veh. Electron. Saf.*, Jul. 2011, pp. 226–231.

[19] K. Kwong, R. Kavaler, R. Rajagopal, and P. Varaiya, "Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 6, pp. 586–606, Dec. 2009.

[20] K. Kwong, R. Kavaler, R. Rajagopal, and P. Varaiya, "Real-time measurement of link vehicle count and travel time in a road network," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 814–825, Dec. 2010, doi: 10.1109/TITS.2010.2050881.

[21] F. Dion and H. Rakha, "Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates," *Transp. Res. Part B, Methodol.*, vol. 40, no. 9, pp. 745–766, Nov. 2006.

[22] R. Li and G. Rose, "Incorporating uncertainty into short-term travel time predictions," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 6, pp. 1006–1018, Dec. 2011.

[23] H. D. Sherali, J. Desai, and H. Rakha, "A discrete optimization approach for locating automatic vehicle identification readers for the provision of roadway travel times," *Transp. Res. B, Methodol.*, vol. 40, no. 10, pp. 857–871, Dec. 2006.

[24] H. Li, C. Li, A. Zheng, J. Tang, and B. Luo, "MsKAT: Multi-scale knowledge-aware transformer for vehicle re-identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19557–19568, Oct. 2022.

[25] S. D. Khan and H. Ullah, "A survey of advances in vision-based vehicle re-identification," *Comput. Vis. Image Understand.*, vol. 182, pp. 50–63, May 2019.

[26] B. He, J. Li, Y. Zhao, and Y. Tian, "Part-regularized near-duplicate vehicle re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3992–4000.

[27] Y. Shen, T. Xiao, H. Li, S. Yi, and X. Wang, "Learning deep neural networks for vehicle re-ID with visual-spatio-temporal path proposals," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1918–1927.

[28] X. Liu, W. Liu, T. Mei, and H. Ma, "PROVID: Progressive and multimodal vehicle reidentification for large-scale urban surveillance," *IEEE Trans. Multimedia*, vol. 20, no. 3, pp. 645–658, Mar. 2018.

[29] X. Liu, W. Liu, T. Mei, and H. Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, 2016, pp. 869–884. [Online]. Available: https://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf

[30] P. Wang, J. Lai, Z. Huang, Q. Tan, and T. Lin, "Estimating traffic flow in large road networks based on multi-source traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 9, pp. 5672–5683, Sep. 2021.

[31] W. Yao, M. Zhang, S. Jin, and D. Ma, "Understanding vehicles commuting pattern based on license plate recognition data," *Transp. Res. C, Emerg. Technol.*, vol. 128, Jul. 2021, Art. no. 103142.

[32] C. An, X. Guo, R. Hong, Z. Lu, and J. Xia, "Lane-based traffic arrival pattern estimation using license plate recognition data," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 4, pp. 133–144, Jul. 2022.

[33] J. Zeng and J. Tang, "Modeling dynamic traffic flow as visibility graphs: A network-scale prediction framework for lane-level traffic flow based on LPR data," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 4173–4188, Apr. 2023.

[34] F. Fu, "Travel time estimation for urban road networks using ANPR data," Ph.D. dissertation, College Civil Eng. Archit., Zhejiang Univ., Hangzhou, China, 2017.

[35] W. Qin, X. Ji, and F. Liang, "Estimation of urban arterial travel time distribution considering link correlations," *Transportmetrica A, Transp. Sci.*, vol. 16, no. 3, pp. 1429–1458, Jan. 2020.

[36] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Trans. Control Syst. Technol.*, vol. 10, no. 1, pp. 55–63, Jan. 2002.

[37] D. Wang, F. Fu, X. Luo, S. Jin, and D. Ma, "Travel time estimation method for urban road based on traffic stream directions," *Transportmetrica A, Transp. Sci.*, vol. 12, no. 6, pp. 479–503, Jul. 2016.

[38] A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations*. Oxford, U.K.: Oxford Univ. Press, 1997.

[39] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Evanston, IL, USA: Routledge, 2018.

[40] N. Watcharapinchai and S. Rujikietgumjorn, "Approximate license plate string matching for vehicle re-identification," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2017, pp. 1–6.

[41] P. Hao, X. Ban, K. P. Bennett, Q. Ji, and Z. Sun, "Signal timing estimation using sample intersection travel times," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 792–804, Jun. 2012.

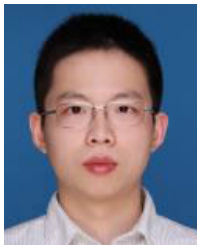
[42] X. Yang et al., "Box-grained reranking matching for multi-camera multi-target tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, New Orleans, LA, USA, Jun. 2022, pp. 3095–3105.

[43] F. Li et al., "Multi-camera vehicle tracking system for AI city challenge 2022," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, New Orleans, LA, USA, Jun. 2022, pp. 3264–3272.

Chunguang He received the Ph.D. degree in highway and transportation engineering from Zhejiang University, Hangzhou, China, in 2023. His research interests include traffic control, traffic flow theory, and intelligent transportation systems.

Dianhai Wang is currently a Professor of traffic engineering with Zhejiang University, Hangzhou, China. His research interests include traffic control, traffic flow theory, and traffic information. He is a member of the China Higher Education Steering Committee, Traffic Engineering Subject; and the China City Planning Academic Committee.

Zhengyi Cai (Member, IEEE) received the B.S. degree from the Department of Transportation, Jilin University, Changchun, China, in 2009, and the Ph.D. degree in highway and transportation engineering from Zhejiang University, Hangzhou, China, in 2018. His current research interests include intelligent transportation systems and urban mobility.



Jiaqi Zeng received the B.S. degree in civil engineering and the Ph.D. degree in highway and transportation engineering from Zhejiang University, Hangzhou, China, in 2018 and 2023, respectively. His research interests include traffic flow theories and traffic state identification.

Fengjie Fu received the Ph.D. degree in road and traffic engineering from Zhejiang University, Hangzhou, China, in 2017. She is currently a Lecturer with the Department of Traffic Management Engineering, Zhejiang Police College, Hangzhou. Her research interests include traffic safety analysis and traffic control.