Multimodal Traffic Speed Monitoring: A Real-Time System Based on Passive Wi-Fi and Bluetooth Sensing Technology

Ziyuan Pu , *Member, IEEE*, Zhiyong Cui, Jinjun Tang, Shuo Wang, and Yinhai Wang, *Senior Member, IEEE*

Abstract—Traffic speed is one of the critical indicators reflecting traffic status of roadway networks. The abnormality and sudden changes of traffic speed indicate the occurrence of traffic congestions, accidents, and events. Traffic control and management systems usually take the spatiotemporal variations of traffic speed as the critical evidence to dynamically adjust the traffic signal timing plan, broadcast traffic accidents, and form a management strategy. Meanwhile, transport is multimodal in most cities, including vehicles, pedestrians, and bicyclists. Traffic states of different traffic modes are usually used simultane- ously as the significant input of advanced traffic control systems, e.g., multiobjective traffic signal control system, connected vehi- cles, and autonomous driving. In previous studies, Wi-Fi and Bluetooth passive sensing technology was demonstrated as an effective method for obtaining traffic speed data. However, there are some challenges that greatly affect the accuracy the esti- mated traffic speed, e.g., traffic mode uncertainty and the errors caused by sensors' detection range. Thus, this study develops a real-time method for estimating the multimodal traffic speed of road networks covered by Wi-Fi and Bluetooth passive sen- sors. To address the two identified challenges, an algorithm is developed to correct the biased estimated traffic speed based on the received signal strength indicator of Wi-Fi and Bluetooth signals, and a novel semisupervised Possibilistic Fuzzy *C*-Means clustering algorithm is proposed for identifying traffic modes of Wi-Fi and Bluetooth device owners. The performance of the proposed algorithms is evaluated by comparing with the selected baseline algorithms. The experimental results indicate the supe- riority of the proposed algorithm. The proposed method of this study can provide accurate and real-time multimodal traffic speed information for supporting traffic control and manage- ment, and, thus, improving the operational performance of the whole road network.

Index Terms—Multimodal traffic speed, real-time monitoring system, received signal strength indicator (RSSI), traffic mode identification, Wi-Fi and Bluetooth passive sensing.

Ziyuan Pu is with the School of Engineering and Advanced Engineering Platform, Monash University, Selangor, Bandar Sunway 47500, Malaysia (e-mail: ziyuan.pu@monash.edu).

Zhiyong Cui is with the School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: zhiyongc@uw.edu).

Jinjun Tang is with the Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China (e-mail: jinjuntang@csu.edu.cn).

Shuo Wang is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: sharon_wang@njust.edu.cn).

Yinhai Wang is with the Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: yinhai@uw.edu).

I. INTRODUCTION

TRAFFIC speed is one of the critical indicators reflecting traffic status of roadway networks [1]. The abnormality and sudden changes of traffic speed indicate the occurrence of traffic congestions, accidents, and events. Traffic control and management systems usually take the spatiotemporal varia- tions of traffic speed as the critical evidence to dynamically adjust traffic signal timing plan, broadcast traffic accidents, and form management strategy. In the meanwhile, traffic is multimodal in most cities, including motorized vehicle, pedes- trian, bicyclist. Multimodal traffic speed is usually used simul- taneously as the significant input of advanced traffic control systems, e.g., multiobjective traffic signal control system [2], connected vehicles, and autonomous driving.

Basically, the traffic speed data collection method can be divided into two categories in terms of sensing tech- nologies. On the one hand, traffic speed can be estimated based on the data transmitted by travelers or vehicles, such as GPS-based traffic estimation [3]–[5] and crowdsourcing methods [6], which are so-called active sensing methods. However, in some cases, the small sample size limits the implementation of these technologies and the reliability of the estimated traffic speed. On the other hand, roadside traffic sen- sors capture valuable data for passively sensing traffic speed. Generally, there are two main methods for passively sensing traffic speed, either calculating traffic speed based on reiden- tification results, or estimating the average speed based on point detection methods. For point detection, multiple meth- ods were developed based on loop detectors [7], microwave sensors [8], etc. However, since the speed information in the middle of road segments is unachieved for point detectors, the estimated speed will be erroneous if the congestion or traffic accidents occur in the middle of road segments. For reiden- tifying vehicles or travelers at different locations, a unique identifier or representative features are required. Then the traffic speed can be calculated based on the detection time

difference. Compared with point detection methods, it can not only observe the traffic status at detection points but the traffic status in the middle of road segments also can be inferred. Previously, multiple sensing technologies were developed for reidentification, e.g., computer vision [9]–[11], RFID-based speed detection [12], license plate recognition [13], and passive Wi-Fi and Bluetooth sensing technology [14]. For computer vision-based traffic speed sensing, the license plate number of motorized vehicles and the extracted nonmotorized travelers' features are treated as unique identifiers [15]. However, several disadvantages make it questionable for traffic speed detection, including privacy issues [16], low accuracy and efficiency (especially for reidentifying nonmotorized travelers) [17], and the influence of environmental conditions (illumination, weather, etc.). Besides, other reidentification methods also have limitations, such as low penetration rate of RFID tags, which impair their applicability and reliability. Recently, as the ubiquitous usage of smart phones, scholars developed traffic status estimation methods based on passively sensing Wi-Fi and Bluetooth mobile device [18].

Jason *et al.* proposed a method for real-time travel time estimation using Media Access Control address matching in 2008 [19], which is the first research applying Wi-Fi and Bluetooth sensing technology in travel time acquisition. Then, a plenty of studies investigated the influential factors of traffic speed detection accuracy and reliability [18], [20]–[30]. It is stated that the accuracy can be affected by the communication quality of Bluetooth sensors, the detection range of the sensors, the level of average traffic speed on a road segment, vertical sensor placement, and the type of antennas using for sensing Wi-Fi and Bluetooth device. To improve the accuracy of traffic speed estimation based on Wi-Fi and Bluetooth sensing data, other data sources, e.g., loop detector data, GPS recordings, were fused with Wi-Fi and Bluetooth sensing data [31]–[33], Hyoshin and Ali proposed a two-stage stochastic model for determining the optimal number and location of Wi-Fi and Bluetooth sensors [34], traffic mode identification model was developed [35], and received signal strength indicator (RSSI) was utilized for improving estimation accuracy [36], [37]. Besides, such technology was implemented for monitoring traffic in different scenarios, including travel time prediction [38], [39], arterial traffic congestion analysis [40], bicycle travel time estimation [41], travel time delay monitoring in work zones [42], roadway system assessment [43], freeway travel time monitoring system development [44], pedestrian network monitoring [45], crowd mobility pattern exploration [46], and transit ridership flow monitoring [47].

In summary, most of the existing studies based on Wi-Fi and Bluetooth passive sensing only involves in traffic speed monitoring on freeways. However, for providing multimodal traffic speed in the urban area, there are two main challenges that have not been addressed, including the biasness of traffic speed estimation caused by the detection range of the sensor, and the traffic modes uncertainty of Wi-Fi and Bluetooth device owners. For the biasness caused by the detection range of the sensor, previous studies tried to mitigate the bias by evaluating the level of the estimated errors by

the first or the last detection [37], and exploring the detection point which is the closet to the sensor [48]. However, those simple methods have not effectively corrected the biasness. For the traffic mode uncertainty, several studies have established algorithms to identify traffic modes based on supervised machine learning algorithms [35], filtering method based on predefined thresholds [49], and logit models [45]. However, the accuracy of those existing methods highly relies on the empirical information extracting from a large set of labeled data or predefined thresholds. Considering that the labeled data are hard to obtain and the optimal values of predefined thresholds are difficult to determine, the existing methods may still need to be improved to release the dependency on larger amount of labeled data or predefined thresholds. Therefore, the major objective of this research is to address the two identified challenges for improving the accuracy and reliability of multimodal traffic speed monitoring systems based on passive Wi-Fi and Bluetooth sensing. The major contributions of this study can be summarized as follows.

1) An algorithm is established to correct the biasness of the estimated traffic speed caused by the detection range of the sensor based on RSSI of Wi-Fi and Bluetooth signals. Ground-truth speed data and the corresponding estimated traffic speed of 408 trips are compared to validate the performance.

2) A traffic mode identification algorithm is proposed based on a designed semisupervised Possibilistic Fuzzy C-Means (PCM) clustering algorithm. Multiple baseline algorithms are selected for the evaluation purpose. The evaluation results demonstrate the advantage of the algorithm in terms of detection accuracy.

3) A real-time multimodal traffic speed estimation algorithm is established for estimating the traffic speed of the road network covered by Wi-Fi and Bluetooth passive sensors. The accuracy of the proposed algorithm is evaluated based on the comparison of estimated results and ground-truth data. The evaluation results indicate the proposed algorithm can accurately estimate all three traffic modes, including walk, bike, and car modes.

II. METHODS

In this section, the proposed system architecture and the algorithm framework are presented. Section II-A introduces the system architecture. Section II-B presents the overall algorithm framework. Sections II-C and II-D show the proposed algorithm for correcting estimated traffic speed and identifying traffic mode. Section II-E introduces the evaluation metrics.

A. System Architecture

Fig. 1 shows the architecture of the proposed system. The system mainly consists of two parts: 1) sensor networks at the edge side and 2) the remote server for data analysis, managing, and visualizing. The data transmission between two sides is supported either by wireless communication or through Ethernet cables. The major role of sensor networks at the edge side is to capture the MAC addresses of the mobile devices within sensors' detection range. Real-time MAC address data

TABLE I
EXTRACTED FEATURES

| Features (Unit)	Unit	Definition
Start Detection Times	Times	The number of times that a unique MAC address is detected at the start sensing location
End Detection Times	Times	The number of times that a unique MAC address is detected at the end sensing location
Start Detection Duration	Seconds	The total amount of time for a unique mac to be detected at the start sensing location
End Detection Duration	Seconds	The total amount of time for a unique mac to be detected at the end sensing location
First Time Difference	Seconds	The detection time difference between the first detected data point at two sensing locations of a road segment
Last Time Difference	Seconds	The detection time difference between the last detected data point at two sensing point of a road segment
Original Speed	Meters/Second	The estimated speed calculating by Equation (1)
Corrected Speed	Meters/Second	The estimated speed calculating by Equation (6)
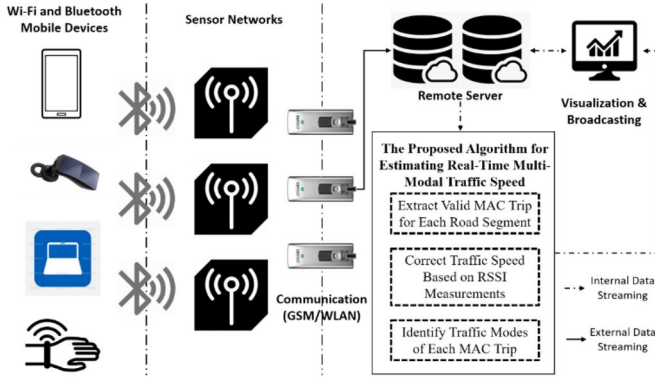
Fig. 1.   System architecture.

Algorithm 1: Real-Time Multimodal Traffic Speed <u>Estimation</u>

Initialization: start time t_0, time interval Δt
for road segments i in $\{1, \ldots, N\}$ **do**:
    Extract all MAC trips M within time interval $[t_0, \ t_0 + \Delta t]$
    for MAC trip j in $\{1, \ldots, M\}$ **do**:
        Correct traffic speed based on RSSI using Equation (6)
        Extract features vector v_j of MAC Trip j
    end for
    Identify travel modes for all MAC trips M using
    Algorithm 2
end for
Output: average multi-modal traffic speed of all road segments for the time window <u>$[t_0, \ t_0 + \Delta t]$</u>

---

collected by all nodes of a sensor network is transmitted to the remote server for further analysis and visualization. At the serve side, the proposed algorithm is implemented to estimate the real-time multimodal traffic speed based on the Wi-Fi and Bluetooth MAC address data. The results are visualized and broadcasted to transportation managers and road users for decision-making and travel scheduling.

B. Real-Time Multimodal Traffic Speed Monitoring

The main objective of the proposed algorithm is to estimate average multimodal traffic speed, including car, bike, and walk modes, of all road segments for the time windows with predefined time interval. In this study, a road segment is defined as the road between two adjacent Wi-Fi and Bluetooth MAC address sensors. Algorithm 1 presents the proposed algorithm. First, the algorithm traverses each road segment of a road network to extract valid MAC trips within a specific time window $[t_0, \ t_0 + \Delta t]$, where t_0 is the start time of a time window, and Δt is a predefined time interval. A valid MAC trip is generated when a unique MAC is detected by two adjacent sensors in chronological order within a reasonable time range. After all valid MAC trips are achieved for a road segment, the traffic speed of each trip will be corrected. The traffic speed correction algorithm will be introduced in the Section II-B. Then, a vector of features which represents the characteristics of each MAC trip will be extracted. The extracted features are presented in Table I. Basically, the features not only contain the travel time and speed attributes, but also include the movement features at two sensing stations of a road segment, e.g., detection times and duration. Then, all extracted features will be used as the input of the traffic mode identification algorithm based on the proposed semisupervised PCM clustering which will be introduced in the Section II-C. Finally, the average multimodal traffic speed of a road segment within the time window $[t_0, \ t_0 + \Delta t]$ is the outcome of the proposed algorithm. The algorithm can be implemented in a real-time way by repeating Algorithm 1 for every Δt.

C. Correcting Traffic Speed Based on Received Signal Strength Indicator

In general, MAC address sensors can capture the MAC address of the mobile device within a specific detection range. Typically, the detection range is about 50–80 m for Wi-Fi sensing and 20 m for Bluetooth sensing. The MAC address of a discoverable mobile device can be detected anywhere
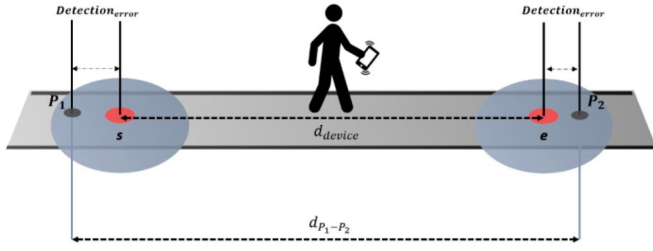
Fig. 2. Estimated traffic speed error caused by detection range.

Fig. 3. RSSI versus distance with regression line. (a) Wi-Fi and (b) Bluetooth.

within the detection range. However, the exact detection location cannot be precisely identified. Such spatial uncertainty causes errors for traffic speed estimation. From the illustration in Fig. 2, the red circles represent the location of two Wi-Fi and Bluetooth sensors which are denoted as s and e. The two blue circles represent the detection ranges of two sensors. While the pedestrian is walking from the start sensing location s to the end sensing location e, the MAC address of the mobile device carried by the pedestrian is detected at P_1 and P_2. However, the exact locations of P_1 and P_2 cannot be detected, and the traffic speed of the pedestrian for moving from s to e can be estimated by

$$\text{Speed}_{\text{original}} = \frac{d_{\text{device}}}{TT_{P_1 - P_2}} \qquad (1)$$

where d_{device} is the distance between sensing points s and e, and $TT_{P_1 - P_2}$ is the travel time moving from P_1 to P_2. Obviously, the estimated traffic speed in (1) is biased and the error can be calculated by

$$\text{Speed}_{\text{error}} = \frac{d_{\text{device}} - d_{P_1 - P_2}}{TT_{P_1 - P_2}} \qquad (2)$$

where $d_{P_1 - P_2}$ is the distance between P_1 and P_2. The magnitude of estimated speed errors depends on the size of the detection range. Sometimes, the detection range can be enlarged to 1000 m by powerful antenna in order to capture the MAC address from the vehicles driving with high speed [26].

For each detected Wi-Fi and Bluetooth probe request frame, an integer ranging from -120 to -30, which is called RSSI, tells the signal strength of a Wi-Fi or Bluetooth signal. In previous studies, distance between sensors and Wi-Fi or Bluetooth mobile device has been demonstrated as one of the major influential factors of RSSI [50]. The developed function which maps RSSI to distance was utilized for indoor localization [51], [52].

In this study, the RSSI is utilized to correct the biased traffic speed estimation based on Wi-Fi and Bluetooth sensing data. To explore the relationship between RSSI and distance, the experiments are conducted to collect RSSI values with different distance. Fig. 3 shows the boxplot of RSSI measurements to distance for Wi-Fi and Bluetooth signal separately. According to the figures, the detection range for Wi-Fi and Bluetooth sensing are 60 and 35 m, respectively. It is obvious that the values of RSSI increase while the mobile device and the sensor are getting closer. The relationship is fitting by an exponential regression since the minimum value of distance is zero. The dashed red lines in Fig. 3 present the fitted relationship. The R-square is larger than 0.95 for both
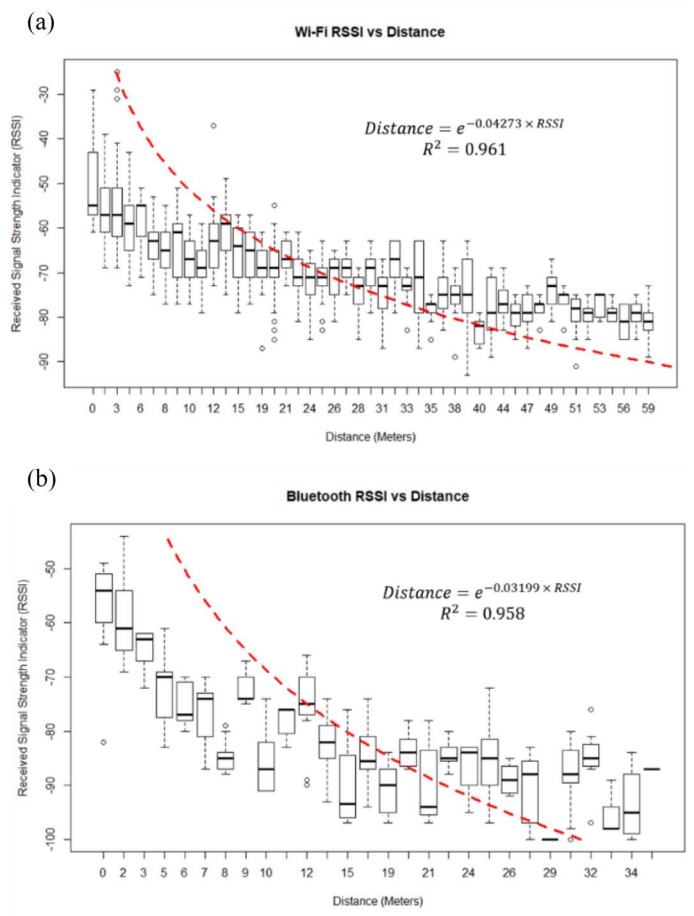
Wi-Fi and Bluetooth sensing, which indicates the well-fitting performance of the regression functions. Then, the distance to sensors can be estimated by (3) and (4) based on RSSI measurements

$$d_{\text{Wi-Fi}} = e^{-0.04273 \times \text{RSSI}} \tag{3}$$

$$d_{\text{Bluetooth}} = e^{-0.03199 \times \text{RSSI}} \tag{4}$$

where $d_{\text{Wi-Fi}}$ and $d_{\text{Bluetooth}}$ are the distance to sensors for Wi-Fi and Bluetooth signal.

To correct the errors based on RSSI, the relative position of detection locations and sensors is another required information. Fig. 4 illustrates the relative position when a mobile device is detected by once, twice and n times. In Fig. 4, Scenarios 1-1 and 1-2 present the case when a mobile device is detected once while passing the detection range, where P_1 is the detection point and d_{P1} is the distance between the sen- sor and P_1. In this case, P_1 is either detected before or after the traveler passing the sensor, and the relative position can- not be determined. When a mobile device has two detection points while passing the detection range, there are four scenar- ios presented in Scenario 2-1 through Scenario 2-4. If P_1 and P_2 are detected in chronological order and $\text{RSSI}_{P_1} < \text{RSSI}_{P_2}$, the relative position of P_1 to the sensor can be determined as that P_1 is detected before the mobile device passing the sensor and vice versa. Thus, in general, if a mobile device is
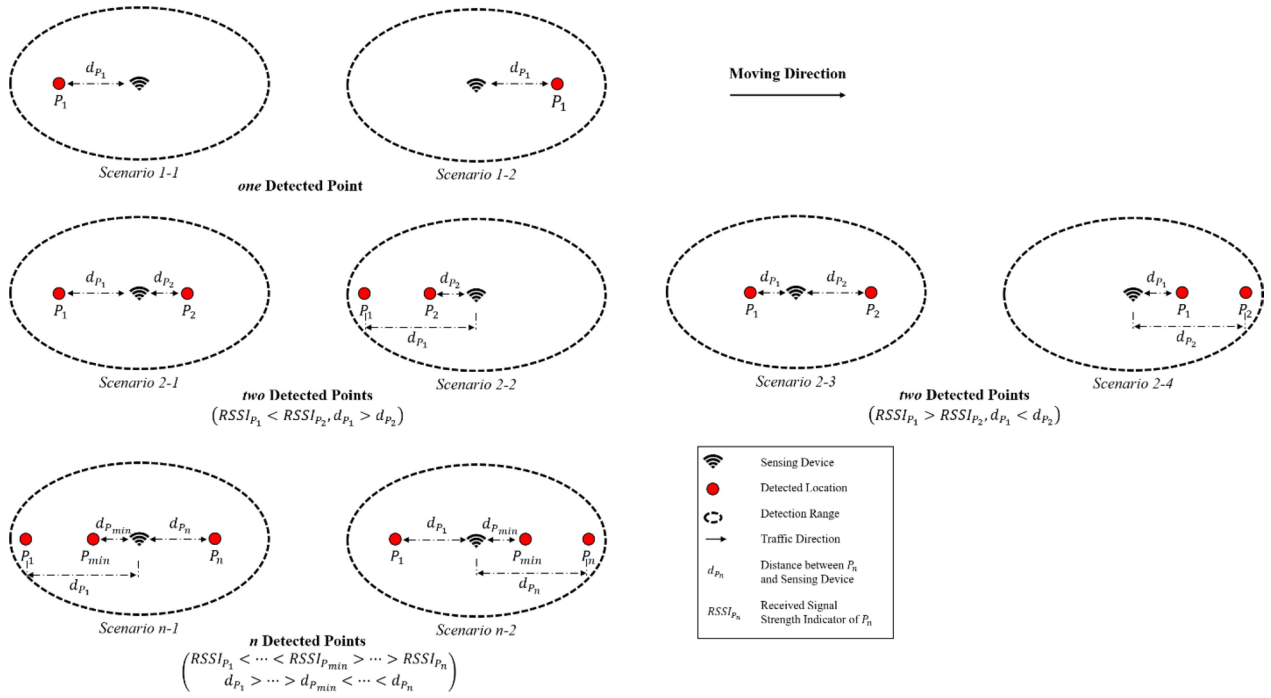
Fig. 4. Illustration of relative position of detection locations and sensors.

detected n times, the detection point with the highest RSSI measurement has the closest distance to the sensor. The relative position of the rest of detection points to the sensor can be determined by the following rules.

1) The detection points are detected before the timestamp of the detection point with the highest RSSI measurements that are detected before the mobile device passing the sensor.

2) The detection points are detected after the timestamp of the detection point with the highest RSSI measurements that are detected after the mobile device passing the sensor.

Thus, if a mobile device is detected more than once either at the starting or ending sensor location, the estimated traffic speed of a single pair of detection points at start and end sensing locations, s and e, can be calculated by where d_{device} is the distance between start and end sensing locations. P^s and P^e are the detection points at start and end sensing locations. d_s and d_e are the distances between P^s and P^e to the sensor, and TT_{s-e} is the detection time difference between the timestamp of P^s and P^e. The traffic speed can be corrected by add or minus d_s and d_e depends on P^s and P^e are detected before or after the detection point $P_{\min}$ which has the closet distance to sensors due to the highest RSSI measurement. Typically, for a specific MAC address, it can be detected multiple times at a single sensor location. Then, the estimated traffic speed of a MAC trip can be corrected by averaging the corrected speed of all pairs of detection points at start and end sensing locations based on

$$\text{Corrected Speed}_{\text{trip}} = \frac{\sum_{i}^{L_S} \sum_{j}^{L_E} \text{Corrected Speed}_{i-j}}{S \times E} \quad (6)$$

where S and E are the total numbers of detection points at start and end sensing locations, i represents the ith detection point at the start sensing location, and j represents the jth detection point at the end sensing location. As it is discussed above, the proposed algorithm for traffic speed correction based on RSSI measurements is not feasible when a mobile device is detected only once at each sensing location. However, Fig. 5(a) shows the boxplot of detection times of the mobile devices when they are carried by car, bike, and pedestrian. The average detection times of a single mobile device is larger than once for all three modes, which indicates the impacts of only one detection points is trivial, and the proposed traffic speed correction algorithm can be implemented under most scenarios.

D. Travel Mode Identification Using Semisupervised Possibilistic Fuzzy C-Means

To monitor multimodal traffic speed by passively sensing Wi-Fi and Bluetooth mobile devices, the traffic modes of valid

$$\text{Corrected Speed}_{s-e} = \begin{cases} \dfrac{d_{\text{device}} - d_s + d_e}{TT_{s-e}}, & \text{if } P^s = \left\{P_n^s \middle| n \in (\min, n]\right\} \text{ and } P^e = \left\{P_n^e \middle| n \in (\min, n]\right\} \\ \dfrac{d_{\text{device}} + d_s - d_e}{TT_{s-e}}, & \text{if } P^s = \left\{P_n^s \middle| n \in [1, \min)\right\} \text{ and } P^e = \left\{P_n^e \middle| n \in [1, \min)\right\} \\ \dfrac{d_{\text{device}} + d_s + d_e}{TT_{s-e}}, & \text{if } P^s = P_n^s \middle| n \in [1, \min) \text{ and } P^e = P_n^e \middle| n \in (\min, n] \end{cases} \quad (5)$$

$$\left\lfloor \frac{d_{\text{device}} - d_s - d_e}{TT_{s-e}} \right., \qquad \text{if } P^s = \left\{ P^s_{\ n} 1 \middle| n \in (\min, n] \right\} \text{ and } P^e = \left\{ P^e \middle|_t n \in [1, \min) \right\}$$
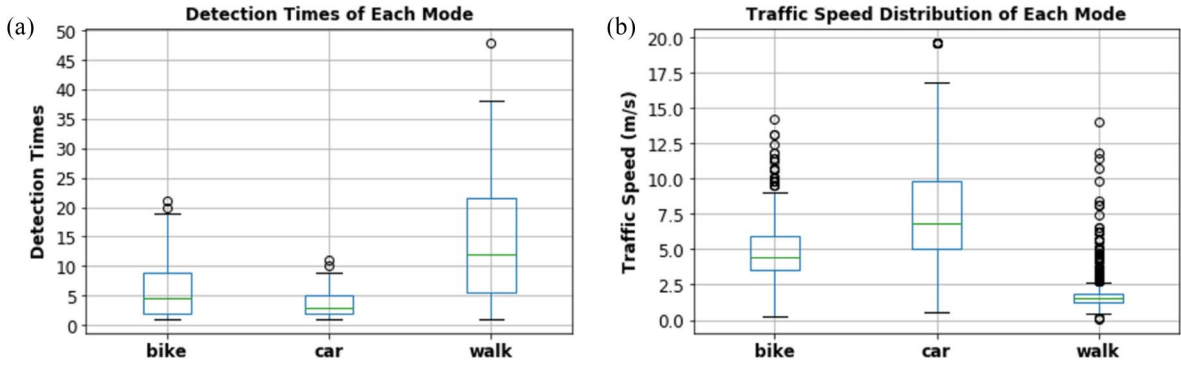
Fig. 5.   Boxplot of key features. (a) Detection times and (b) ground truth of traffic speed of each mod.

Algorithm 2: Semisupervised PCM Clustering

Initialization:
  The number of clusters C
  The maximum number of iterations L
  The fuzzification parameter m
  Calculate the center of each cluster in labelled data $v_i^{(0)}$
  Initialize η_i using Equation (13)
  Initialize $U^0 \in R^{C \times N}$, $u_i^0 = U^0(i,j) \in [0, 1]$ using
  Equation (11)

Repeat:
  Update U^t using (10), and Increment L

Until: $kU^t - U^{t-1}k \le \varepsilon$ or $t \ge L$

MAC trips need to be identified based on the features of each trip. In this study, the traffic mode of each MAC trip is identified by the proposed semisupervised PCM cluster- ing algorithm. The proposed algorithm is introduced in the following sections.

1) Fuzzy C-Means Clustering: Fuzzy C-Means (FCM) clustering is a widely used fuzzy-based clustering algorithm [53]. Comparing to traditional hard clustering algorithms, e.g., K-Means clustering, FCM assigns a certain membership function of all clusters to each data point which allows the ambiguous boundaries among the features of

different clusters [54]. The objective function of FCM and the constraints for assigned membership functions of clusters

are shown in the following:

$$J_{FCM} = \sum_{j=1}^{N} \sum_{i=1}^{C} u_{ij}^m d_{ij}^2$$

$$u_{ij} \in [0, 1] \text{ for all } i \text{ and } j \quad (7)$$

$$0 < \sum_{j=1}^{N} u_{ij} < N \text{ for all } i, \text{ and}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \text{ for all } j. \quad (8)$$

points, and C is the total number of clusters. It is noted that the last constraints in (8) restricts the memberships to lie on the hyperplane defined by $\sum_{i=1}^{C} u_{ij} = 1$. For the traffic modes identification, this constraint is too restrictive and lead to the memberships only represent relative numbers dependent on C rather than the real possibility of a data point belonging to clusters. Thus, the PCM clustering algorithm was developed to improve the algorithm by releasing this constrain. The detailed introduction of PCM clustering is presented in the next section.

2) Possibilistic Fuzzy C-Means Clustering: PCM clus- tering was developed by releasing the constraint of $\sum_{i=1}^{C} u_{ij} = 1$ [55]. The constrains of PCM is shown in the following:

$$u_{ij} \in [0, 1] \text{ for all } i \text{ and } j$$

$$0 < \sum_{j=1}^{N} u_{ij} \le N \text{ for all } i, \text{ and}$$

$$\max_i u_{ij} > 0 \text{ for all } j. \quad (9)$$

where the memberships of each cluster are restricted by the constrain of $\max_i u_{ij} > 0$. In this case, the membership functions represent the degree of possibility of a data point belonging to clusters. The objective function of PCM is shown in (10)

$$J_{PCM} = \sum_{i=1}^{C} \sum_{j=1}^{N} u_i^m d_{ij}^2 + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} \left(1 - u_{ij}\right)^m \quad (10)$$

where η_i are suitable positive numbers which determine the distance at which the values of the membership function of a data point in a cluster becomes 0.5. The selection of its value will be introduced later. The interpretation of other parameters is the same as those of the FCM clustering. The solution of the objective function can be achieved by

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_i}\right)^{\frac{1}{m-1}}}. \quad (11)$$

where $u_{ij}\in[0, 1]$ is the membership function of the jth data point belonging to cluster i, d_{ij}^2 is the distance of the jth data point to the center of the ith cluster, m is the parameter for controlling the fuzzification, N is the total number of data

It is obvious that the membership function of PCM satisfies the constrains in (9). In each iteration, the updated value of u_{ij} depends only on the distance between the center of the ith cluster and the jth data point. The membership of a data point in a cluster should be determined solely by how far it is

from the center of the cluster and should not be coupled with its location with respect to other clusters. Then, the solution of the objective function of PCM clustering allows optimal membership solutions to lie in the entire unit hypercube rather than restricting them to the hyperplane given by $\sum_{i=1}^{L_C} u_{ij} = 1$. For the selection of the value of η_i, it should represent the possibility distribution for each cluster. The typical selection of η_i is shown in the following:

$$\eta_i = K \frac{\sum_{j=1}^{L_N} u_{ij}^m d_{ij}^2}{\sum_{j=1}^{L_N} u_{ij}^m} \qquad (12)$$

where K is a constant number which is typically selected to be 1 [55]. Equation (12) makes η_i proportional to the average fuzzy intracluster distance of clusters. The initialized value of η_i depends on the initialization of u_i^m and the center of clusters by randomly selection. However, random selection of the initialized parameters can generate unexpected errors to clustering accuracy [56]–[58]. In this study, a semisupervised PCM clustering is proposed to mitigate the errors utilizing a small set of labeled data.

3) Semisupervised PCM Clustering: In (11) and (12), η_i and u_{ij} are iteratively updated until the converging criteria are met. To utilize the prior information of labeled data, the equation of η_i is designed as in (13) for semisupervised PCM clustering

$$\eta_i = \frac{\sum_{x_j \in (\text{labelled})}^{L_{n_{\text{labelled}}}} d_{ij}^2}{n_{\text{lablled}}} \qquad (13)$$

where $x_j \in (\text{labelled})$ are the data points in the labeled dataset and n_{labelled} is the total number of data points in labeled datasets. In this case, the memberships and u_{ij} and η_i can be initialized by the labeled data. The semisupervised PCM clustering algorithm is trained by Algorithm 2.

E. Results Evaluation

To evaluate the performance of traffic speed correction based on RSSI and the final estimation of multimodal traffic speed estimation, mean absolute error (MAE), mean square error (MSE), and mean absolute percentage error (MAPE) are used to compare the accuracy of corrected traffic speed and original speed. The following equations present the metrics formulation:

$$\text{MAE} = \frac{\sum_{i=1}^{L_N} |\hat{Y_i} - Y_i|}{N} \qquad (14)$$

$$\text{MSE} = \frac{\sum_{i=1}^{L_N} (\hat{Y_i} - Y_i)^2}{N} \qquad (15)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\hat{Y_i} - Y_i}{Y_i} \right| \times 100\% \qquad (16)$$

(a)

(b)

(c)

Fig. 6. System deployment. (a) Wi-Fi and Bluetooth sensing device. (b) Installation. (c) Data storage and analysis server.

MAPE usually expresses accuracy as a percentage. The model with a smaller value of MAE, MSE, and MAPE performs better in the prediction of observed data.

The performance of the proposed semisupervised PCM clustering algorithm for identifying traffic modes is evaluated by comparing the accuracy of the selected unsupervised and semisupervised clustering algorithms, including *K*-Means [59], Constrained *K*-Means [60], FCM clustering [53], semisupervised FCM [61], and PCM clustering [55]. The evaluation metrics are presented in the following:

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \qquad (17)$$

where TP is short for True Position which is the number of MAC trips belonging to traffic mode *i* with correctly assigning traffic mode *i*, and FN is short for false negative, which is the number of MAC trips belonging to traffic mode *j* with wrongly assigning traffic mode *i*.

III. SYSTEM DESIGN AND DATA COLLECTION

A. System Deployment

A prototype system based on the designed framework was established in this study for the data collection purpose. Fig. 6 shows the customized Wi-Fi and Bluetooth sensor, and the remote server of the prototype system. The detailed description of the customized sensing device can be found in our previous study [47]. Typically, sensors are installed at road intersections and get power supply from the roadside cabinets. However, the metal shell of cabinets generates sever influ-

where $\hat{Y}_i$ is the estimated average traffic speed for time window i, Y_i is the ground-truth value, and N is the number of time windows in testing data set. Typically, the MAE presents a measure of the average misprediction of the model, the MSE is used to measure the error associated with a prediction, and the ence for Wi-Fi and Bluetooth signal communicating due to signal shielding. Thus, Wi-Fi and Bluetooth antenna should be extended to the outside of the cabinet with waterproof measures. The sensor installation is shown in Fig. 6(b). The Wi-Fi and Bluetooth antennas are put in a waterproof box

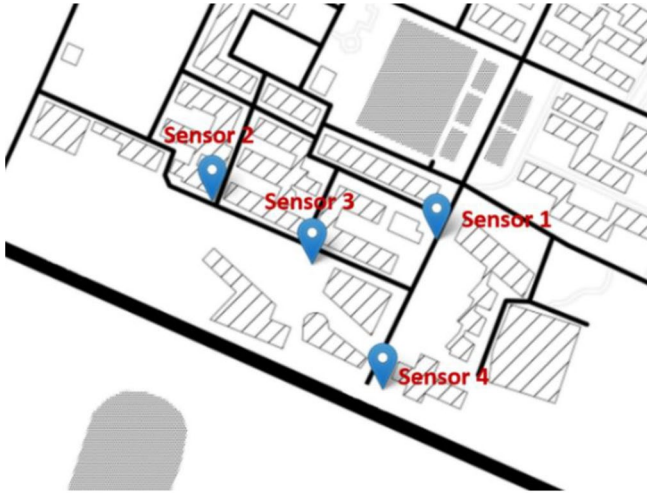| Sensing Points	Data Points					Unique MAC				
	1	2	3	4	Total	1	2	3	4	Total
Wi-Fi	22074	4136	6185	45820	78215	6596	1233	2193	14241	24263
Bluetooth	4771	3182	10485	10273	28711	602	395	422	914	2333
Total	26845	7318	16670	56093	106926	7198	1628	2615	15155	26596

Fig. 7.   Study Site.

which is attached to the side of the cabinet. The remote server of the prototype system is a general PC with data listening, managing, analyzing, and visualizing programs deployed.

B. Data Collection

The data used in this study were collected by four Wi-Fi and Bluetooth sensors at Tongji University in Shanghai City, China. The sensors' locations are shown in Fig. 7. Besides the Wi-Fi and Bluetooth MAC address data, the ground truth of traffic speed data and traffic mode data was collected for the validation purpose. 4-h ground truth of multimodal traffic speed was collected by video camera from 16 P.M. to 18 P.M. on June 2, 2019 and June 3, 2019. The MAC address data collecting in the same time periods are used for the analysis which is shown in Table II.

In Table II, the number of data points and the number of unique MAC address collected by each sensor are presented. In total, 106 926 data points and 26 596 unique MAC addresses were detected by four sensors within the 4-h data collection period. Among them, 24 263 unique MAC addresses were detected through Wi-Fi and 2333 were collected via Bluetooth. The data volume of Sensor 4 is the most among four sensors and Sensor 2 obtained the least amount of data.

In addition, the true traffic mode of 408 MAC trips were collected by ten volunteers. They traveled back and forth among four sensors by different traffic modes. Meanwhile, the high-resolution GPS trajectories of their movements were collected by a phone app to calculate the ground-truth traffic speed of

TABLE III
NUMBER OF MAC TRIPS WITH TRAFFIC MODE LABELS

Traffic Modes	Car	Bike	Walk	Total
Wi-Fi	52	42	66	160
Bluetooth	70	108	70	248
Total	122	150	136	408

each trip for evaluating the proposed algorithm of correcting traffic speed based on RSSI measurements. Table III shows the number of MAC trips with traffic mode labels for each mode.

IV. EXPERIMENTAL RESULTS

A. Traffic Speed Correction Based on RSSI

To evaluate the proposed algorithm for correcting estimated traffic speed based on RSSI measurements, the estimation accuracy is calculated based on the estimated traffic speed before and after the correction. MSE, MAE, and MAPE are utilized as the evaluation metrics. The original speed is calculated using the data points with the highest RSSI value at the end or start sensing location. The corrected speed by RSSI is estimated based on the proposed method describing in Section II-C. The evaluation results are presented in Table IV. According to the evaluation results, the proposed algorithm for correcting traffic speed is highly effective for improving the estimation accuracy. The estimated traffic speed of all three traffic modes is closer to the ground truth after the correction based on RSSI measurements. All three metrics decrease a lot comparing with the original estimated speed. Among three traffic modes, the walk mode achieved the most accurate estimation, in which the accuracy is about 95%, since the relative more detection points of each trip provide more information about the locations of travelers within the detection range. The estimated speed of the car mode and bike mode after correction also performed well, in which the accuracy is about 85%.

B. Travel Mode Identification

In this section, the evaluation result of traffic mode identification is presented. Totally, the data of 408 labeled MAC trips were used for evaluating traffic mode identification algorithms. The total data set was randomly split into two parts with the ratio of 4:6 as training data and testing data. The evaluation metrics were calculated only using testing data. For unsupervised clustering algorithms, the data labels of training data were not used as the input of the algorithms. For semisupervised clustering algorithms, 40% of training

PERFORMANCE EVALUATION OF CORRECTING TRAFFIC SPEED BY RSSI

Mode	Car		Bike		Walk	
	Original Speed	Corrected Speed by RSSI	Original Speed	Corrected Speed by RSSI	Original Speed	Corrected Speed by RSSI
MSE	7.1854	**1.7194**	1.3099	**0.9208**	0.0986	**0.0121**
MAE	2.4608	**1.1672**	0.8898	**0.7160**	0.2421	**0.0863**
MAPE	0.2909	**0.1511**	0.2005	**0.1626**	0.1869	**0.0663**

TABLE V
PERFORMANCE EVALUATION OF TRAVEL MODE IDENTIFICATION

K-Means		Predicted			Recall
		Walk	Bike	Car	
Ground Truth	Walk	72.06%	26.47%	1.47%	72.06%
	Bike	0.00%	50.00%	50.00%	50.00%
	Car	0.00%	50.00%	50.00%	50.00%

COP K-Means		Predicted			Recall
		Walk	Bike	Car	
Ground Truth	Walk	98.53%	1.47%	0.00%	98.53%
	Bike	0.00%	75.00%	25.00%	75.00%
	Car	0.00%	6.56%	93.44%	93.44%

FCM		Predicted			Recall
		Walk	Bike	Car	
Ground Truth	Walk	85.29%	13.24%	1.47%	85.29%
	Bike	0.00%	39.33%	60.67%	39.33%
	Car	0.00%	1.64%	98.36%	98.36%

Semi-FCM		Predicted			Recall
		Walk	Bike	Car	
Ground Truth	Walk	98.53%	1.47%	0.00%	98.53%
	Bike	0.00%	78.67%	21.33%	78.67%
	Car	0.00%	1.64%	98.36%	98.36%

PCM		Predicted			Recall
		Walk	Bike	Car	
Ground Truth	Walk	85.29%	13.24%	1.47%	85.29%
	Bike	0.00%	63.63%	36.38%	63.62%
	Car	0.00%	0.82%	99.18%	99.18%

Semi-PCM		Predicted			Recall
		Walk	Bike	Car	
Ground Truth	Walk	99.18%	0.19%	0.00%	**98.53%**
	Bike	0.00%	82.67%	17.33%	**82.67%**
	Car	0.00%	0.82%	99.18%	**99.18%**

data was randomly selected as labeled data. The labeled training data are used to initialize the hyperparameters, e.g., η_i in Semi-PCM. The experiment was repeated five times. The presented result is the average performance of all five experiments.

Table V is the confusion matrix which shows the evaluation results of the proposed algorithm and other baseline cluster- ing algorithms. The values in the table represent how much percentage trips in a specific traffic mode are assigned to the traffic mode. For example, 72.06% in the first row of the *K*-Means table represent 72.06% trips by walking are assigned to walk mode. According to the results, the proposed semisu- pervised PCM clustering algorithm outperformed all other baseline algorithms in terms of the highest accuracy of traffic mode identification for all three traffic modes. The average identification accuracy of the walk mode and car mode is closer to 100%. For the bike mode, there are about 17% trips were identified as car mode. The major reason is that the traffic speed of car mode is relative lower than the usual condition due to the lower speed limits on the university campus. As shown in Fig. 5(b), the third quantile of the traffic speed of car mode is about 10 m per second. In the urban area, the average traffic speed is usually 14 m per second or higher. Such a low traffic speed distribution of car mode made the features of car mode and bike mode similar.

For other clustering algorithms, *K*-Means, FCM clustering, and PCM clustering performed with the accuracy in ascend- ing order. The main reason is that the mechanism of the FCM clustering algorithm allows the features space of clusters to

be partially overlapped, and PCM clustering improves the membership function by reflecting the essential of "possibil- ity." It should be noticed that all clustering algorithms trained by semisupervised learning performed better than they were trained by the unsupervised learning strategy. As it is dis- cussed in the section of methodology, the better performance is attributed to avoiding random initialization based on the prior information in the labeled dataset.

C. Multimodal Traffic Speed Estimation

After the traffic mode of each MAC trip is identified, the average traffic speed of each road segment within a predefined time window can be calculated. 15 min was selected for evaluating the estimated multimodal traffic speed. The 4-h time period from 16 P.M. to 18 P.M. on June 2, 2019 and June 3, 2019, was divided into 16 time windows. The ground truth of multimodal traffic speed was calculated based on the video recording. The trips in video data were extracted by manually identification. In this study, multimodal traffic speed was esti- mated for the road segments between two adjacent sensors. The road segments with sensing points in the middle were not considered. Actually, if the traffic modes can be identi- fied accurately based on the data collecting by two sensing points, adding more sensing points in the middle of road seg- ments will definitely make the estimated accuracy higher [35]. In the 16 time windows, not every time window had valid trips of every traffic mode for each road segment. Thus, the eval- uation results are calculated only based on the time windows
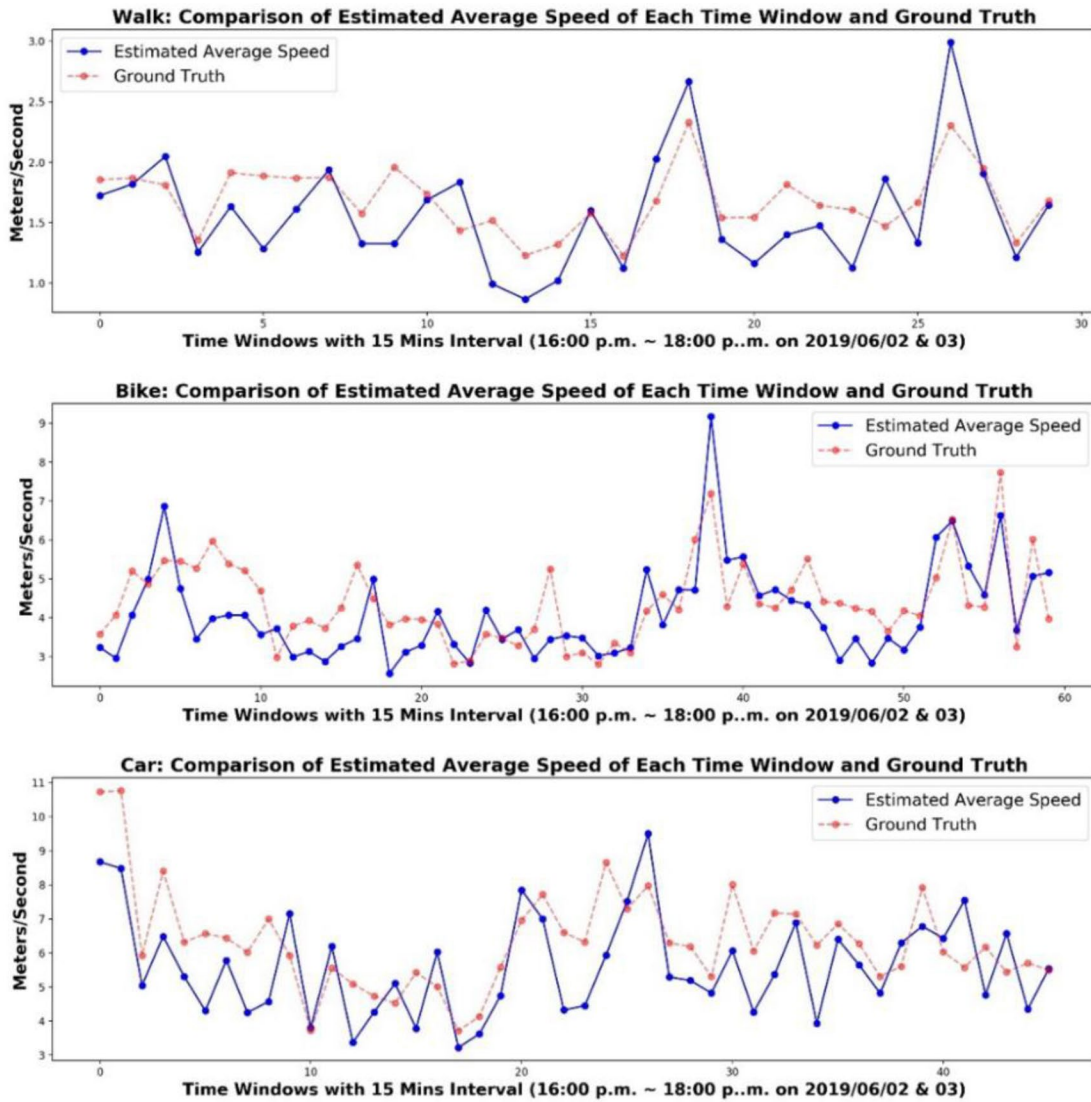
Fig. 8. Comparison of estimated multimodal traffic speed and ground truth.

TABLE VI
PERFORMANCE EVALUATION OF ESTIMATED MULTIMODAL TRAFFIC SPEED

Mode	Car		Bike		Walk	
	Original Speed	Corrected Speed by RSSI	Original Speed	Corrected Speed by RSSI	Original Speed	Corrected Speed by RSSI
MSE	2.2898	**1.5235**	1.0612	**0.6867**	0.1383	**0.0943**
MAE	1.2768	**1.0507**	0.8688	**0.7094**	0.3258	**0.2537**
MAPE	0.2001	**0.1594**	0.1926	**0.1606**	0.1947	**0.1519**

with valid measurements of both the ground truth and estimated multimodal traffic speed. Fig. 8 shows the comparison of the ground truth and the estimated multimodal traffic speed for the time windows with 15 min interval. In the figures, the solid blue lines present the estimated traffic speed of three modes and the dashed red lines show the ground truth. During the 4-h time period, the traffic speed of three modes fluctuated. The estimated multimodal traffic speed is highly close to the ground truth and reflect the flotation very well. The evaluation metrics are shown in Table VI.

In Table VI, the evaluation metrics calculated by the estimated speed before and after correcting based on RSSI measurements are presented. According to the evaluation results, traffic speed correcting based on RSSI improved the estimation accuracy a lot for all three traffic modes. The values of the evaluation metrics calculated by the corrected multimodal speed is largely reduced compared with those calculating by the original multimodal traffic speed. The overall traffic speed estimation accuracy is around 85% for all three traffic modes.

V. Conclusion

This study proposed a real-time multimodal traffic speed monitoring system based on passive Wi-Fi and Bluetooth sensing technology. An algorithm was established to correct the estimated traffic speed based on RSSI measurements. The traffic mode of each trip is identified by the proposed semisupervised PCM clustering algorithm. The performance of the proposed system was evaluated based on the ground-truth data. The evaluation results indicated the effectiveness and accuracy of the proposed system. The future research direction includes urban mobility recognition based on the multimodal traffic status estimating by Wi-Fi and Bluetooth sensing data. In addition, the communication range of 5G and 6G is much shorter than 3G/4G communication [62]–[64], the street-level movements probably can be reflected by phone communication data. Thus, exploring the feasibility of utilizing the cell phone signal of 5G and 6G for traffic speed monitoring is also one of the future research directions.

References

[1] C.-H. Wu, C.-C. Wei, D.-C. Su, M.-H. Chang, and J.-M. Ho, "Travel time prediction with support vector regression," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, vol. 2. Shanghai, China, 2004, pp. 1438–1442, doi: 10.1109/ITSC.2003.1252721.

[2] J.-D. Schmöcker, S. Ahuja, and M. G. H. Bell, "Multi-objective signal control of urban junctions–Framework and a London case study," *Transp. Res. C, Emerg. Technol.*, vol. 16, no. 4, pp. 454–470, 2008.

[3] C. Nanthawichit, T. Nakatsuji, and H. Suzuki, "Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway," *Transp. Res. Rec.*, vol. 1855, no. 1, pp. 49–59, 2003.

[4] E. Jenelius and H. N. Koutsopoulos, "Travel time estimation for urban road networks using low frequency probe vehicle data," *Transp. Res. B, Methodol.*, vol. 53, pp. 64–81, Jul. 2013.

[5] M. Chen and S. I. J. Chien, "Dynamic freeway travel-time prediction with probe vehicle data: Link based versus path based," *Transp. Res. Rec.*, vol. 1768, no. 1, pp. 157–161, 2001.

[6] D. Vij and N. Aggarwal, "Smartphone based traffic state detection using acoustic analysis and crowdsourcing," *Appl. Acoust.*, vol. 138, pp. 80–91, Sep. 2018.

[7] Y. Wang and N. L. Nihan, "Can single-loop detectors do the work of dual-loop detectors?" *J. Transp. Eng.*, vol. 129, no. 2, pp. 169–176, 2003.

[8] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015, doi: 10.1016/j.trc.2015.03.014.

[9] B. Abdulhai and S. M. Tabib, "Spatio-temporal inductance-pattern recognition for vehicle re-identification," *Transp. Res. C, Emerg. Technol.*, vol. 11, nos. 3–4, pp. 223–239, 2003.

[10] X. Liu, W. Liu, T. Mei, and H. Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 869–884.

[11] S. Wang, Z. Pu, Q. Li, Y. Guo, and M. Li, "Edge computing-enabled crowd density estimation based on lightweight convolutional neural network," in *Proc. IEEE Int. Smart Cities Conf. (ISC2)*, Manchester, U.K., 2021, pp. 1–7.

[12] J. Pérez, F. Seco, V. Milanés, A. Jiménez, J. C. Díaz, and T. De Pedro, "An RFID-based intelligent vehicle speed controller using active traffic signals," *Sensors*, vol. 10, no. 6, pp. 5872–5887, 2010.

[13] 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.

[14] A. Haghani, M. Hamedi, K. F. Sadabadi, S. Young, and P. Tarnoff, "Data collection of freeway travel time ground truth with Bluetooth sensors," *Transp. Res. Rec.*, vol. 2160, no. 1, pp. 60–68, 2010.

[15] 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, 2017.

[16] C. Slobogin, "Public privacy: Camera surveillance of public places and the right to anonymity," *Mississippi Law J.*, vol. 72, p. 213, Feb. 2003.

[17] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, 2013, pp. 3318–3325.

[18] Y. Wang, Y. Malinovskiy, Y.-J. Wu, U. K. Lee, and M. Neeley, "Error modeling and analysis for travel time data obtained from Bluetooth MAC address matching," Dept. Civil Environ. Eng., Univ. Washington, Seattle, WA, USA, Rep. Agreement T4118 Task 46, 2011.

[19] J. S. Wasson, J. R. Sturdevant, and D. M. Bullock, "Real-time travel time estimates using media access control address matching," *ITE J.*, vol. 78, no. 6, pp. 20–23, 2008.

[20] B. N. Araghi, J. H. Olesen, R. Krishnan, L. T. Christensen, and H. Lahrmann, "Reliability of Bluetooth technology for travel time estimation," *J. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 240–255, 2015.

[21] M. Martchouk, F. Mannering, and D. Bullock, "Analysis of freeway travel time variability using Bluetooth detection," *J. Transp. Eng.*, vol. 137, no. 10, pp. 697–704, 2011.

[22] Z. Pu, Q. Zhang, Y. Lv, and Y. Wang, "A device-free Wi-Fi sensing method for pedestrian monitoring using channel state information," in *Proc. Int. Conf. Transp. Develop.*, 2020, pp. 207–220.

[23] Y. Malinovskiy, U.-K. Lee, Y.-J. Wu, and Y. Wang, "Investigation of Bluetooth-based travel time estimation error on a short corridor," in *Proc. Transp. Res. Board 90th Annu. Meeting*, 2011, p. 19.

[24] T. M. Brennan Jr., J. M. Ernst, C. M. Day, D. M. Bullock, J. V. Krogmeier, and M. Martchouk, "Influence of vertical sensor placement on data collection efficiency from Bluetooth MAC address collection devices," *J. Transp. Eng.*, vol. 136, no. 12, pp. 1104–1109, 2010.

[25] A. Bhaskar and E. Chung, "Fundamental understanding on the use of Bluetooth scanner as a complementary transport data," *Transp. Res. C, Emerg. Technol.*, vol. 37, pp. 42–72, Dec. 2013.

[26] Y. Malinovskiy, Y.-J. Wu, Y. Wang, and U. K. Lee, "Field experiments on Bluetooth-based travel time data collection," in *Proc. Transp. Res. Board 89th Annu. Meeting*, 2010, p. 17.

[27] A. M. Hainen, J. S. Wasson, S. M. L. Hubbard, S. M. Remias, G. D. Farnsworth, and D. M. Bullock, "Estimating route choice and travel time reliability with field observations of Bluetooth probe vehicles," *Transp. Res. Rec.*, vol. 2256, no. 1, pp. 43–50, 2011.

[28] Y. Aliari and A. Haghani, "Bluetooth sensor data and ground truth testing of reported travel times," *Transp. Res. Rec.*, vol. 2308, no. 1, pp. 167–172, 2012.

[29] N. Abedi, A. Bhaskar, E. Chung, and M. Miska, "Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses," *Transp. Res. C, Emerg. Technol.*, vol. 60, pp. 124–141, Nov. 2015.

[30] J. D. Porter, D. S. Kim, M. E. Magaña, P. Poocharoen, and C. A. G. Arriaga, "Antenna characterization for Bluetooth-based travel time data collection," *J. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 142–151, 2013.

[31] A. Bhaskar, M. Qu, and E. Chung, "Bluetooth vehicle trajectory by fusing Bluetooth and loops: Motorway travel time statistics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 113–122, Feb. 2015.

[32] L. Jie, H. Van Zuylen, L. Chunhua, and L. Shoufeng, "Monitoring travel times in an urban network using video, GPS and Bluetooth," *Procedia Soc. Behav. Sci.*, vol. 20, pp. 630–637, Jan. 2011.

[33] C. Bachmann, M. J. Roorda, B. Abdulhai, and B. Moshiri, "Fusing a Bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation," *J. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 152–164, 2013.

[34] H. Park and A. Haghani, "Optimal number and location of Bluetooth sensors considering stochastic travel time prediction," *Transp. Res. C, Emerg. Technol.*, vol. 55, pp. 203–216, Jun. 2015.

[35] S. Yang and Y.-J. Wu, "Travel mode identification using Bluetooth technology," *J. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 407–421, 2018.

[36] D. Schrank, B. Eisele, T. Lomax, and J. Bak, *2015 Urban Mobility Scorecard*, Texas A M Transp. Inst., College Station, TX, USA, 2015.

[37] D. D. Puckett and M. J. Vickich, "Bluetooth⌐-based travel time/speed measuring systems development," TTI, College Station, TX, USA, Rep. UTCM Project #09-00-17, 2010.

[38] W. Qiao, A. Haghani, and M. Hamedi, "A nonparametric model for short-term travel time prediction using Bluetooth data," *J. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 165–175, 2013.

[39] J. Barceló, L. Montero, L. Marqués, and C. Carmona, "Travel time forecasting and dynamic origin-destination estimation for freeways based on Bluetooth traffic monitoring," *Transp. Res. Rec.*, vol. 2175, no. 1, pp. 19–27, 2010.

[40] T. Tsubota, A. Bhaskar, E. Chung, and R. Billot, "Arterial traffic congestion analysis using Bluetooth duration data," in *Proc. 34th Aust. Transp. Res. Forum*, 2011, pp. 1–14.

[41] Z. Mei, D. Wang, and J. Chen, "Investigation with Bluetooth sensors of bicycle travel time estimation on a short corridor," *Int. J. Distrib. Sens. Netw.*, vol. 8, no. 1, 2012, Art. no. 303521.

[42] R. J. Haseman, J. S. Wasson, and D. M. Bullock, "Real-time measurement of travel time delay in work zones and evaluation metrics using Bluetooth probe tracking," *Transp. Res. Rec.*, vol. 2169, no. 1, pp. 40–53, 2010.

[43] C. M. Day, T. M. Brennan, A. M. Hainen, S. M. Remias, and D. M. Bullock, *Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data*. West Lafayette, IN, USA: Purdue Univ., 2012, doi: 10.5703/1288284314988.

[44] J. J. V. Díaz, A. B. R. González, and M. R. Wilby, "Bluetooth traffic monitoring systems for travel time estimation on freeways," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 123–132, Jan. 2016.

[45] A. Lesani and L. Miranda-Moreno, "Development and testing of a real-time WiFi-Bluetooth system for pedestrian network monitoring, classification, and data extrapolation," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1484–1496, Apr. 2019.

[46] Y. Zhou, B. P. L. Lau, Z. Koh, C. Yuen, and B. K. K. Ng, "Understanding crowd behaviors in a social event by passive WiFi sensing and data mining," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4442–4454, May 2020.

[47] Z. Pu, M. Zhu, W. Li, Z. Cui, X. Guo, and Y. Wang, "Monitoring public transit ridership flow by passively sensing Wi-Fi and Bluetooth mobile devices," *IEEE Internet Things J.*, vol. 8, no. 1, pp. 474–486, Jan. 2021.

[48] A. Saeedi, S. Park, D. S. Kim, and J. D. Porter, "Improving accuracy and precision of travel time samples collected at signalized arterial roads with Bluetooth sensors," *Transp. Res. Rec.*, vol. 2380, no. 1, pp. 90–98, 2013.

[49] B. N. Araghi, R. Krishnan, and H. Lahrmann, "Mode-specific travel time estimation using Bluetooth technology," *J. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 219–228, 2016.

[50] W. Xue, W. Qiu, X. Hua, and K. Yu, "Improved Wi-Fi RSSI measurement for indoor localization," *IEEE Sensors J.*, vol. 17, no. 7, pp. 2224–2230, Apr. 2017.

[51] M. Quan, E. Navarro, and B. Peuker, "Wi-Fi localization using RSSI fingerprinting," California Polytech. State Univ., San Luis Obispo, CA, USA, Rep., 2010.

[52] S. Yiu, M. Dashti, H. Claussen, and F. Perez-Cruz, "Wireless RSSI fingerprinting localization," *Signal Process.*, vol. 131, pp. 235–244, Feb. 2017.

[53] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy *c*-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.

[54] R. Xu and D. C. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[55] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.

[56] B. Kulis, S. Basu, I. Dhillon, and R. Mooney, "Semi-supervised graph clustering: A kernel approach," *Mach. Learn.*, vol. 74, no. 1, pp. 1–22, 2009.

[57] S. Basu, M. Bilenko, and R. J. Mooney, "A probabilistic framework for semi-supervised clustering," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2004, pp. 59–68.

[58] S. Basu, A. Banerjee, and R. J. Mooney, "Semi-supervised clustering by seeding," in *Proc. 19th Int. Conf. Mach. Learn.*, 2002, pp. 27–34.

[59] A. Likas, N. Vlassis, and J. J. Verbeek, "The global *k*-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, 2003.

[60] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained K-means clustering with background knowledge," in *Proc. ICML*, vol. 1, 2001, pp. 577–584.

[61] L. Tari, C. Baral, and S. Kim, "Fuzzy c-means clustering with prior biological knowledge," *J. Biomed. Informat.*, vol. 42, no. 1, pp. 74–81, 2009.

[62] M. Shafi *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[63] Y. Zhou, L. Tian, L. Liu, and Y. Qi, "Fog computing enabled future mobile communication networks: A convergence of communication and computing," *IEEE Commun. Mag.*, vol. 57, no. 5, pp. 20–27, May 2019.

[64] Y. Zhou *et al.*, "Service-aware 6G: An intelligent and open network based on the convergence of communication, computing and caching," *Digit. Commun. Netw.*, vol. 6, no. 3, pp. 253–260, 2020.

Ziyuan Pu (Member, IEEE) received the B.S. degree in transportation engineering from Southeast University, Nanjing, China, in 2010, and the M.S. and Ph.D. degrees in civil and environmental engineering from the University of Washington, Seattle, WA, USA, in 2015 and 2020, respectively.

He is a Lecturer (Assistant Professor) with Monash University, Bandar Sunway, Malaysia. His research interests include transportation data science, smart transportation infrastructures, connected and autonomous vehicles, and urban.

Zhiyong Cui received the B.S. degree in software engineering from Beihang University, Beijing, China, in 2012, the M.S. degree in software engineering from Peking University, Beijing, in 2015, and the Ph.D. degree in civil engineering from the University of Washington, Seattle, WA, USA, in 2021.

He is currently an Associate Professor with Beihang University. His research interests include intelligent transportation systems, deep learning, machine learning, and traffic data mining.

Jinjun Tang received the Ph.D. degree in transportation engineering from Harbin Institute of Technology, Harbin, China, in 2016.

He is currently an Associate Professor with the School of Traffic and Transportation Engineering, Central South University, Changsha, China. His research interests include traffic flow prediction, data mining in the transportation systems, and intelligent transportation systems.

Shuo Wang received the B.Sc. degree from Nanjing University of Science and Technology, Nanjing, China, in 2016, where she is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering.

Her research interests include Internet of Things, edge computing security, and deep learning.

Yinhai Wang (Senior Member, IEEE) received the master's degree in computer science from the University of Washington (UW), Seattle, WA, USA, in 2013, and the Ph.D. degree in transportation engineering from the University of Tokyo, Tokyo, Japan, in 1998.

He is a Professor of Transportation Engineering and the Founding Director of the Smart Transportation Applications and Research Laboratory, UW. He also serves as the Director of USDOT University Transportation Center for Federal Region 10 (PacTrans). His active research fields include traffic sensing, e-science of transportation, and transportation safety.