

Research Article

Freeway Multisensor Data Fusion Approach Integrating Data from Cellphone Probes and Fixed Sensors

Shanglu He,¹ Jian Zhang,¹ Yang Cheng,² Xia Wan,² and Bin Ran¹

¹Jiangsu Key Laboratory of Urban ITS, Southeast University, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Collaborative Innovation Center for Technology and Application of Internet of Things, Nanjing, Jiangsu 210096, China

²Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI 53705, USA

Correspondence should be addressed to Jian Zhang; jianzhang@seu.edu.cn

Received 19 May 2016; Revised 1 September 2016; Accepted 5 September 2016

Academic Editor: Eduard Llobet

Copyright © 2016 Shanglu He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Freeway traffic state information from multiple sources provides sufficient support to the traffic surveillance but also brings challenges. This paper made an investigation into the fusion of a new data combination from cellular handoff probe system and microwave sensors. And a fusion method based on the neural network technique was proposed. To identify the factors influencing the accuracy of fusion results, we analyzed the sensitivity of those factors by changing the inputs of neural-network-based fusion model. The results showed that handoff link length and sample size were identified as the most influential parameters to the precision of fusion. Then, the effectiveness and capability of proposed fusion method under various traffic conditions were evaluated. And a comparative analysis between the proposed method and other fusion approaches was conducted. The results of simulation test and evaluation showed that the fusion method could complement the drawback of each collection method, improve the overall estimation accuracy, adapt to the variable traffic condition (free flow or incident state), suit the fusion of data from cellphone probes and fixed sensors, and outperform other fusion methods.

1. Introduction

accurate estimation of traffic speed is important information for the operation and management of freeway traffic. It is also essential for the study of freeway traffic flow theory, that is, modeling of traffic flow dynamics [1, 2]. Kinds of dedicated sensing infrastructure could collect real-time traffic information, but they still have the limitation in full-size coverage. To achieve a complete description of freeway traffic state, the technique of traffic state estimation is developed. Generally, the majority of estimation approaches [3, 4] are based on the data from a single source. Since a broad spectrum of data from multisources is becoming available, data fusion techniques are essential to integrate and translate all available traffic measurements into a consistent picture of the dynamic traffic state [5]. Various previous works contributed to the data fusion and presented various algorithms, that is, convex combination algorithm, Kalman

filter technique, fuzzy integral approach, artificial intelligence technique, and Dempster-Shafer theory. The comparison study indicated that the majority of data fusion approaches perform reasonably well and decrease estimation error [6]. However, it is still difficult to make a judgment on the best fusion technique. Besides, the newly developed collection methods raise new challenges for the traffic data fusion.

The following reasons motivate this study to make a further investigation into the data fusion technique. First, cellular handoff probe system (CHPS) is gaining popularity and has been implemented to collect data in Jiangsu Freeway (China), while the studies about the integration of these measurements and other sensors are rare [7]. Although CHPS takes advantage of large spatial coverage and low cost in the implementation and maintenance, it still suffers from the lower accuracy compared with GPS-based probe methods and fixed detectors. On the other hand, fixed sensor such as microwave sensor has limited spatial coverage. Therefore,

although each of these sensors provides an exclusive stream of traffic surveillance data, the inherent drawbacks of each sensor affect the full-scale traffic surveillance [8]. Since these two collection methods have been implemented on Jiangsu Freeway, it is practical to integrate the two data sources to provide more precise traffic information.

Second, the direct measurements from CHPS and microwave sensor differ in semantics and might contradict each other. For instance, the measurements of the microwave sensors include the count of vehicles and spot speed, while CHPS is capable of extracting travel time from the cellphone communication records, that is, handoff records. Handoff records are a sequence of messages which are recorded when a phone on call moves between two adjacent towers' areas [9, 10]. Thus, a vehicle with an on-board phone on call and being recorded can be regarded as a probe vehicle. By tracking the probe vehicle, the corresponding travel time of the probe could be calculated. Obviously, the semantical meanings of the measurements from these two sensors are different. An even worse situation is that the measurements from different sensors conflict with each other when they should represent the same traffic state.

Another important motivation is that there are other factors besides the traffic measurements from CHPS or microwave sensors that might affect the fusion accuracy. And it is significant to figure out what the influential factor is and how it will affect the fusion results. For fixed detectors, high failure ratio and inaccurate traffic state conversion arithmetic are two types of principal problems [11]. In addition, the ignorance of traffic variance (i.e., shockwave) due to the limited coverage of fixed detectors is also an issue. The factors affecting the precision of CHPS include sample size, handoff accuracy, handoff consistency, and handoff link length.

Based on these motivations, this study primarily attempts to fuse a cellphone handoff probe system with microwave sensors. The remaining sections provide the detailed study and investigation which are organized as follows: The next section is a literature review about the traffic data fusion techniques. The third section provides detailed illustration of the proposed data fusion approach based on neural network. The fourth section is a brief introduction to the simulation model generating the test data. Then, these data were applied to investigate the influential factors and find out the optimal neural network model for data fusion. And an evaluation of the data fusion method is followed. Finally, the paper summarizes the main conclusions.

2. Literature Review

2.1. A Brief Overview of Data Fusion Techniques. Varieties of traffic engineering areas involve the data fusion techniques [12]. This study focused on the application of data fusion techniques in dynamic traffic state estimation. We summarized these techniques into four categories as (1) statistical-based, (2) probabilistic-based, (3) artificial-intelligence-based, and (4) estimation-based [13].

Among statistical techniques, a weighted combination of measurements from different sources is the most common approach. Tarko and Roupail proposed a regression model

for travel time data fusion in the early nineties [14]. Li et al. used a weighted combination to fuse data from probe vehicles and loop detectors for the estimation of the queue length, while the weights were from the covariance based on Kalman filtering [15]. Hellinga and Gudapati concluded that nonlinear regression appeared to be an adequate method of estimating arterial link delay with detector data [8]. Bachmann et al. test the simple convex combination and Bar-Shalom/Campo combination to fuse the Bluetooth system and loop detectors [6, 16, 17]. Because of the simplicity in computing, this kind of methods is widely used in practice. The challenging issue is to determine the appropriate weights for measurements from different collection methods.

Probabilistic approaches, such as Bayesian approach and Dempster-Shafer inference, are widely used. For instance, El Faouzi et al. investigated the application of DS evidence theory to address the travel time estimation, using data from loop detectors and toll collection stations [18]. Based on evidence theory, Kong et al. added an estimating process to dynamic reliability and thus proposed an improved evidence theory to provide real-time traffic state estimation with data from loop detector and taxi GPS [11]. Mengying et al. proposed a recursive-Bayesian inference data fusion model for signalized arterials, which is based on simulated loop detector data, probe data, and historical data [19]. The most critical factor that impacts the effectiveness of evidence theory based fusion techniques is generating a proper liability of measurements from different sources.

Artificial intelligent based method includes expert systems, fuzzy inference engines, and neural networks [13]. Choi applied the framework of fuzzy operator logic to integrate data sources of fixed traffic detectors, CCTVs, and probe vehicles [20]. Cheu et al. used a neural network to fuse the estimations from loop detectors and GPS probes in Singapore [21]. van Lint et al. used neural networks for the prediction of travel times with gaps in the data, obtaining satisfactory results in spite of partial information [22]. For this kind of fusion method, adequate training data could generate a more robust fusion model and thus provide accurate results. On the other hand, more training data adds the cost of calculation time and memories. Therefore, a balance between the fusion accuracy and the time-consuming problem should be done.

Estimation-based method includes recursive approaches, such as Kalman or particle filter. Kalman filter has been used in the traffic state estimation for decades [3, 4, 23]. There are some attractive developments for the Kalman filter in data fusion. For example, based on the traffic flow theory, Lint and Hoogendoorn proposed an extended generalized Treiber-Helbing filter for the fusion of any data sources [13] and results indicated the method was capable of reconstructing accurate traffic condition. Byon et al. combined multiple data sources to estimate the current traffic state using a single-constraint-at-a-time (SCAAT) Kalman filter [24]. Bachmann et al. also investigated the fusion method of SCAAT Kalman filter [6, 16, 17]. It is critical for Kalman filter based method or its extended versions to use a proper traffic flow model, that is, the relation model of traffic state variables. Besides, the difficulty exists in determining the initial inputs, boundary variables, and so forth.

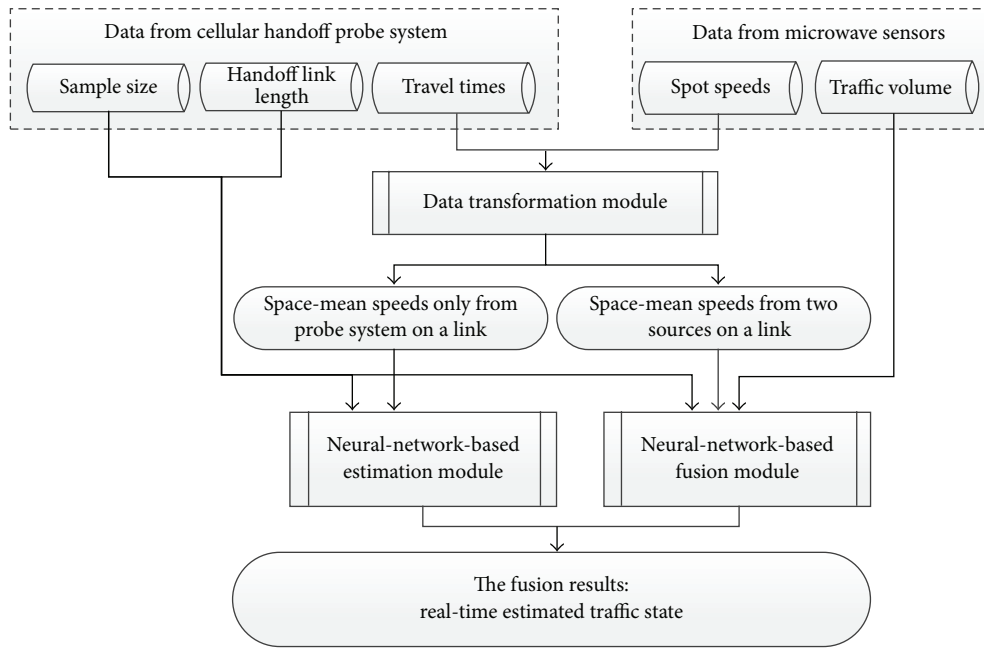


FIGURE 1: The flowchart of data fusion process.

There are other valuable works about making a comparison or summarizing different data fusion techniques. For example, Choi and Chung used three different techniques to fuse and estimate the link travel time of urban road network [25]. And their results indicated that all these fusion methods were validated to outperform the simple arithmetic mean of each traffic source, using data from field experimental GPS probes and detectors. A comprehensive overview paper, written by Faouzi, described the amalgamation problem of traffic data from various sources in road traffic problems [12]. For the estimation of freeway traffic speed, Bachmann et al. conducted a comparative assessment of multisensor data fusion techniques, that is, distribution fusion techniques (simple convex combination and Bar-Shalom/Campo combination), Kalman filter techniques, ordered weighted averaging, fuzzy integral, and artificial neural network [6]. It can be inferred that data fusion is promising and the majority of the techniques improve the accuracy compared with a single sensor. However, before any widespread deployment of data fusion in the field, there are some challenges, such as the accuracy necessary for the effective application, dynamic aspect of data quality [12], and the fusion of data from newly developed sensors. So it is still a meaningful task to improve the data fusion method.

2.2. Motivations. Although data fusion in traffic state estimation has been tackled with various techniques, the preferred fusion approach still has some critical difficulties. These existing works motivated us to optimize the data fusion method in the following aspects. First, a new combination of data collection techniques, that is, cellular handoff probe system and microwave sensors, will be investigated in this study. Although data from loop detectors, GPS probe system, toll collection stations, or Bluetooth probe system have been

widely used in the data fusion study, different data collection techniques would provide distinguished data quality and have different precision. Therefore, it is interesting to investigate the combination of CHPS and microwave sensors. Second, dynamic aspect of data quality will be investigated. The sensitivity of different factors that give an indication of the precision of sensors will be investigated. For instance, it is important to find out the influence of probe vehicle sample size [21] and handoff link length from CHPS and traffic volume from microwave sensors on the fusion results. Third, to achieve a more accurate estimation of dynamic traffic state, this paper will focus exclusively on the improvement of neural-network-based data fusion model. As summarized in the comparative result of Bachmann et al., neural networks sometimes outperform some other data fusion methods [6]. Another reason for choosing the neural network is that it can take into account other influential factors as the fusion inputs besides the traffic state measurements, while these factors indicate the precision of sensors and might have impact on the fusion. Therefore, it is meaningful to improve the neural network model and evaluate the accuracy and adaptability of the neural network model for new data sources. Moreover, the CHPS is becoming an important data source in the practical application of Jiangsu Freeway, while fixed sensors are still in use. Although this study uses a freeway segment in Jiangsu, the method can be applied to other places.

3. Methodology

Figure 1 illustrates the data flow in this study of fusion process and the main components in the fusion, including data transformation module, neural-network-based estimation module, and neural-network-based fusion module. The original inputs include traffic information (e.g., travel time from

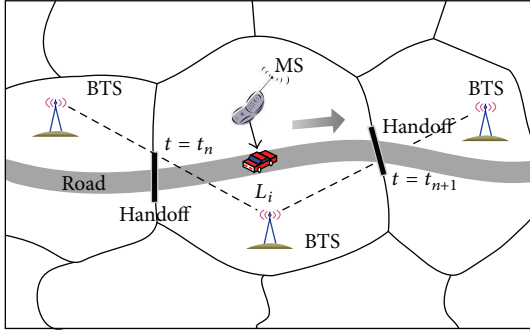


FIGURE 2: Use handoff system to track vehicle [7].

CHPS and spot speed from microwave sensors) and other potential influential information (i.e., the length of handoff link, cellphone probe sample size, and traffic counts from microwave sensors). First, the data transformation module converts the direct traffic measurements from heterogeneous sources into a consistent space-mean speed in the same time interval. Then data will be separated and transferred to different fusion module based on whether there are one or two data sources on the link. The neural-network-based estimation module will estimate the state of links with information and transformed space-mean speed only from CHPS, while the neural-network-based fusion module integrates the available data from two sources. The combination of these two modules generates a composite space-mean speed estimation for each link. The following subsections describe the detailed methodologies in each module.

3.1. Data Transformation Module. The data transformation module is proposed to synchronize the traffic measurements from multisource data. Specifically, spot speeds from microwave sensors are converted to space-mean speed (SMS) of a freeway link in a time interval. Similarly, the locations and time stamps of cellular handoff probe system are also processed to extract the space-mean speed.

3.1.1. SMS from Cellular Handoff Probe System. When a cellphone is turned on and in use (i.e., for a call or text message), all the location-related records of this cellphone can be retrieved from real-time cellphone communication data from cell towers [9]. Figure 2 shows an example of using handoff system to track a vehicle. Each hexagon in Figure 2 is a cell with a base transceiver station (BTS) in the center. When a vehicle with an in-vehicle phone (Mobile Equipment, ME) is on call, a Base Station Controller (BSC) hands off the phone call from one cell to another new cell to keep the cellphone connected. The handoff is performed so quickly that user usually never notices, and the BSC records each handoff once it occurs as cellphone communication data. The recorded information includes anonymous cellphone number, cell ID, handoff timestamp, and event ID. Then the space-mean speed of the roadway segments between two successive handoff points can be obtained from the following equation [7]:

$$\bar{v}_{cj} = \frac{L_i}{t_{n+1} - t_n}, \quad (1)$$

where L_i is the length of handoff link i which is a freeway link between two adjacent handoffs; \bar{v}_{cj} is the space-mean speed of cellular probe vehicle j ; t_n and t_{n+1} are the handoff timestamps as shown in Figure 2.

There are several dynamic factors proved to have an influence on the precision of CHPS, that is, sample size, the length of handoff link, and so forth. Obviously, as a probe technique, the sample size of CHPS is also a time-varying variable that affects the accuracy of CHPS. The length of handoff link is determined by the coverage of the tower signal. However, the signal of cell towers varies depending on the location, terrain, and capacity demands; thus the length of handoff link L_i can be several hundred meters or several kilometers. Unfortunately, handoffs do not always occur at the same location matched to the roadway section. Rather, there is a range of freeway section where the handoff might take place. The consistency of handoff location is one of the factors that might affect the estimation accuracy [10]. So this study takes a further look at the impact of the space-varying handoff link length and the time-varying sample size on the accuracy of fusion.

3.1.2. SMS from Microwave Sensors. Generally, the distance between two successive microwave sensor stations is over 1 km on Jiangsu Freeway. As a result, the simple average of spot speed measurements from two adjacent microwave sensor stations to calculate space-mean speed is not suitable for this situation. Another transforming method based on traffic volume and occupancy is also unsuitable, since the measurements from microwave sensor are traffic volume and spot speed only. To transform spot speed to space-mean speed directly, we apply the method proposed by Rakha and Zhang [26]. They derived a modified relationship formulation as follows:

$$\bar{v}_{ms}^j \approx \bar{v}_{mt}^j - \frac{\sigma_t^{2(i)}}{\bar{v}_{mt}^j}, \quad (2)$$

where \bar{v}_{mt}^j is the time-mean speed in the time interval i ; \bar{v}_{ms}^j is the space-mean speed in the time interval i . $\sigma_t^{2(i)}$ is the variance of the time-mean speeds in the time interval i .

Based on the transformation model, the distance between the adjacent microwave sensors will not affect the space-mean speed. On the other hand, the errors caused by the transformation or sensor failure are difficult to measure, although these errors indicate the precision of sensors. Therefore, this study does not take the distance between sensors, sensor errors, and transformation error into consideration. Instead of these influential factors, another traffic measurement, that is, traffic volume, is regarded as an influential factor. And the sensitivity of these factors is investigated in the following section.

3.2. Neural-Network-Based Fusion Module. The neural network technique can learn how to classify and associate input/output patterns, which makes them suitable for solving problems estimating current link traffic speed from several different sources including measurements and other influential factors. This research applied the multilayer feedforward

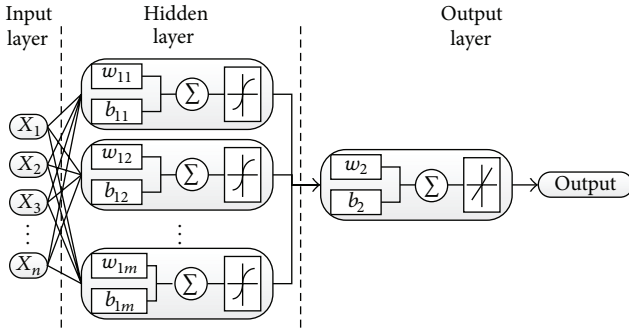


FIGURE 3: The architecture of neural network.

and backpropagation neural network. As shown in Figure 3, the applied neural network includes an input layer, a hidden layer, and an output layer. Different combinations of inputs illustrated in Table 1 will be tested to identify which is the really influential factor that contributes to the precision of fusion. The output layer including one neuron is for the fused result. There is a general conclusion after the several test: more hidden neurons generate less estimation error, while it takes more iterations to train the network. To keep a balance of fusion accuracy and processing time cost, the number of hidden neurons was set to 10 at last.

The neural network in this research is implemented by the function “feedforwardnet” in MATLAB 2014a. And the network training function is “trainbr” which updates weight and bias values according to Levenberg-Marquardt optimization and the processing is called Bayesian regularization. Each neuron in the hidden layer carries a sigmoid transfer function, and the linear transfer function is used for the output layer.

3.3. Neural-Network-Based Estimation Module. Due to the limited coverage of microwave sensor station in practice, there are few links covered by two sensors. The neural-network-based fusion module is designed for estimating the traffic state of these links with two data sources. When only one data source (CHPS) is available, this paper proposed an estimation module to reconstruct the dynamic state. This module shares the same neural network structure as the fusion module. The main difference between estimation module and fusion module is the input data source. The input neurons in the input layer of estimation module are only collected from CHPS. The detailed neural network structure is decided by the sensitivity test of the influential factors.

4. Simulation Settings

Both the training and validation of neural network models require the ground truth speed. The training process makes the neural network perform a particular function by adjusting the values of weights between inputs, so that a particular input leads to a specific output [16]. In the training process, the ground truth is the target output, while, in the validation stage, the ground truth is implemented to evaluate the output

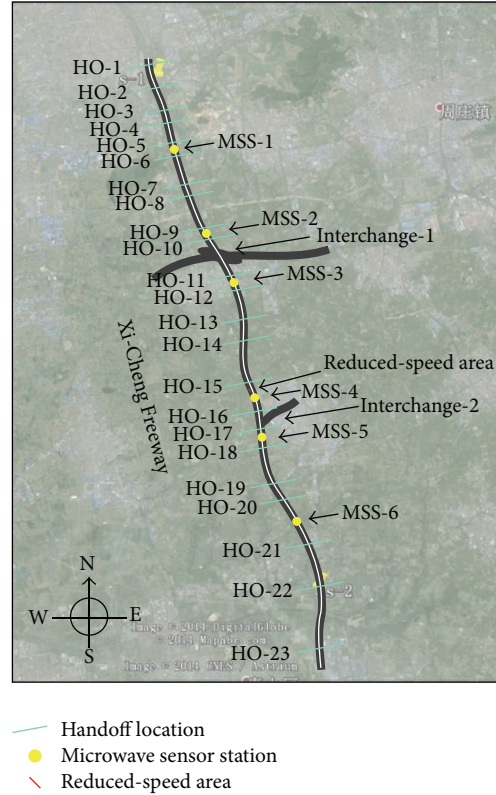


FIGURE 4: The test stretch of Xi-Cheng Freeway.

of fusion model by comparison. However, it is still difficult to obtain the ground truth data in practice.

Microsimulation allows for robust experimentation for traffic studies in general and for data fusion research in particular [6]. In the simulation, the “ground truth” speed could be achieved by counting the average travel time of all vehicles on a link for a given time interval. Meanwhile, some specific type of vehicles is set as the probe vehicles. And the travel time records of those vehicles are similar to the data from cellular handoff probe system. The measurements of microwave sensors could be collected by the function of “Data Collection” in the simulation. The simulation software, VISSIM, is used to build the microscopic traffic model.

4.1. Freeway Geometry and Sensor Layout. A 22 km stretch with three lanes in each direction on Xi-Cheng Freeway, Jiangsu, China, is the simulation test bed. In practice, cellular handoff probe system and microwave sensors have been implemented to collect traffic information on Xi-Cheng Freeway. In the simulation, the handoff locations are decided by the most recent signal test. Six microwave sensor stations are installed on the freeway. The detailed locations of handoffs and microwave sensor stations are shown in Table 2. Figure 4 provides a drawing of the test freeway stretch. Obviously, the length of handoff link that is the distance between two adjacent handoffs is variable (varying from 300 m to 1540 m).

4.2. Traffic Condition Settings. The construction and calibration of the simulation model are based on the field

TABLE 1: Different test scenarios of input neurons in data fusion module.

Scenarios	Number of input neurons	SMS from CHPS	Handoff sample size	Handoff link length	SMS from MS	Volume from MS
1	5	*	*	*	*	*
2	3	*	*		*	
3	2	*			*	
4	4	*	*		*	*
5	4	*	*	*	*	

* indicates the element being chosen as one of the input neurons.

TABLE 2: Xi-Cheng Freeway handoff link and microwave sensor details.

Link index (southbound)	Link index (northbound)	Link length (m)	MS station	From handoff point	To handoff point
1	44	1195		HO-1	HO-2
2	43	1100		HO-2	HO-3
3	42	600		HO-3	HO-4
4	41	950		HO-4	HO-5
5	40	600	MSS-1	HO-5	HO-6
6	39	1250		HO-6	HO-7
7	38	450		HO-7	HO-8
8	37	1100		HO-8	HO-9
9	36	300	MSS-2	HO-9	HO-10
10	35	1210	Interchange-1	HO-10	HO-11
11	34	790	MSS-3	HO-11	HO-12
12	33	1040		HO-12	HO-13
13	32	880		HO-13	HO-14
14	31	1380		HO-14	HO-15
15	30	1140	MSS-4	HO-15	HO-16
16	29	760	Interchange-2	HO-16	HO-17
17	28	770	MSS-5	HO-17	HO-18
18	27	1110		HO-18	HO-19
19	26	690		HO-19	HO-20
20	25	1520	MSS-6	HO-20	HO-21
21	24	1390		HO-21	HO-22
22	23	1540		HO-22	HO-23

data collected from 07:00 to 19:00 including two peaks (a morning peak from 08:00 to 09:00 and an evening peak from 15:00 to 17:00). To create a dynamic traffic condition, a 300 m reduced-speed area is set on the two inside lanes within handoff link 15 of the southbound Xi-Cheng Freeway, specifically between handoff point HO-15 and microwave sensor station MSS-4, as shown in Figure 4. And the speed reduction lasts 2 hours from 12:00 to 14:00.

4.3. Measurements. Microwave sensor stations upload traffic measurements every 5 minutes to the Jiangsu Freeway Operation Center, while cellular handoff probe system collects cellphone communication records every 1 minute. In this study, measurements from both sources are aggregated every 15 minutes from the same start time. The statistical result of CHPS from Jiangsu Freeway Operation Company shows

that the ratio between handoff probe samples and traffic flow volume is variable. In general, the sample proportion varies from 1% to 15%. To simulate the randomness of sample size, the percentage of probe vehicles in every interval uses a random number between 1% and 15%.

The function of “Data Collection” in VISSIM is capable of recording the spot speed and counting the number of each pass-by vehicle on different lanes. These records are used as the measurements of microwave sensors and are integrated every 15 minutes.

The ground truth space-mean speed of each link is transformed from the average travel time of each link. The average travel time of all vehicles is obtained by the function of “travel time” in VISSIM. All this information will be processed to test, validate, and evaluate the neural-network-based data fusion method.

5. Investigation and Evaluation

This section includes (1) an investigation of factors that will influence the fusion accuracy by sensitivity analysis; (2) determining the final fusion model based on the previous investigation; (3) using the simulation data and fusion model to generate the fusion results; (4) a comparison between the proposed fusion method and other approaches. Both investigation and evaluation of the proposed method are based on the simulation data. More specifically, the simulation results of the whole 12 hours are divided into two data subsets, that is, one subset for model training and investigation from 07:00 to 13:00 and another subset for evaluation and comparison from 13:00 to 19:00. Each subset covers a peak hour and an hour affected by the incident (a speed deceleration). Thus, the first subset could train the fusion method adapting to dynamic traffic conditions, while the second subset would test the adaptability of the proposed method to dynamic traffic conditions. As a measure of accuracy, this paper utilizes the Root Mean Square Error (RMSE) which can be calculated by the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N (\hat{v}_k - v_k)^2}, \quad (3)$$

where \hat{v}_k is the fused speed in time interval k ; v_k is the surrogate ground truth speed in time interval k ; N is the number of time intervals.

5.1. Sensitivity Analysis of Influential Factors. The majority of existing works on traffic data fusion focus on the direct integration of traffic variables (e.g., traffic speeds from two sources). The objective of this subsection is to investigate other factors besides traffic variables that might affect the accuracy of fusion results. The following reasons show why some factors deserve a deeper investigation.

Microwave sensor is a kind of nonintrusive technology for traffic detection. The traffic agencies and researchers [27, 28] evaluated this collection technique and indicated that factors influencing its precision included (1) sensor failures; (2) the installation positions and density; (3) traffic condition, such as traffic volume level; (4) environmental conditions, such as weather and lighting. For the first factor, in this study, the training of neural-network-based fusion model could find out the relation between inputted speed from microwave sensors and the ground truth speed. To some extent, the training process considers the situation when the sensor fails. Since the location of each microwave sensor is fixed, the second factor about the sensor installation or the location optimization problem is not in the scope of this study. For the third factor (traffic condition), considering that it affects the accuracy of microwave sensor, this study aims to investigate whether this factor will have an effect on the fusion results. The fourth factor (e.g., weather) does have an impact on the freeway traffic [29, 30] as well as the performance of traffic sensors. It will be more meaningful to evaluate the influence of environmental conditions on the fusion. Although this study has not taken the environmental conditions into consideration due to the limited data source,

the proposed neural-network-based model has an open structure. It provides an opportunity for us to study the impact of environmental factors by adding weather factors (such as the degree of rain, snow, temperature, and visibility) into the input layer in the future. In summary, the traffic counts besides the traffic speed from microwave sensors are taken into the investigation.

Cellular handoff probe system is a kind of probe-based technique for traffic detection. The factors proven to have an influence on the accuracy of CHPS include (1) call duration; (2) handoff link length; (3) the accuracy of handoff location; (4) handoff consistency; (5) the valid number of cellphone probes [31, 32]. It is hard to study the first factor because the existing CHPS does not provide the information about the call duration. Generally, longer call duration could guarantee more valid probes, since a valid probe requires that the on-call cellphone passes at least two handoff locations and long call duration can ensure the probe phone generates at least two handoff records. It can be inferred that the call duration has an impact on the fifth factor (sample size). For the second factor (the length of handoff link), the existing work [31] shows that it has a negative impact on the handoff sample size, which means that handoff links with shorter length produce more sample size. The third and fourth factors depend on the wireless operation company, which means that these two factors are determined by the location of cell towers and the signal strength of antenna. CHPS does not provide such information relevant to these two factors, and thus it is difficult to study the impact of these factors. However, some studies [32] validated that handoff location accuracy and consistency are sufficient to estimate useful travel time. As for the fifth factor (sample size), obviously, it is a critical factor to impact the CHPS accuracy as a kind of probe-based collection technique. In summary, handoff link length and sample size besides traffic speed from CHPS are selected to study their impact on the precision of fusion.

The sensitivity of different factors on the accuracy of fusion model was tested by the various combinations of inputs in neural network models. Figure 5 illustrates RMSE of different neural networks on southbound links and northbound links, respectively. Figure 5(c) is just an enlarged version of the bottom lines in Figure 5(b) for distinction. Based on Figure 5, the average RMSE of all links during the test time are 0.456, 0.404, 1.193, 0.673, and 0.399 for neural network with five inputs (two speed values from multisource, handoff link length, sample size of CHPS, and traffic volume from microwave sensor), three inputs (two speed values from multisource and sample size of CHPS), two inputs (two speed values from multisource), four inputs (two speed values from multisource, sample size of CHPS, and traffic volume), and four inputs (two speed values from multisource, handoff link length, and sample size of CHPS), respectively. The following are some findings about the sensitivity of the investigated factors.

Firstly, the factors of sample size of CHPS, handoff link length, or traffic volume are all influential factors that increase the fusion accuracy. Although speeds from CHPS and microwave sensors are essential inputs, the fusion of only speeds from two sources has the worst performance.

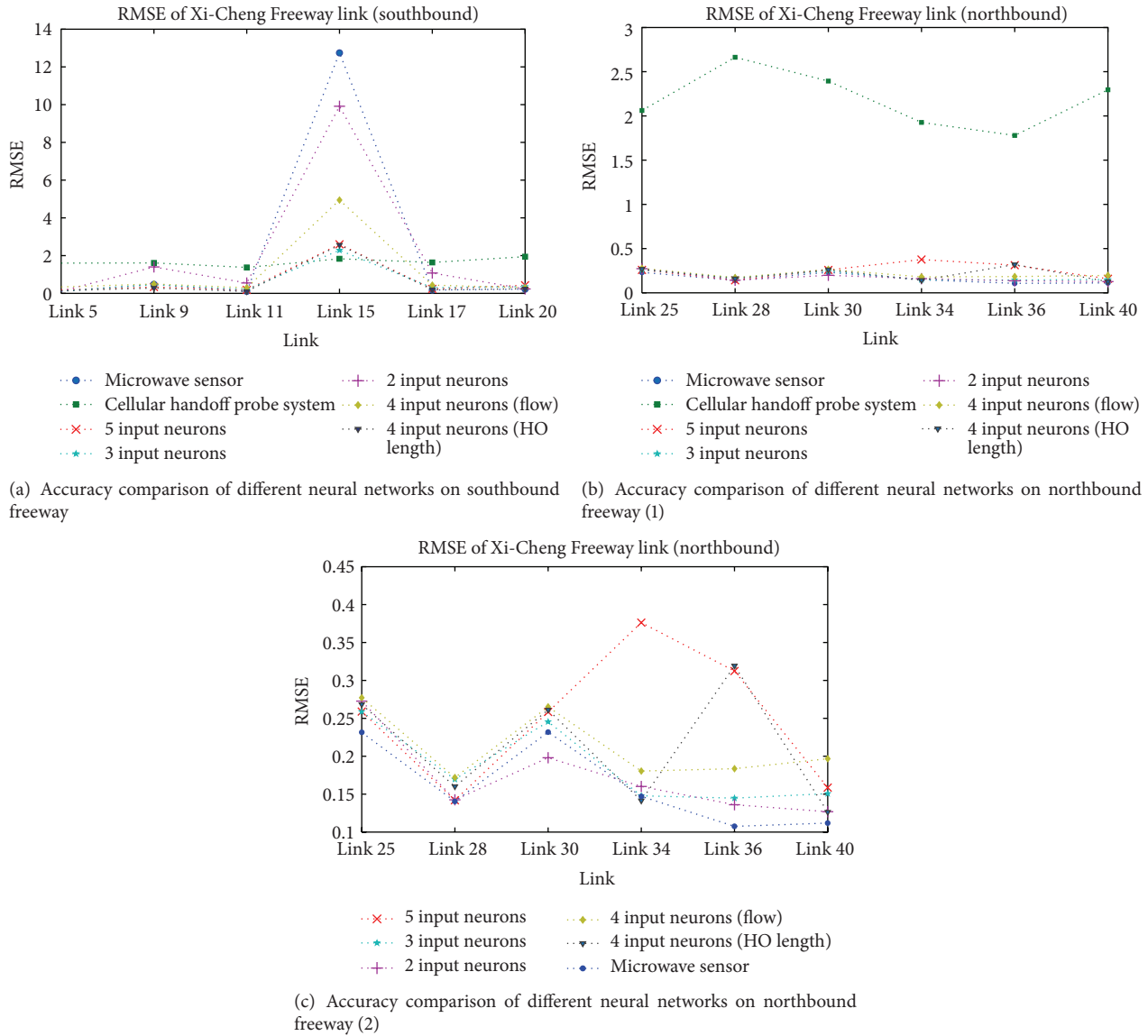


FIGURE 5: Accuracy comparison of different neural networks in fusion module.

The second finding is that the importance of every influential factor on the fusion accuracy can be inferred. With the smallest average RMSE (0.399), the combination of sample size and handoff link length has the strongest influence on the fusion accuracy. When only sample size is taken into consideration, the average RMSE (0.404) is still low comparatively. The average RMSE of the combination of sample size and traffic volume increases. The combination of sample size, handoff link length, and traffic volume can reduce the error of the former combination error. It can be inferred that the combination of sample size and handoff link length makes the most influential factor and has more positive impact on the fusion accuracy.

The third finding is related to the traffic condition. It can be observed that speeds from microwave sensors can achieve high accuracy under free flow condition. However, the error of microwave sensor is high under the incident condition.

The reason for the inaccuracy under incident condition is the speed deceleration happening on the upstream of microwave sensor station. The traffic at the sensor point is recovered according to the propagation of shockwave. On the other hand, the accuracy of CHPS is relatively stable under different traffic condition. As a result, under free flow condition, the fused results reduced the error compared with measurements only available from cellular handoff probe system. And, under the incident condition, the fused results reduced the error caused by the microwave sensors.

5.2. Determining the Neural Network Fusion Model Structure.

Based on the previous investigation, the influential factors sample size and the length of handoff link have the manifest effect on the fusion accuracy, and thus the neural network model with four input neurons (speed, sample size and handoff link length from cellular handoff probe system, and speed

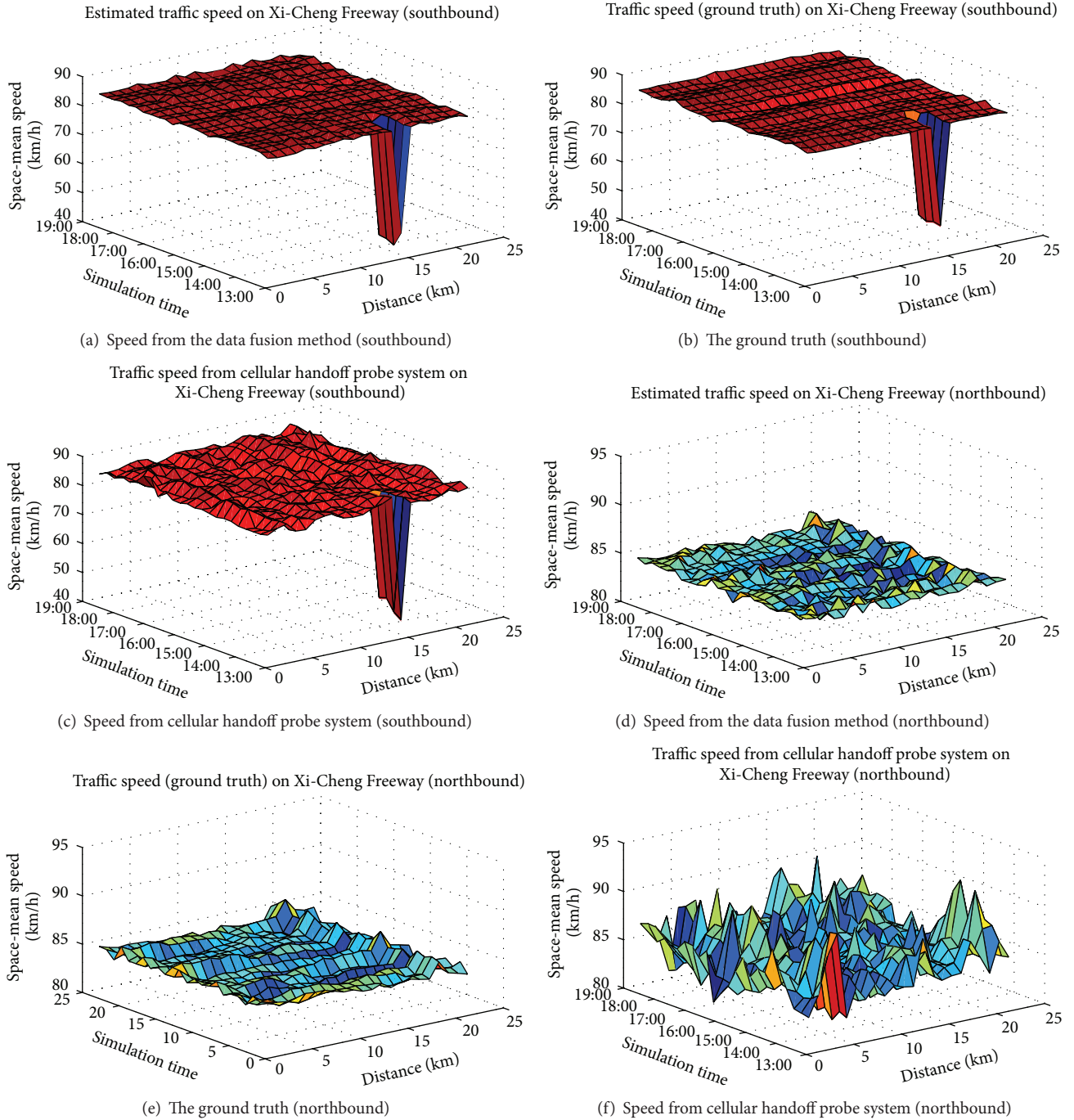


FIGURE 6: Results of data fusion method.

from microwave sensors) is trained as the optimal fusion model. Neural network model with three input neurons (two-source speeds and sample size) is suboptimal. Comparatively, the sample size is an important factor for the improvement of fusion accuracy. Handoff link length also plays a decisive role in improving the fusion performance. It suggests that speed, sample size, and handoff link length should also be the inputs for the neural-network-based estimation module. Except for the sample size and handoff link length, the effect of another influential factor (traffic volume) is not as critical

as the other two factors. Therefore, the neural network model with four input neurons (speed, sample size and handoff link length from cellular handoff probe system, and speed from microwave sensors) in the fusion module is applied, while three input neurons (speed, sample size, and handoff link length) are used in the estimation module.

5.3. Evaluations of the Proposed Data Fusion Approach. Simulation data from 13:00 to 19:00 is applied to evaluate the proposed data fusion approach in this section. Figure 6 shows

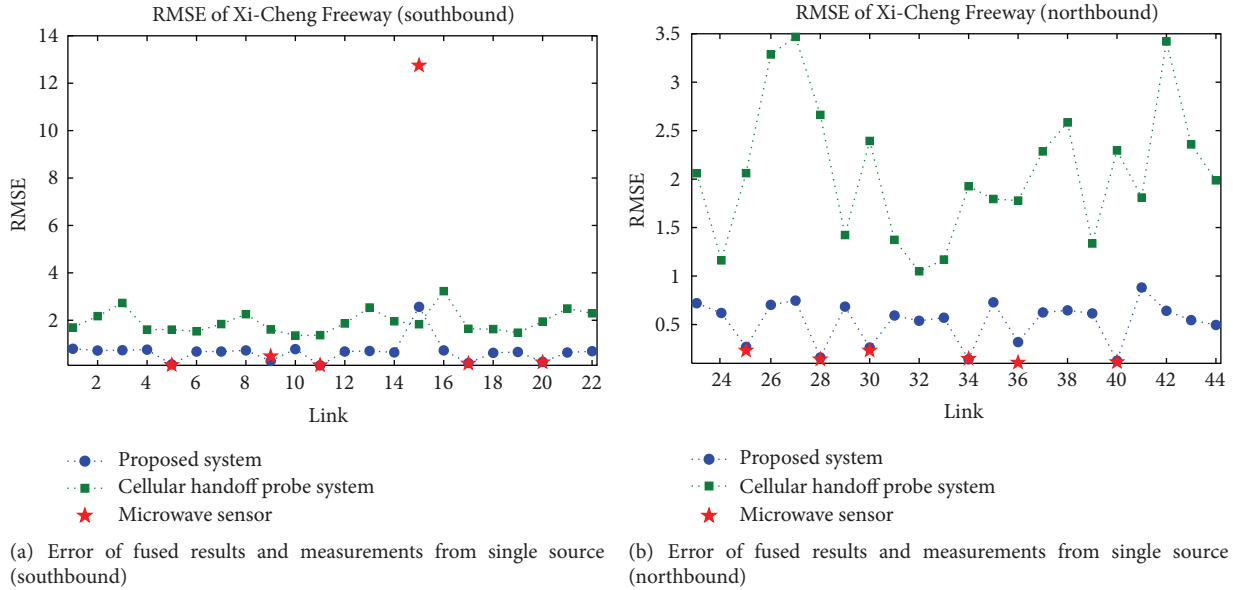


FIGURE 7: Accuracy comparison between fused results and measurements from single source.

the outputs of the data fusion method, the ground truth speed, and the speed measurements from cellular handoff probe system. Figures 6(a), 6(b), and 6(c) are speeds on southbound Xi-Cheng Freeway, whereas Figures 6(d), 6(e), and 6(f) are speeds on the northbound freeway. It can be inferred that the data fusion approach produced the traffic state variables which were much closer to the ground truth on both southbound and northbound freeways. This approach is capable of tracking dynamic characteristics of traffic speed as shown in Figure 6.

For a detailed evaluation of the proposed data fusion approach, RMSE of different links during test period is calculated, as illustrated in Figure 7. Although microwave sensors can provide more accurate estimation under free flow condition, their performance under the congested condition is unstable. In this study, the microwave sensor stations are located downstream of the traffic incidents. And the microwave sensor did not monitor the speed deceleration on the upstream location, which could be explained by the propagation of the shockwave. Thus, it leads to high estimation error under this congested condition. Besides, their spatial resolution is limited due to the high cost of installation and maintenance. Comparatively, the error of CHPS is stable under various traffic conditions which is a little higher than the error of microwave sensors. Integrating the merit of both detectors (i.e., the high accuracy of microwave sensors under free flow condition, the stability precision, and large spatial coverage of CHPS), the proposed data fusion method could produce a better overall traffic state estimation.

5.4. Comparison between the Proposed Method with Other Fusion Methods. The existing research has made a comparison between various fusion methods for traffic state estimation [6]. The authors have made several conclusions about six fusion methods. (1) The authors do not recommend

the ordered weight average (OWA) method and the Choquet fuzzy integral algorithm for data fusion. Both of these two methods require the inherent relationship between sensors. For instance, OWA method needs to know the significance order of measurements from different sensors, which is hard to achieve if the precision of sensors is rearranged frequently and their inherent properties are different. The fuzzy integral suffers from the similar problem for not fully finding out the relationship between sensors. (2) That research found that the simple convex combination, Bar-Shalom/Campo combination, and single-constraint-at-a-time Kalman filter had a similar performance which significantly improved the fusion accuracy. If Kalman filter does not use other traffic variables (flow and density) to build up more advanced state-space model, the two combination methods take more advantages for much simpler computations. (3) Neural networks performed as well as other fusion methods and were even better sometimes. Our previous investigation and evaluation present a proof that the neural networks can perform better with more inputs, such as the handoff probe sample size, handoff link length, and traffic volume. Although Kalman filter can also add more traffic variables by applying kinematic traffic dynamic model, it is difficult to add nontraffic variables (e.g., handoff link length) into the Kalman filter. Following the existing research findings, the neural-network-based method is an optimal choice for the fusion of data from CHPS and microwave sensors.

However, the existing work did not make a comparison of Dempster-Shafer (DS) evidence theory based fusion approach. The reason might be the different output form of DS fusion process [18] compared with other fusion methods. For instance, when the speed measurements from CHPS and microwave sensors are going to be fused, the first step of a DS fusion process is to divide measurements into predefined classes. Each class reflects a kind of traffic conditions, such

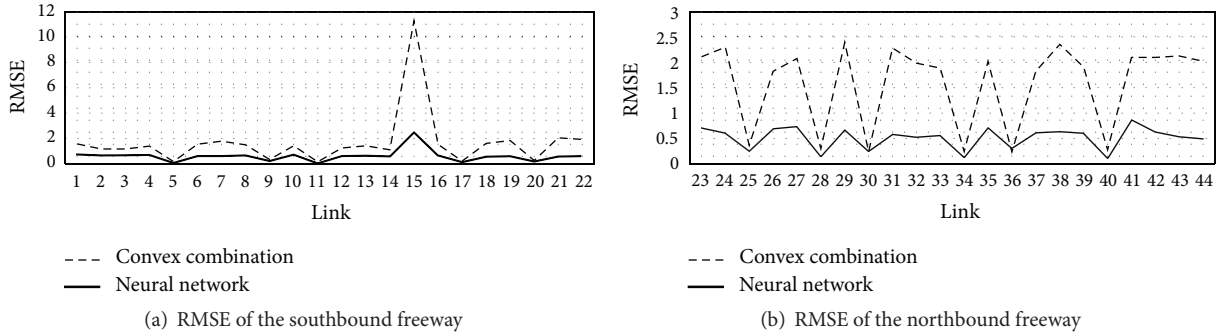


FIGURE 8: Accuracy comparison of different fusion method.

as free flow, crowded, and congested. Then, the following processes include generating probability mass functions, using Dempster's rule to fuse mass value matrix, and choosing the class with the largest probability as the output of the DS fusion process. Thus, the direct output is not a fused speed value. In some senses, DS fusion method is more suitable for the state fusion, while this study aims to solve the data fusion problem. Besides, the challenges of determining the degree of belief and dealing with the conflict evidence are still issues of DS fusion method.

For the above reasons, this study chooses the convex combination as a representation to make a numerical comparison with the proposed method. A simple convex combination is a linear combination of measurements from different sensors. The linear coefficients are obtained from the covariance between sensors and the ground truth. The same training set of data as neural network methods is applied to achieve the coefficients. And the same testing set of data is used to fuse and calculate the RMSE. In the proposed method, the neural networks are applied to fuse both multisource data and single source data, while the convex combination is only applicable when two sources are available. The fusion function and covariance are as follows:

$$\begin{aligned}
 v_f &= [(p_c)^{-1} + (p_m)^{-1}]^{-1} \cdot (p_c)^{-1} \cdot v_c \\
 &\quad + [(p_c)^{-1} + (p_m)^{-1}]^{-1} \cdot (p_m)^{-1} \cdot v_m, \quad (4) \\
 p^{-1} &= (p_c)^{-1} + (p_m)^{-1},
 \end{aligned}$$

where v_f is the fused speed; v_c is the space-mean speed from CHPS; v_m is the space-mean speed from microwave sensor; p_c is the covariance between speed measurements from CHPS and the ground truth speed; p_m is the covariance between speed measurements from microwave sensor and the ground truth speed; p is the covariance of fused speed.

Figure 8 shows the RMSE of the convex combination and the proposed method. Obviously, the proposed neural-network-based fusion method outperforms the convex combination. On the other hand, the training dataset is not very large, so there is no significant difference between the time cost to train the neural network model and that to calibrate the coefficients of the convex combination function. With the increasing volume of training data, the neural network will be

more robust but, at the same time, the training will be more time-consuming. Therefore, it is needed to seek a balance of accuracy and time-consuming problem.

6. Conclusions

This paper has described a data fusion approach designed to provide freeway speed estimation through fusing data from cellular handoff probe system and microwave sensors. It includes three main modules, that is, data transformation module, neural-network-based estimation module, and neural-network-based fusion module. The neural-network-based fusion module is designed to fuse traffic information collected from two sources. Another neural-network-based estimation module is developed to estimate the link speed when only cellular handoff probe data is available. The proposed method has been developed, trained, tested, and validated with the data from the simulation.

The main contributions of this study include finding out the sensitive factors that influence the fusion accuracy, determining the optimal architecture of neural network model, validating and evaluating the capability of fusion approach, and comparing the proposed method with other fusion methods. Specifically, the sensitivity investigation was conducted using several factors including speeds from two sources, sample size and the length of handoff link from CHPS, and traffic volume from microwave sensors. These factors have an impact on the single collection technique, and then the investigation proves that they also influence the fusion accuracy. The results indicate that the combination of handoff sample size and handoff link length is most effective and influential to the fusion precision. And using this neural network architecture (speeds from two data sources, probe sample size, and handoff link length as inputs), the proposed method is capable of improving the estimation accuracy under different traffic conditions (free flow or congested condition). The performance of the proposed fusion method is evaluated by more detailed analysis and comparison. First, the fusion results indicate that the combining two data sources can complement the drawback of an individual sensor and thus improve the overall estimate accuracy. Moreover, the proposed method is effective under different traffic conditions. Finally, an analysis of the proposed method and other data fusion approaches

shows that the neural-network-based method is optimal. A valid comparison furtherly proved that the proposed method is superior to the convex combination. It is constructive and significant for the development of traffic data fusion theory and field engineering application.

In this study, the validation and evaluation of proposed fusion approach were based on the traffic simulation model. Although simulation could obtain a high degree of similarities to the reality, the complexities of probe vehicle behavior, weather, incident, and so forth are extremely difficult to simulate. So, the further evaluation of the data fusion method is critical for its validation before it can be applied in practice. For instance, the data fusion method should be tested using field data and also in other road networks. More types of traffic conditions also need to be explored. Besides, with the increase of training dataset, it is necessary to find a balance among the size of data to be trained, the computation time, and the fusion accuracy.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this article.

Acknowledgments

This work was supported by the National Key Basic Research Development Program of China under Grant no. 2012CB725405 and the National Natural Science Foundation of China under Grant no. 51308115.

References

- [1] D. Ngoduy, S. P. Hoogendoorn, and R. Liu, "Continuum modeling of cooperative traffic flow dynamics," *Physica A: Statistical Mechanics and Its Applications*, vol. 388, no. 13, pp. 2705–2716, 2009.
- [2] T.-Q. Tang, L. Chen, Y.-H. Wu, and L. Caccetta, "A macro traffic flow model accounting for real-time traffic state," *Physica A*, vol. 437, Article ID 16169, pp. 55–67, 2015.
- [3] D. Ngoduy, "Applicable filtering framework for online multiclass freeway network estimation," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 2-3, pp. 599–616, 2008.
- [4] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, 2005.
- [5] Q. Ou, *Fusing heterogeneous traffic data parsimonious approaches using data-data consistency [Ph.D. thesis]*, Delft University of Technology, 2010.
- [6] C. Bachmann, B. Abdulhai, M. J. Roorda, and B. Moshiri, "A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling," *Transportation Research—Part C: Emerging Technologies*, vol. 26, pp. 33–48, 2013.
- [7] S. He, W. Wang, J. Zhang, F. Zhan, and B. Ran, "Evaluation of freeway sensor placement based on the aggregation of cellular probe system and loop detectors," in *Proceedings of the Transportation Research Board 93rd Annual Meeting*, Washington, DC, USA, 2014.
- [8] B. Hellenga and R. Gudapati, "Estimation link travel times for advanced traveller information systems," in *Proceedings of the 3rd Transportation Specialty Conference*, Canadian Society for Civil Engineering, Ontario, Canada, 2000.
- [9] Z. T. Qiu, *Macroscopic traffic state estimation for large scale freeway network [Doctor of Philosophy]*, University of Wisconsin-Madison, Madison, Wis, USA, 2007.
- [10] N. Caceres, L. M. Romero, F. G. Benitez, and J. M. Del Castillo, "Traffic flow estimation models using cellular phone data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1430–1441, 2012.
- [11] Q.-J. Kong, Y. Chen, and Y. Liu, "An improved evidential fusion approach for real-time urban link speed estimation," in *Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC '07)*, pp. 562–567, IEEE, Seattle, Wash, USA, October 2007.
- [12] E. Faouzi, "Data fusion in road traffic engineering: an overview," in *Proceedings of the SPIE: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications*, vol. 5434, pp. 360–371, Orlando, Fla, USA, 2004.
- [13] J. W. C. V. Lint and S. P. Hoogendoorn, "A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 8, pp. 596–612, 2010.
- [14] A. Tarko and N. Roupail, "Travel time fusion in advance," in *Proceedings of the Pacific Rim TransTech Conference: A Ride Into the Future*, pp. 36–42, Settle, Wash, USA, 1993.
- [15] J.-Q. Li, K. Zhou, S. E. Shladover, and A. Skabardonis, "Estimating queue length under the connected vehicle technology: using probe vehicle, loop detector, and fused data," in *Proceedings of the Transportation Research Board 92nd Annual Meeting*, Washington, DC, USA, 2013.
- [16] C. Bachmann, *Multi-sensor data fusion for traffic speed and travel time estimation [M.S. thesis]*, Department of Civil Engineering, University of Toronto, Toronto, Canada, 2011.
- [17] 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," *Journal of Intelligent Transportation Systems*, vol. 17, no. 2, pp. 152–164, 2013.
- [18] N.-E. El Faouzi, L. A. Klein, and O. De Mouzon, "Improving travel time estimates from inductive loop and toll collection data with dempster-shafer data fusion," *Transportation Research Record*, no. 2129, pp. 73–80, 2009.
- [19] C. Mengying, K. Liu, and J. Chen, "A recursive-Bayesian inference model for dynamic travel time estimation using fusion of simulated loop detector and probe data," in *Proceedings of the Transportation Research Board 93rd Annual Meeting*, Transportation Research Board, Washington, DC, USA, 2014.
- [20] K. Choi, "Data fusion methodology for link travel time estimation for advanced traveler information system," *KSCE Journal of Civil Engineering*, vol. 3, no. 1, pp. 1–14, 1999.
- [21] R. L. Cheu, D.-H. Lee, and C. Xie, "An arterial speed estimation model fusing data from stationary and mobile sensors," in *Proceedings of the IEEE Intelligent Transportation Systems Proceedings*, pp. 573–578, Oakland, Calif, USA, August 2001.
- [22] J. W. C. van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 5-6, pp. 347–369, 2005.
- [23] 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,” *Transportation Research Record*, vol. 1855, pp. 49–59, 2003.
- [24] Y.-J. Byon, A. Shalaby, B. Abdulhai, and S. Elshafiey, “Traffic data fusion using SCAAT Kalman filters,” in *Proceedings of the Transportation Research Board 89th Annual Meeting*, Washington, DC, USA, 2010.
- [25] K. Choi and Y. Chung, “A data fusion algorithm for estimating link travel time,” *Journal of Intelligent Transportation Systems*, vol. 7, no. 3-4, pp. 235–260, 2002.
- [26] H. Rakha and W. Zhang, “Estimating traffic stream space-mean speed and reliability from dual and single loop detectors,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1925, pp. 38–47, 2005.
- [27] D. K. Chang, *Evaluation of the accuracy of traffic volume counts collected by microwave sensors [Master of Science]*, Brigham Young University, Provo, Utah, USA, 2015.
- [28] E. Minge, J. Kotzenmacher, and S. Peterson, *Evaluation of Non-Intrusive Technologies for Traffic Detection*, SRF Consulting Group, Minneapolis, Minn, USA, 2010.
- [29] M. Agarwal, T. H. Maze, and R. Souleyrette, “Impact of weather on urban freeway traffic flow characteristics and facility capacity,” in *Integration of Road Weather Information with Traffic Data*, Aurora project, Center for Transportation Research and Education, Iowa State University, Ames, Iowa, USA, 2005.
- [30] U.S. Department of Transportation, “F. H. A. How do weather events impact roads?” http://www.ops.fhwa.dot.gov/weather/ql_roadimpact.htm.
- [31] F. Yang, Y. Jiang, and L. Wang, “Traffic probe sample size experiment based on mobile phone handover information,” in *Proceedings of the 2nd International Conference on Transportation Engineering (ICTE '09)*, pp. 681–686, Chengdu, China, July 2009.
- [32] D. Gundlegård and J. M. Karlsson, “Handover location accuracy for travel time estimation in GSM and UMTS,” *IET Intelligent Transport Systems*, vol. 3, no. 1, pp. 87–94, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

