Hybrid Multi-Sensor Data for Traffic Condition Forecasting

Qinglu Ma
College of computer science, Chongqing University, Chongqing, 400044, China
Email: MQL360@gmail.com

Weining Liu and Dihua Sun
Key Laboratory of Dependable Service Computing in Cyber Physical Society of Ministry of Education, Chongqing University, China
Email: {lwn, d3sun} @ cqu.edu.cn

***Abstract*—Existing time-series models that are used for short-term traffic condition forecasting are mostly monophyletic detector in nature. Generally, information-fusion online from multi-source sensors to more accurately and completely obtains traffic condition than using either of them alone. In this paper, a prediction method for real-time traffic condition prediction is presented, in which the multi-source detection data are collected by different components of a time-series data sets, such as Global Position System (GPS), Radio Frequency Identification (RFID), video camera. Finally, a case study at the Chongqing, China, city center with serious traffic congestion is performed to illustrate the forecasting strategy. The results indicate that the presented forecasting strategy is one of the effective approaches to predict the real-time traffic condition in a road network, especially at the locations where no continuous data collection takes place.**

***Index Terms*—traffic condition, forecasting, GPS, RFID, video camera**

I. INTRODUCTION

CONGESTION caused by the significantly increased in mobility demand with respect to transportation system capabilities has become a serious problem, not only for the economic impacts of uncertain travel time but also for the ecological aspects involved (e.g., energy consumption, pollution, and many more). An important requirement is the availability of detailed information about present and potential traffic demand. The efficient use of existing infrastructure by dynamic traffic management is one of the strategies to reduce congestion.

In general, the current traffic information can directly be measured with different detectors but the future needs can be estimated by present and historic data. Therefore, Short-term traffic forecasting is an important tool in following the evolution of traffic conditions over time in any transport network. The major aim of anticipation is the provision of real-time and accurate road network traffic state information, i.e., traffic counts, travel speed and so on.

Practically, the short-term traffic condition prediction exists a time horizon of 15 min or less, supports short-range operational modifications to improve the efficiency of the network at a finer scale. The ability to provide a dynamic traffic control requires continuous forecasting of traffic conditions in the near future. With the increasing demand for the development of more efficient Intelligent Transportation System (ITS) based traffic management systems, considerable research attention has been focused on short-term traffic information forecasting. An extensive review of this subject has been given by Bidisha Ghosh et al. (2009)[1] and Vlahogianni et al. (2005) [2].

In addition, traffic data collection under complex traffic conditions is one of the major problems faced by researchers as well as traffic regulatory authorities. Study and analysis of traffic information forecasting is critically dependent on the availability of observed traffic data. In modern transportation systems, information about the position and the orientation of the vehicles should be accurate. More and more different types of sensors (e.g., GPS[3]–[6], RFID[7],[8], video cameras[9],[10], and cell phones [11],[12]) have been employed to detect traffic parameters and forecast traffic condition by virtue of the rapid progress of sensor technology[13]. Multiple sources may provide complementary data, and multi-source data fusion can produce a better understanding of the observed situation by decreasing the uncertainty related to the individual sources. Therefore, the multi-source information fusion technology[14] has been expected to get more accurate and comprehensive forecasting of traffic information by combining data from different detectors.

Several methodologies have been proposed in the literature for the purpose of multi-sensor fusion and aggregation under heterogeneous data configurations[15]-[24]. Although application of data fusion techniques in complex systems modeling is not new, there is a growing interest in their use in transportation systems. Road traffic could be considered as a field where benefits expected from the application of data fusion techniques are fruitful. However, the benefits come with challenges in assessing feasibility, effectiveness and usefulness of such

doi:10.4304/jcp.7.8.1870-1879

approaches [14].

Among these sensors, GPS, RFID and video camera are most widely used at present. However, each of them has some inherent drawbacks, such as the GPS is poor statistical representation and errors in the map-matching process, the RFID is high failure ratio and inaccurate arithmetic conversion of traffic states and the video camera is much expensive and so on. Therefore, In order to guarantee the quality of traffic forecasting information and improve its reliability and availability, we need to use multi-source data for mutual testing, complementing and integrated processing. Multi-source data are the object of manufacture in data fusion, while optimization and integrated processing is the core of data fusion.

In this paper, the multi-source information fusion technology has been introduced to solve the problem of traffic information prediction. The goal is to get more accurate and comprehensive traffic condition by combining data from the speed surveillance of GPS, RFID and video camera.

II. Multi-Sensor Data Collection and Processing

In most cases, it was inadequate that existing methods were used to forecast complex traffic condition, especially, when the traffic data were incomplete. i.e., partially missing or substantially contaminated by noises, while this situation often occurs in practice. The incomplete data can be caused by malfunctions or measurement errors in data collection and recording systems, i.e. faulty GPS locations, failed RFID detectors, or signal communication and processing errors, etc. Although the historical average method (filling up the incomplete data with their historical averages) could be applied to deal with this matter, the forecasting performance was quite limited. For the topic of traffic flow forecasting with incomplete data, the proposed method about data collection and processing as follow.

1) Data Collection

In any transportation network, a huge number of traffic detector data are collected, processed, and transmitted to allow both traffic agencies and travelers to comprehend traffic conditions at any moment. Meanwhile, it also lays a foundation for the automatic control and real-time guidance of road network traffic.

As different types of traffic data collection techniques, such as GPS, RFID and video cameras, have their strengths and weakness respectively. The GPS-equipped vehicles acted as moving sensors traveling in a traffic stream and do not require instrumentation to be set up on the roadway. This technique can easily provide the real-time traffic performance data (e.g., real-time speed, position and direction etc.) on any part of large networks and offer a viable way to complement fixed-point traffic sensors, such as RFID and video cameras, which involving high installation and maintenance cost. The RFID technology allows non-line-of-sight, non-contact, and multiple-tag simultaneous-reading capabilities, which is more efficient than scanning license-plate for vehicle tracking. However, RFID readability can be affected by the relative position and orientation of the tag antenna and the reader antenna, because antenna orientation affects its power pattern[7]. Video camera can through analyze traffic videos to provide macroscopic traffic characteristics such as classified vehicle flows, average vehicle speeds and average occupancies, and microscopic characteristics, i.e. individual vehicle trajectories, lateral, and longitudinal spacing.

In addition to the traffic data, it's also possible in the prediction procedure to use the information of weather and road conditions from intersecting measurement points near the test areas. Both can be changed quickly. Although we try to collect more information as soon as possible, it is not enough to forecast traffic condition if these factors play a role. Therefore, we offer the following rules that the weather and road conditions were classified into three categories:

a. Normal

Both lanes were dry, moist, wet or wet and salted, and there was either no rain or little rain and the warning sensors showed no warnings.

b. Poor

The visibility was 150–299 m or the average wind speed was 12.0–16.9 m/s or none of the conditions above were fulfilled.

c. Hazardous

Both lanes were covered with snow or ice and the road surface temperature was also below +2℃. On one of the lanes, visibility was <150 m, or the average wind speed was over 16.9 m/s.

The weather information was obtained from the China Meteorological Data Sharing Service System (CMDSSs) (http://cdc.cma.gov.cn/index.jsp), which includes hourly temperature (e.g., dry bulb temperature), precipitation (e.g., rainfall and snowfall), and sky conditions (e.g., visibility, wind speed). Meteorological data were selected as a training factor in prediction model development as it was found to have the most significant impact on traffic condition through preliminary analysis. The hybrid data were merged with the weather and the state of road data as consistent data sets for developing prediction models.

The road condition was also an influence factor in improving the accuracy of traffic information forecasting. The qualitative description of evaluation was shown in the table 1.

TABLE 1
Qualitative description of evaluation sets

Factor	Set of Evaluation		
Road Status	Good	Fair	Poor
Road Gradient	Class A ($<10^o$)	Class B ($>10^o$ and $<20^o$)	Class C ($>20^o$)

Intersection Split Ratio	1:1	1:2	1:3 or 1:4

2) Data processing

A. GPS data

The road is segmented into a set of sequential straight lines for ease of calculation. Consider object moving along path P and traveling the distance Δl during the time interval $\Delta t = t_{i+1} - t_i$.

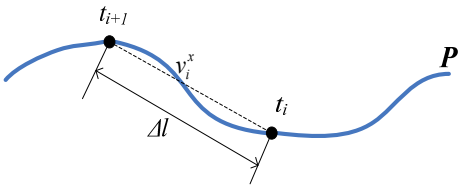


Figure1. Travel detecting through GPS

Then, the average speed $\bar{v}$ over time T is defined as a time average of v as follows:

$$\bar{v} = \frac{\int_{t=0}^{T} v dt}{T}$$

$$= \frac{v(0,t)(\frac{t_1 - t_0}{2}) + \sum_{i=1}^{n-1} v(i,t)(\frac{t_{i+1} - t_{i-1}}{2}) + v(p,t)(\frac{t_n - t_{n-1}}{2})}{\frac{1}{v(x,0)}(\frac{x_i^1 - x_i^0}{2}) + \sum_{j=1}^{n-1} \frac{1}{v(x,j)}(\frac{x_i^{j+1} - x_i^{j-1}}{2}) + \frac{1}{v(x,n)}(\frac{x_i^n - x_i^{n-1}}{2})} \quad (1)$$

Where, x_i^j (j=0,1,2, …, n) and $v(x,t)$ are series in the length and spot speeds, which are collected by running the vehicle on the road segment L_i(measured along the path P).

Because of the uncertain nature of traffic conditions, the travel speed of a bus on route segment normally fluctuates around the average speed of the route segment. To take into account this inherent variation, the average speeds of individual segments are obtained from a dedicated traffic information system which monitors traffic using GPS equipped vehicles as probes. Travel speeds are grouped by time of day and updated continuously to reflect changes in traffic congestion.

In actual measurements, each discrete instantaneous speed sample v_i^x contains measurement error v_i^e. Hence true speed v_i is $v_i = v_i^x - v_i^e$. For a discrete series of speed samples v_i^x acquired at N uniform time intervals over the time T we can approximate the integral in the equation (2) by a sum and the expression for the unknown average speed,

$$\bar{v} = \frac{\sum_{i=1}^{N} v_i^x}{N} - \frac{\sum_{i=1}^{N} v_i^e}{N} = \bar{v}_x - \bar{v}_e \quad (2)$$

Where, $\bar{v}_x$: Exact average of all measured GPS samples. $\bar{v}_e$: Measurement error.

B. RFID data

In the area of transportation, RFID technology with appropriate algorithm and network was applied to a multi vehicle, multi lane and multi road junction area to provide an efficient traffic management scheme, i.e. smart transit card, no parking, parking management, vehicle type and traffic information collection, analysis, calculation of vehicle speed etc.

The raw data from the readers was accumulated in the centralized database, which recorded the average speed of all vehicles traveling between two readers within the area of the controlled zone. Hence accumulated speed average in the area is calculated as:

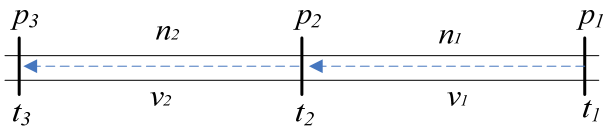


Figure2. Reader reading the passing through tag

$$v_1 = \frac{p_2 - p_1}{t_2 - t_1} \quad (3)$$

$$\bar{v}_1 = \frac{\sum_{i=1}^{i=n_1} v_i}{n_1} = \frac{\sum_{i=1}^{i=n_1} (\frac{p_{i+1} - p_i}{t_{i+1} - t_i})}{n_1} \quad (4)$$

$$\bar{v} = \frac{\sum_{j=1}^{N} \bar{v}_j}{N} = \frac{\sum_{j=1}^{N} \left(\sum_{i=1}^{n_j} v_i \Big/ n_j \right)}{N}$$

$$= \frac{\sum_{j=1}^{N} \left(\sum_{i=1}^{n_j} \left(\frac{p_{i+1} - p_i}{t_{i+1} - t_i} \right) \Big/ n_j \right)}{N} \quad (5)$$

Where, v: Speed of the vehicles between two readers.

p: Location of the RFID reader.

t: Time tagged the RFID tag at the particular reader.

n: Total number of vehicles traveling between two readers

N: Number of reader.

The average speed and the predicted information are important parameter in determining the proper routing of the traffic utilizing the intelligent algorithm for the traffic flow system.

C. Video Camera data

Traffic images can be collected either by remote sensing or by using video camera. In both ways, traffic information can be captured over a certain length of the road, thus useful in obtaining vehicle trajectory data. The geometry of the camera and roadway are shown in Figure3.

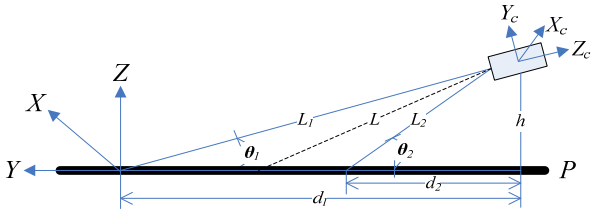


Figure3. Camera and roadway geometry

In the scene, the line P was viewed as one of roads, and assumed that the camera was located at a height h above the ground plane and a horizontal distance d from the foot of the video camera. The camera was oriented at a tilt (down) angle θ such that a point on the ground (X, Y, Z) was transformed into the camera coordinate system (X_c, Y_c, Z_c) by only a translation and a rotation. The camera was oriented along the negative Z_c axis and its line of sight intersected the ground plane a distance $L = h \cdot csc(\theta)$ *away.* Therefore, accumulated average speed in the area is calculated as:

$$v_1 = \frac{d_1 - d_2}{\Delta t} \tag{6}$$

$$\bar{v} = \frac{\sum_{i=1}^{N} v_i}{N} = \frac{\sum_{i=1}^{N} \frac{d_{i+1} - d_i}{t_{i+1} - t_i}}{N} \tag{7}$$

Where, v_1: Speed of a vehicle from time t_1 to t_2.
d_i: Distance from the foot of the video camera.

In addition, the video camera was selected to obtain the real-time traffic images from the traffic road section. Where the aims were judged the current traffic condition and predicted the next traffic information by manual analysis.

D. Data fusion

The purpose of using data fusion was to improve the performances of the traffic information forecasting that combined traffic data from different sources. When the traffic state was obtained according to each data source with Bayesian inference, the mutual information between the judged traffic state and real traffic state can be calculated according to Shannon Entropy[28].

According to the Shannon theorem, the entropy of information source was the limit of lossless coding for the information source.

$$H(X) = -\sum_{i}^{m} P(a_i) \log_2 P(a_i) \quad \text{(bit/ symbol)} \tag{8}$$

Where, X is one of the feature variables set $X=\{X_1, X_2, \ldots, X_n\}$, which has m_i values i.e. $x \in \{a_i\}$, $i=1,2,\ldots,m$. $p(x)$ is probability density function respectively.

If class variable is denoted by Y whose probability density function is $p(y)$ and variable Y has k values i.e. $y \in \{c_i\}$, $i=1, 2, \ldots, k$. that means all features are projected to k different classes. The joint probability distribution of X and Y is denoted by $p(x, y)$, and the joint entropy between the feature variable X and the Y can be described as,

$$H(X,Y) = -\sum_{i=1}^{m}\sum_{j=1}^{k} p(a_i, c_j) \log p(a_i, c_j) \tag{9}$$

Where X_i can be substituted by a subset of X, i.e. the joint entropy can be generalized to the condition of n variables.

The mutual information between feature variable X_i and class variable Y is,

$I(X, Y)=H(X)+H(Y)-H(X, Y)$

$$= \sum_{i=1}^{m}\sum_{j=1}^{k} p(a_i, c_j) \log \frac{p(a_i, c_j)}{p(a_i)p(c_j)} \tag{10}$$

Where, $H(X)$ and $H(Y)$ respectively measure the entropy of variables X and Y, and $H(X,Y)$ corresponds to the cross entropy for these variables. Mutual information values were defined between 0 and 1, and higher the value, the stronger correlation among events. When its value is zero, the events were independent.

According to the result, the weight of each data source based on mutual information is,

$$w_i = \frac{m_i}{m_1 + m_2 + \ldots + m_k} \tag{11}$$

Where, k is the data source number. The mutual information between each data source and real traffic state is denoted by m_i.

The fused posterior probability distribution was denoted,

$$p(Y|X_1, X_1, X_2, \ldots, X_k) = \sum_{i=1}^{k} w_i p(Y \mid X_i) \tag{12}$$

Where, $p(Y|X_i)$, $i=1,2,\ldots,k$ denote the posterior distribution of class variable Y based on data source i.

E. Pre-evaluation of Road Traffic Information

The pre-evaluation of road traffic information aims to estimate traffic condition in real time, both in current and future environments. In the first case the task was particularly challenging because of the low average speeds typical of the city centers, especially at rush hours.

At the end of each time interval, the module produces a score that can be related to the possible presence of traffic slow down. The score can be easily mapped into an indicator of the real traffic situation, for example by associating different traffic information to different score ranges. The traffic level can be proposed for four conditions as follow.

a. no traffic

In general, the average speed was more than 40km/h, few vehicles, typically night traffic. The Green was selected to map in a road section that has a good weather and road condition, relatively. In this case, it's can free travel and without any restrictions either to decrease or increase speed.

b. slow traffic

The average speed was in the range 20 to 40 km/h, tending to increase in traffic. The Yellow was selected to map in a road segment. The traffic is flowing smoothly.

c. heavy traffic

The average speed was in the range 10 to 20 km/h and the Orange was selected to map in a road segment. Traffic moving slowly and congestion will forthcoming.

d. stationary traffic

The average speed was less than 10km/h and the Red was selected to map in a road section. In this case, the traffic is almost status of standstill. Congestion has occurred.

III. FORECASTING METHODOLOGY

In this section, the basic architecture of the traffic information forecasting has been designed and implemented. The core was a data fusion that manages data collection, processing and transmission according with the policy established by the different data sources, respectively, through the activation of the proper modules and the coordination of their actions. For fusing with RFID data, the GPS data collection interval was set 3 minutes. The travel speed of the road section was obtained through calculating the average value of the 3 minutes' GPS data. Accordingly, traffic information is obtained from different assessment methods containing a simple extrapolation model based on single-species traffic detector, the results of forecasting based on multi-detector extrapolation, and a voting model based on simple extrapolation model.

Mutual information[26] between the forecasting result and real traffic state was used to calculate the weight of each data source. Mutual information can measure arbitrary correlation of variables[27]. The weight based on mutual information can display the real correlation of each data source. The traffic information was predicted with simple extrapolation model according to fused probability distribution.

According to the result of extrapolation, the ballot method was adopted to judge the traffic condition among different data source respectively.

In conclusion, the real-time traffic video images, which were backed from the interested area, were selected to predict the traffic condition by manual. After the contrast between the result of ballot and the result of manual, a final traffic condition was decided.

Given the above, the main modules of the hybrid architecture and their interactions were depicted in Figure4.

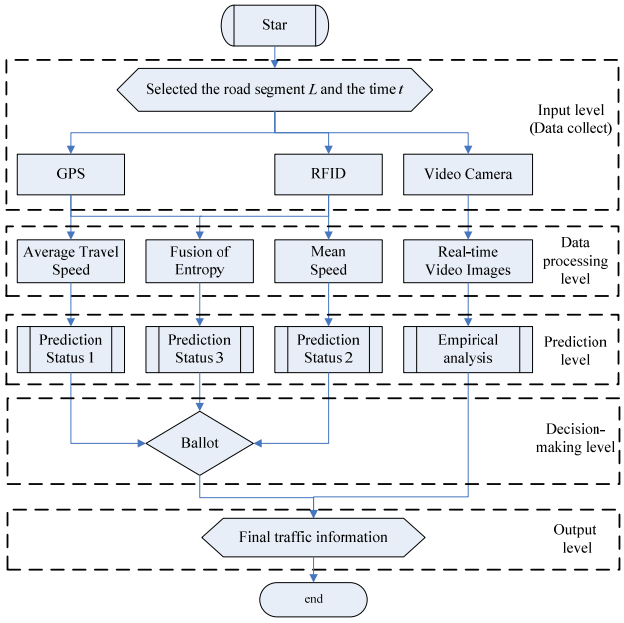


Figure4. Algorithm flowchart for fusion prediction.

In this proposed method, different data-sources have different prediction models to deal. Naturally, different results will be obtained by different forecasting methods in the same road segment at the same time. Then, the ballot theory was selected to deal with these results, and according the real-time video images, we can obtain the better final predictive validity.

There were three steps to show the traffic condition by forecasting information.

Firstly, to calculated the average speed of the road section from the real-time speed of hybrid sensor network.

Secondly, to predicted the travel condition of the road section in next time according to the current information of traffic.

Finally, to show the traffic information by the four traffic level in the E-map.

The average speed is a vital component of traffic information prediction, and the speeds were grouped by time of day and were updated continuously to reflect changes in traffic congestion. To take into account the uncertain nature of traffic conditions (e.g., weather, road condition, etc), the average speed of the route segment presented here was one of the first attempts at real-time short-term prediction of travel time for scheduled public transport.

Li etc. [29] proposed an algorithm that estimates this speed based on historical bus travel speed along the route segment and the current travel speed of the bus derived from GPS data, that is:

$$\hat{v}(x,t) = \frac{\sum_{x=1}^{N-1} v_a(x,t) + v(x,t)}{N} \tag{13}$$

Where, $\hat{v}(x,t)$: predict the speed of route segment x at t time (to the immediate downstream route segments).

$v_a(x,t)$: Historical average speed of route segment x ($x = 1,2,\cdots,N-1$).

$v(x,t)$: Current speed of the bus obtained from GPS data.

N : Number of route segments before reaching the station of interest.

The implication of this algorithm was that when the bus was far away from the station, its predicted speed v would depend primarily on its historical average speed along the route v_{ax} ($x=1, 2, \ldots N-1$) rather than its current speed v_x. In real operating conditions, however, the current speed of the bus was usually a more important factor influencing how fast the bus will travel over the distance to the destination station of interest. Furthermore, this method would predict a non-zero speed even when v_x=0.

This model could therefore perform poorly in applications where the bus speed changes frequently, the GPS speed was not sufficiently accurate or the historical average speed of route does not adequately reflect the current speed of the traffic.

Sun and Luo et al. [3] based on average speed estimate updated dynamically as follow,

$$v(x,t+\tau) = \frac{L_{xe}v(x,t) + L_{xb}v_a(x,t-\tau)}{L_{xe} + L_{xb}} \quad (14)$$

Where, L_{xe}: distance from the current bus location to the end of the route segment x.

L_{xb}: distance from the beginning of the route segment x to the current bus location.

It was more reasonable for proposed algorithm to develop prediction model based on the average speeds of route segments rather than the historical average velocity. And it leads to superior result, because this model only estimates the speed to the end of route segment instead of the speed to downstream bus stations, when the bus was far from the station.

To take into account the uncertain nature of traffic conditions, this paper proposed a new scheme for predicting travel time to a downstream stop. The proposed method first estimated the objective factors (e.g., weather, road condition etc.) to the immediate downstream route segments, denoted by i, based on the weather adjustment factors estimate updated dynamically as follows:

$$v(x,t+\tau) = pv(x,t) + (1-p)v(x,t-\tau) \quad (15)$$

Where, p: Objective adjustment factors (0⩽p⩽1).

Then, based on (13), (14) and (15),

Let $A = pL_{xe}, B = (1-p)L_{xf}$, then average speed estimate updated dynamically as follow,

$$v(x,t+\tau) = \frac{Av(x,t) + Bv_a(x,t-\tau)}{A+B} \quad (16)$$

Obviously, where, if p=0.5 then (16) will back into Sun et al (14). Similarly, when p=1, $v(x,t+\tau)$ is only related with current data, when p=0, $v(x,t+\tau)$ is only related with historical data. In generally, $v(x,t+\tau)$ will depend on the fusion of current and historical data.

IV. Experiment

The study area for this research consists of the urban network of Chongqing, The China, which is a Largest-sized municipality with about eight millions inhabitants and a cross section of about 640 square kilometers. Data were collected at about 20 signalized urban intersections from September 2009 until September 2010. Vehicles were detected by a real-time collection technique. The data were processed into volume measurements per link per time interval. A link represents a unidirectional road segment that could contain more than one lane. We define a volume profile $Q_{dl} = (q_{dl1}, \ldots, q_{dln})$ as a time series of n intervals for day d and link l. In most cases, measurements were provided in 5-min intervals so that n = 288. However, for about 30% of the links, only measurements in 30-min intervals were provided. For these links, n = 48.

The traffic information measurement of traffic network (as shown Figure5.) used hybrid sensor devices was determined by a real-time collection technique. The real-time actual data were used to test the effectiveness of the proposed integration forecasting approach in the busy city centers of Chongqing, China, where is a biggest-sized Municipality with about 8 million inhabitants.

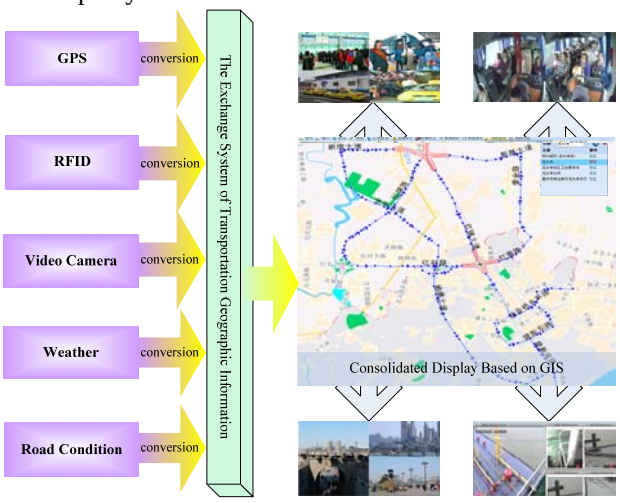
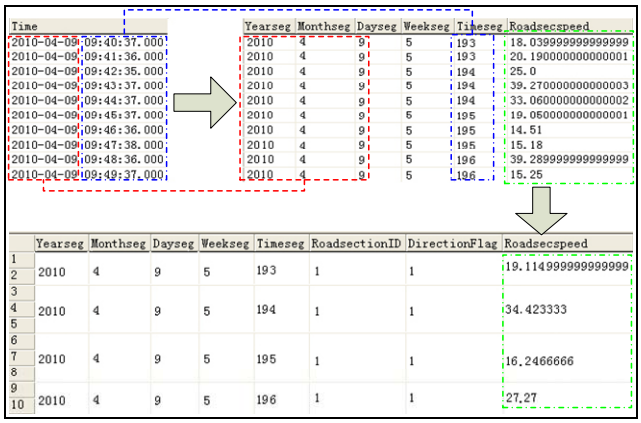


Figure5. Consolidated display of transport network based on GIS

The GPS data collection interval was set 10 seconds. As a key traffic parameter, the average travel speed can analyze the traffic condition of complex urban road. RFID and video data will be provided in the partial crossway. During the collection time, sample time interval is 3 minutes, then the whole day, from 00:00:00 to 23:59:59, can be divided into 480 time range, i.e. each one hour was divided into 20 time range and so forth, time range can be obtained by following formulation (17)

$$t_m = H_t \times 20 + \frac{M_t}{3} \quad (H_t\text{:Hour},\ M_t\text{:Minute}) \qquad (17)$$

Figure 6 shown the differences were significant before and after time format conversion experiment.



	Yearseg	Monthseg	Dayseg	Weekseg	Timeseg	RoadsectionID	DirectionFlag	Roadsecspeed
1								
2	2010	4	9	5	193	1	1	19.1149999999999999
3								
4	2010	4	9	5	194	1	1	34.423333
5								
6								
7	2010	4	9	5	195	1	1	16.2466666
8								
9								
10	2010	4	9	5	196	1	1	27.27

Figure6. Status before time format translation

RFID readers were fixed along street networks, and continuously collect and broadcast several kinds of data on traffic movement. The data were processed into a separate set of files (e.g., the information of vehicles, the velocity, etc.) per link per time interval. A link represents a unidirectional road segment that could contain more than one lane, where each sensor collects data every 3 min.

Meanwhile, the images used in this work were taken from roadside video cameras installed by Chongqing Transportation Authority in their traffic management role. The video was digitized at a rate of five frames per second and stored in files using grey scale, 320 by 240 pixels, and JPEG image format. The resolution and sample rate were selected to provide sufficient detail in the image to identify individual vehicles and to capture sequential images rapidly enough so that individual vehicles can be tracked between images without pattern recognition techniques (e.g., the vehicles move no more than about one vehicle length between images). The frame of video shot was shown in Figure7 at the same time in three different points.

Figure7. Frame of the traffic surveillance video shot at the same time in three different monitoring points

Intuitively, the average speed was proposed over all days in a period. Figure 8 illustrates velocity time series measured by GPS(red curves) and RFID(green curves) chosen, and their forecasting (blue curves), such approximations adequate for traffic analysis and trend forecasting, in spite of divergences at specific points for this specific day. These differences were due to the fact that we might have just considered shorter periods, for every day, in which case the results would be more precise.

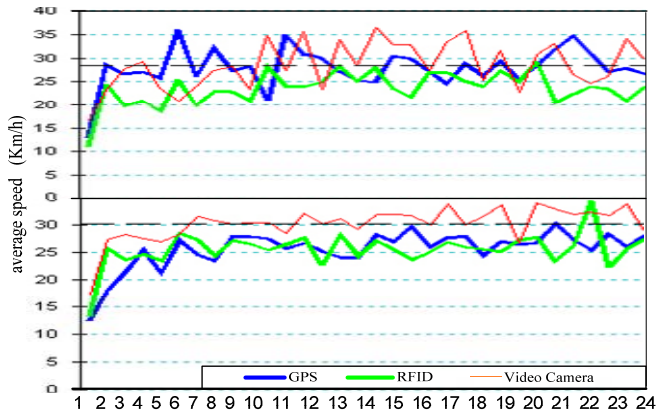


Figure8. Trend analysis of traffic status

Assume that the discernment framework of the four traffic states was S = {S1, S2, S3, S4}, and the evidence set was E = {E1, E2, E3}, where S1 → S4 orderly indicate the four different traffic states, and E1, E2, E3 denotes the evidence provided by the GPS, RFID and Video Camera data, respectively. Then the traffic condition would be shown in E-map by four different colors.

The road average-speeds and free speeds are used as the estimating evaluation index of traffic for traffic condition identification. Figure 9 shows the identification results of traffic congestion by evaluation index in an interested road segment.

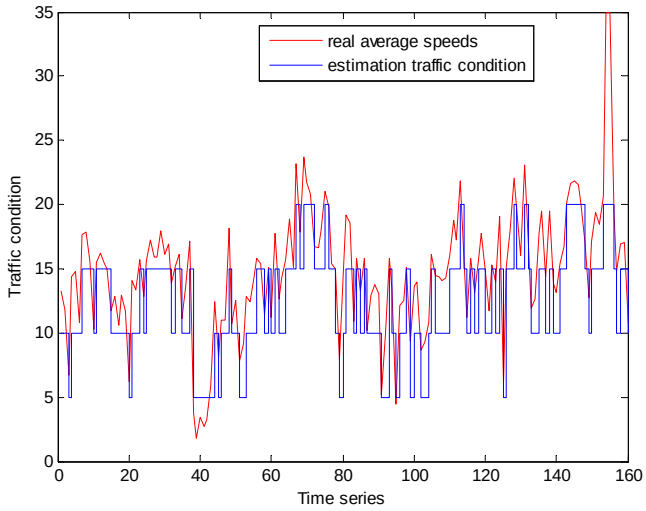


Figure9. Traffic condition and the evaluation index

The research evaluated the performance of the proposed approach was measured on the urban traffic state surveillance system. The real-time data were used to update trajectory forecast, and thus improve future trajectory studies. These predictions results were

illustrated in Figure 10, using real-time GPS-data, RFID-data and Video-data. the main urban roads for which the maximum allowed speed lied between 40 and 70 km/h, the average travel speed in the study area was often less than 40 km/h. Compared with the real traffic conditions, the integrate method was capable of capturing the spatial characteristics of the network as well as the temporal evolution of traffic in different locations in the network and giving better predictions.

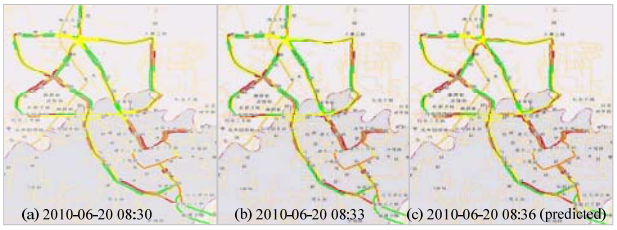


Figure10. Traffic condition at different time

The judgment process was that we first picked out the object link such as the link shown in figure 10. Then, we played the surveillance video of the first 3 min again and again. Meanwhile, we asked a skilled engineer to make a decision about the traffic state shown in the video, comprehensively considering the situation during the entire 3 min. The referenced evidences involve the length of vehicle queues, the speed of vehicles, the length of time of the red light, etc. This decision process was repeated every 3 min until the end of the video, and all these decisions were recorded. After that, we input the traffic detector data during the same period into the system with the proposed algorithm. When the computation was over, we recorded the traffic states judged by the system on this link.

Figure11 shows the comparison results between the estimated value and the true value of traffic conditions. It is obvious that the road average-speeds estimation is efficient for identifying traffic condition.

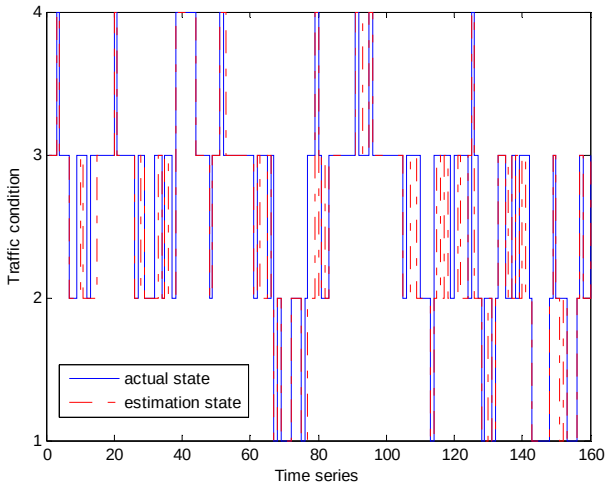


Figure11. Comparison results between the estimated value and the true value of traffic conditon

There were some data strategies based on statistics, such as used the mean or median of several individual forecasts as the final forecast. However the mean or median may not result in improvement over the individual models[30]. In this paper, the research project on Intelligent Transportation System for traffic information prediction was described. Different data source (GPS, RFID, and Video Camera) was selected to forecast the time series that were derived from the database. The system provides a flexible mechanism to deal with data, and each prediction model run as independent. It may be supported by various transit networks, forecasting algorithms or models, as well as different data sources. By analysis the forecasting performance of experiment, it can be shown that the hybrid-sensor-network can provide more accurate prediction results comparing with the original models.

V. DISCUSSION

In order to evaluate the veracity of the proposed approach, three error statistics, mean state decision error (MSDE), mean absolute percentage error (MAPE), and absolute relative error(ARE) were presented to demonstrate the predictive performance of the proposed methodology.

The MSDE was defined by the percent of the error times between the state decisions of the proposed approach and the real traffic states.

The prediction accuracy was evaluated by computing the *MAPE*, which was calculated for each bus route segment (*i*, defined as the segment between time-points *i* and *i* + 1). For comparison purpose, we had also compute the *ARE* with respect to the estimated travel speed $V_r(i,t)$ in the below equation.

$$MAPE = \frac{1}{n}\sum_{i}^{n}\frac{|t_i - t_o|}{t_o}\times 100\% \tag{18}$$

$$ARE = \left|\frac{V_e(i,t) - V_r(i,t)}{V_r(i,t)}\right|\times 100\% \tag{19}$$

Where, t_i : Predicted travel time.

t_o : observed real travel time and n is the number of samples.

$V_e(i, t)$: Estimation speed.

$V_r(i, t)$: Real traffic flow speed.

As is well know, it was difficult to obtain the real values of the spatiotemporal mean speed. The real traffic conditions were judged by skilled engineers, according to the traffic surveillance videos. Table II shown the experimental results of the accuracy measured by the three error estimates.

TABLE II
Results of accuracy

Method	MSDE (%)	MAPE (%)	ARE (%)
GPS	18.9	15.06	35.4
RFID	13.5	10.3	10.1

Camera	4.3	8.2	9.8
Hybrid sensor	2.3	5.06	6.35

The results have shown reasonably good predictive performance of the proposed forecasting strategy. The prediction errors indeed may be attributed to a number of factors, which include manual errors during data collection and possible presence of faulty loop detectors and so on. Additionally, first it should be noticed that these results were obtained from a specific setup in this specific case, and they might not be generalized in other situations.

Unlike the previous Multisource traffic data fusion methods, such as the literature [13], this method focuses on integration of the results from various detection sources, and the operation of each detection was independent each other. Therefore, it was easy to accomplish in engineering, especially when all kinds of detection technology were universally developed. Moreover, the results shown that it had stronger practicality and robustness.

VI. CONCLUSION

Traffic information forecasting is an important issue of intelligent transportation systems. Accurate prediction of traffic condition can not only help passengers time their travel plans from work places and homes and make successful transfers by reducing travel times at road, but also help transit agencies manage and operate their systems in a more responsive manner such as real-time dispatching and scheduling. For improving the accuracy and reliability, much work has been done. Multisource data fusion is the best choice. Many multisource data fusion concentrate on feature levels. Little work has been done on the decision level.

In this paper, a model was presented for real-time prediction of traffic condition, which is been implemented in an intelligent prediction system. The system is capable of tracking a large number of buses simultaneously, detecting their service routes and directions automatically, and predicting their average travel speed with an acceptable accuracy.

From the analysis, it shows that the data integration method can improve prediction efficiency of traffic condition. It is feasible to apply entropy based data fusion method to traffic information prediction. It is feasible to apply voting-theory based data integration method to traffic information prediction.

ACKNOWLEDGMENT

The authors would like to thank the Chongqing Municipal Transportation Information Centre and the Chongqing SUOMEI Intelligent Transportation Communications Services CO. LTD for providing real traffic data used in this paper.

This research is jointly sponsored by Research Fund for the Doctoral Program of Higher Education of China(RFDP) (Grant No. 20090191110022) and the Chongqing Scientific and Technological Project (Grant No. 2007ChongJiaoZi40). The writers are also grateful for the constructive comments of the referees and editors.

REFERENCES

[1] B. Ghosh, B. Basu, and M. O'Mahony. Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis [J]. *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 246–254, Sep. 2009.

[2] Eleni I. Vlahogianni, Matthew G. Karlaftis, John C. Golias. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach [J]. *Transportation Research Part C* vol.13, 2005, 211–234.

[3] Dihua Sun, Hong Luo, Liping Fu et al. An Intelligent System for Predicting Bus Arrival Time Based on GPS Data [J]. in *Transp. Res. Rec. 1879, J. Transp. Res. Board, TRB*, Washington, D.C., 2008, pp. 62–72.

[4] C. A. Quiroga and D. Bullock. Travel time studies with global positioning and geographic information systems: An integrated methodology [J]. *Trans. Res. C,* vol. 6, no. 1/2, pp. 101–127, Feb. 1998.

[5] B. R. Hellinga and L.-P. Fu. Reducing bias in probe-based arterial link travel time estimates [J]. *Trans. Res. C,* vol. 10, no. 4, pp. 257–273, Aug. 2002.

[6] Y. Li and M. McDonald. Link travel time estimation using single GPS equipped probe vehicle [C]. in *Proc. 5th Int. IEEE Conf. Intell. Transp. Syst.*, Singapore, 2002, pp. 932–937.

[7] N.C. Wu, M.A. Nystrom, T.R. Lin, H.C. Yu. Challenges to global RFID adoption[J]. *Technovation*, vol. 26, no. 12, pp. 1317-1323, Dec. 2006.

[8] B. Coifman. An intelligent traffic management expert system with RFID technology [J]. *Expert Systems with Applications*, vol. 37, no. 10, pp. 3024-3035, Apr. 2010.

[9] Y. Cho and J. Rice. Estimating velocity fields on a freeway from low resolution videos [J]. *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 463–469, Dec. 2006.

[10] B. T. Morris and M. M. Trivedi. Learning, modeling, and classification of vehicle track patterns from live video [J]. *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 425–437, Sep. 2008.

[11] Sun, D. and Fu, L. Cellular Phone Based Real-Time Bus Arrival Information System [C]. in Proc. 8th Int. *Conf. Applications of Advanced Technologies in Transportation Engineering*, Beijing, China, 2004.

[12] K. Sohn and K. Hwang. Space-based passing time estimation on a freeway using cell phones as traffic probes [J]. *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 559–568, Sep. 2008.

[13] Qing-Jie Kong, Zhipeng Li, Yikai Chen, et al. An Approach to Urban Traffic State Estimation by Fusing Multisource Information [J]. *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 499–511, Sep. 2009.

[14] Nour-Eddin EI Faouzi, Henry Leung, Ajeesh Kurian. Data fusion in intelligent transportation systems: Progress and challenges – A survey [J]. *Information Fusion* 12 (2011) 4–10

[15] S. Hashem, Optimal linear combinations of neural networks, Neural Networks 10 (1997) 599–614.

[16] N.-E. El Faouzi, Heterogeneous data source fusion for impedance indicators, in: IFAC Symposium on Transportation Systems, Chania, Greece, vol. 3, 1997, pp.1375–1380.

[17] I. Okutani, The Kalman filtering approaches in some transportation and road traffic problems. in: Gartner, Wilson (Eds.), Proceedings of the 10th ISTTT,1987, pp. 397–416.

[18] D. Huang, H. Leung, EM-IMM based land-vehicle navigation with GPS/INS, in:IEEE International Conference on Intelligent Transportation Systems,Washington, DC, October 2004.

[19] D. Huang, H. Leung, An expectation maximization based interactive multiple model approach for collaborative driving, IEEE Transactions on Intelligent Transportation Systems 6 (2005) 206–228.

[20] D. Dubois, H. Prade, Possibility Theory, Plenum Press, New York, 1988.

[21] A.P. Dempster, Upper and lower probabilities induced by multivalued mapping, Annals of Mathematical Statistics 38 (1967) 325–339.
[22] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society Series, B 30 (1968) 205–247.
[23] G. Shafer, Mathematical Theory of Evidence, Princeton 2702, Princeton University Press, 1976.
[24] N.-E. El Faouzi, L.A. Klein, O. De Mouzon, Improving travel time estimates from inductive loop and toll collection data with Dempster–Shafer data fusion, Transportation Research Record: Journal of the Transportation Research Board 2129 (2009) 73–80.
[25] N.-E. El Faouzi and E. Lefevre. Classifiers and distance-based evidential fusion for road travel time estimation [C]. in *Proc. SPIE Multisensor, Multisource Inf. Fusion: Architectures, Algorithms*, Appl., Orlando, FL, 2006, pp. 1–16.
[26] Shannon CE, Weaver W. The mathematical theory of communication [M]. *University of Illinois Press, Champaign*,1963
[27] J.N. Kapur. Entropy Optimization Principles with Applications [M]. *Boston: Academic Press*, 1992.
[28] C.E. Shannon. A Mathematical Theory of Communication [J]. *Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, July, October, 1948.
[29] Li W., M. W. Koendjbiharie, R. C. Juca, et al. Algorithms for Estimating Bus Arrival Times Using GPS Data [C]. in *Proc. 5th Int. IEEE Conf. Intell. Transp. Syst.*, Singapore, 2002, pp. 868–873.
[30] L. See and R.J. Abrahart, Multi-model data fusion for hydrological forecasting, *Computers & Geosciences*, vol.27, 2001, pp 987-994.

MA Qinglu (M'09) was born in Shan'xi, China, in 1980. He received the B.S. degree in mining engineering in 2003 from the Heilongjiang Institute of Science and Technology, Harbin, China, and he is currently working toward the Ph.D. degree in Computer Science and Technology from Chongqing University, Chongqing, China.

His main research activities are in the field of Intelligent Transportation System (ITS), Data Processing and Chaotic Prediction.