PERFORMANCE COMPARISON OF GPS-PROBE-VEHICLE BASED METHODS IN URBAN TRAFFIC STATE ESTIMATION

Qian-kun ZHAO¹, Qing-jie KONG¹, Ying-jie XIA¹,², Yun-cai LIU¹

¹ Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, 800 Dong Chuan Road, Shanghai, 200240, China; PH 021-34205294; FAX 021-34204022; email: qjkong@sjtu.edu.cn
² Hangzhou Institute of Service Engineering, Hangzhou Normal University, Hangzhou, 310012, China

ABSTRACT

Global Positioning System (GPS)-equipped probe vehicles have been one of the principal techniques in detection of traffic flow. In order to guide the application of the different methods using data from GPS probe vehicles, this paper compares two of the most popular methods using the GPS data to estimate traffic state- the method of tracking probe vehicles, and the one of fitting the model with the sample points of probe vehicles. In the experiment, data from thousands of GPS-equipped taxies were taken as the probe vehicles, and the estimate accuracy and operation speed of the two different methods were systematically measured. For the accuracy experiment, the ground truth was obtained by replaying the videos which were shot on 24 links in Shanghai downtown. The experiment results illustrate that the tracking-based method usually bears higher estimate accuracy, and slower operation speed, compared to the model-fitting method.

1. INTRODUCTION

In the past few years, many researchers have proposed different methods to estimate traffic states by using GPS information. For instance, Quiroga and Bullock (1998), Hellinga and Fu (2002) proposed an integrated GPS-GIS method to obtain travel time. However, these implementations either confine their focus to highways or assume refined time resolution of the GPS signals (Byon et al., 2006). Therefore, various methodologies and systems have been developed to solve this problem. Shi et al. (2008) and Kong et al. (2009) described an estimation method by fitting an adaptive traffic flow model with the sample points of probe vehicles. Also, (Chen et al.2007) rendered a tracking-based method to estimate the traffic state. These works discussed the theoretical basis and implementation of the two methods, and also verified feasibility of them with
the simulation results. However, they did not perform the experiments with enough ground truths of traffic state. Besides, the two methods should be compared in various road conditions to guide their application.

Therefore, this paper systematically compares the two typical methods with massive data from the GPS probe vehicles in Shanghai. The two methods are the model-fitting-based method and vehicle-tracking-based method. With an experiment system employing the latest GIS-T digital map, several examples using real GPS and GIS data are implemented to compare the accuracy and operation speed of the two methods. The ground truths were obtained by repeating the videos that were shot on 24 surface links in Shanghai downtown.

2. MODEL-FITTING METHOD

In this section, we present the model-fitting-based method evaluated in the comparison experiment.

2.1 Traffic Speed Model-The state of traffic flow on each road link during a specified period of time can often be represented by the average velocity of this road link. Generally, the velocity distribution of traffic flow has certain spatiotemporal regularity. Therefore, the velocity could be represented by time and distance.

For each link of the urban road network, the sample points of the GPS probe vehicles form a sample set \( \Psi \) (space \( l \), time \( t \), and speed \( v \)): \( \Psi = \{(l, t, v)|0 \leq l \leq L, t \in T\} \), where \( L \) denotes the length of the link, and \( T \) represents a computing period.

Then, the spatiotemporal speed distribution of the vehicles can be obtained by

\[
v(l, t) = \sum_{i=0}^{a} \sum_{j=0}^{b} c_{ij} l^i t^j
\]  

(1)

Where \( c_{ij} \) is the parameter to be solved, and \( (a, b) \) is the maximum degree of the polynomial model.

According to (1), the spatiotemporal average velocity \( \overline{\nu_G} \) along the whole link during a period can be easily obtained by

\[
\overline{\nu_G} = \left( \int_{(k-1)T}^{kT} \overline{\nu_0} dt \right) / T = \left( \int_{(k-1)T}^{kT} \nu(l, t) dt \right) / T \cdot L
\]  

(2)

2.2 Model Fitting- In order to calculate the average speed of one certain road link, the pending parameters \( c_{ij} \) should be determined first.

Formula (1) can be rewritten in vector form:

\[
v(l, t) = \varphi(l, t) \cdot \tilde{c}
\]  

(3)
Where \( \hat{c} = (c_{i0}, c_{i1}, \ldots, c_{i9}, c_{i10}, \ldots, c_{ij})^T \);
\( \overrightarrow{\varphi(l, t)} = (l, l^2 \ldots l^i, t, t \ldots t^i \ldots t^i l^j) \).

Assuming fitting a sample set from \( n \) probe vehicles, which is \( Y = \{(l_1, t_1, v_1), (l_2, t_2, v_2), \ldots, (l_n, t_n, v_n)\} \), according to (3), we can get the following expression:

\[
\vec{V} = \Phi \cdot \vec{c}
\]

(4)

Where \( \vec{V} = v_1, v_2, \ldots, v_n \); \( \Phi = \overrightarrow{\varphi}, \overrightarrow{\varphi_2}, \ldots, \overrightarrow{\varphi_n} \).

Then, \( \vec{c} \) can be obtained by solving (4) by using the least-square method.

3. TRACKING-BASED METHOD

In this section, we employ the A* heuristic search algorithm to determine the optimal tracking path, and calculate the average velocity of a road link by considering the velocities of the tracks passing along the road link, as well as their corresponding credibility factors.

3.1 Judgment of the optimal tracking path—We need to determine the vehicle driving path according to the GPS data. There are several approaching paths from the start point to the end point, and we usually repute that the path which needs shorter travel time is more credible. In fact, the driving distance and the crossroad number are two major factors to affect the time cost of a path. However, these two factors are always contradictory: a path with shorter distance usually contains more crossroads. Therefore, the factors should be traded off to determine the optimal tracking path.

The A* heuristic search algorithm is optimal upon the searching time, and can balance multiple optimum criteria by adjusting the factors of its evaluation function. Therefore, we employ the A* algorithm to judge the optimal tracking path, and we define the evaluation function for track \( k \) as:

\[
f^k(N) = g^k(N) + h^k(N)
\]

(5)

Where \( g^k(N) \) is the cost of the path from the start location to crossroad \( N \); \( h^k(N) \) is the heuristic function that estimates the cost of the cheapest path, which is the Euclidean distance from crossroad \( N \) to the end location.

More specifically, we construct

\[
g^k(N) = \overrightarrow{d^k(N)} + \overrightarrow{\alpha^k \cdot c^k(N)}
\]

(6)

Where \( \overrightarrow{d^k(N)} \) is the total journey from the start location to crossroad \( N \), \( \overrightarrow{\alpha^k} \) is the
number of crossings through which the path has already traversed, and $\alpha_k$ is the scale factor between them, which is subject to adjustment from case to case.

### 3.2 Average velocity of traffic flow

We first calculate the average velocity along track $k$ ($k = 1, 2, \ldots, m$), that is:

$$\bar{v}^k = \frac{\left(\sum_{j=1}^{n} d_j^k\right)}{(t_{end}^k - t_{start}^k)} \quad (7)$$

Where $\bar{v}^k$ is the average velocity along the track; $d_j^k$ ($j = 1, 2, \ldots, n$) is the passing distance on road link $j$; $t_{start}^k$ and $t_{end}^k$ are the time stamps.

Besides, the track $k$ may just partly cover some of the road links. Hence, it is intuitive to distribute $\bar{v}^k$ to each involved road link along with a coverage proportion. That is

$$\lambda_j^k = \frac{d_j^k}{l_j} \quad (8)$$

Where $\lambda_j^k$ ($j = 1, 2, \ldots, n$) is defined as the credibility factor of $\bar{v}^k$ on road link $j$; $d_j^k$ is the corresponding passing distance; and $l_j$ is the total length of road link $j$.

During an analysis cycle, the average velocity of a road link should consider all these distributed velocity contributions, as well as their corresponding credibility factors. That is

$$v_j = \frac{\left(\sum_{i=1}^{m} v_j^i \cdot \lambda_j^i\right)}{\sum_{i=1}^{m} \lambda_j^i} \quad (9)$$

Where $v_j$ is average speed of road link $j$; $v_j^i$ is the distributed velocity contribution of road link $j$ given by track $i$, with the corresponding credibility factor $\lambda_j^i$.

### 4. EXPERIMENTS AND DISCUSSION

In this section, the accuracy and computing speed of the model-fitting method and tracking-based method are both evaluated with the real-data experiments.

#### 4.1 Experiments Data

The GIS-T data of Shanghai used in the system contains more than 15,000 links and over 170,000 nodes, which is extremely consistent with the real urban road network. The GPS data used in these experiments are provided by Shanghai Municipal Traffic Information Center, from over 16,000 GPS-equipped taxies. Herein each GPS signal includes taxi ID, time, instant velocity, current location, and its load condition etc.

In order to obtain the real link average speed, we chose 24 surface links in Shanghai downtown, and then shot videos at the entrance and exit of these links with high clear...
cameras. By this means, we can easily find the same vehicle by matching the license plates with human eyes in the two videos recorded at the two positions. If the same vehicle is found, then we can obtain its travel time on this link. After recording the travel time of many vehicles in a computing period, we can work out an average speed on this link for one period. This average speed is taken as the ground truth in the accuracy experiment.

4.2 Accuracy Result of the Methods.-We employed historical GPS signals from June 1st to June 5th, 2010, to verify the accuracy of the two algorithms. The processing cycle was 4 minutes.

First of all, we compare the estimate results of the two methods with the real value. Figure 1 shows the comparison results on Jian-Guo-Xi-Road link from Yue-Yang Road to Tai-Yuan Road. From the figures, we can see that the two estimation results both response well to the changes of real value.

![Comparison result between real value and the methods’ results](image1)

(a) Model-fitting method  
(b) Tracking-based method

Figure 1 Comparison result between real value and the methods’ results

Next, we calculate the estimation errors of the two algorithms to compare their performance. As shown in Figure 2, the error fluctuation of the tracking-based method is much smoother than that of the model-fitting method.

![Errors of the methods on different types of road](image2)

(a) Jian-Guo-Xi Road  
(b) Jiao-Zhou Road

Figure 2 Errors of the methods on different types of road
The accuracy of methods on the different types of roads or under the different conditions of traffic flow is shown in Table 1.

**Table 1. Traffic estimation accuracy of the two methods in different situation**

<table>
<thead>
<tr>
<th></th>
<th>Model-Fitting</th>
<th>Tracking-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arteries</td>
<td>75.81%</td>
<td>83.54%</td>
</tr>
<tr>
<td>Branches</td>
<td>74.06%</td>
<td>81.25%</td>
</tr>
<tr>
<td>One-way Roads</td>
<td>78.90%</td>
<td>87.41%</td>
</tr>
<tr>
<td>Two-way Roads</td>
<td>73.31%</td>
<td>79.90%</td>
</tr>
<tr>
<td>Peak Hour</td>
<td>72.73%</td>
<td>80.86%</td>
</tr>
<tr>
<td>Non-peak Hour</td>
<td>77.72%</td>
<td>85.51%</td>
</tr>
</tbody>
</table>

From Figure 2 and Table 1, we can find that:
1) The tracking-based method has higher accuracy than the model-fitting method in all cases. The tracking-based method can generally obtain over 70% accuracy while the accuracy of model-fitting method is always over 80%.
2) For either of the two methods, the following conclusion can be drawn: the accuracy on arteries is higher than that on branches; the accuracy on one-way roads is higher than that on two-way roads; the accuracy is higher during non-peak hour than that during peak hour.

4.3 **Time cost of the methods.** We employed a personal computer with an Intel Core i5 Quad-core 2.67G processor and 4-GB DDR3 1333MHz RAM to operate the system using the two methods. The information of GPS probe vehicles was collected and processed every 4mins. The calculations of the model-fitting method in one cycle require 34 seconds, while the tracking-based method needs 71 seconds. That is because the latter method spends much more time to judge the optimal tracking path.

5. **CONCLUSION**

This paper has compared the performance of the two methods based on GPS-equipped probe vehicles in estimation of traffic state on urban surface road network. The ground truth is obtained by replaying the monitoring videos. The experiment results illustrate that the tracking-based method usually bears higher estimate accuracy and slower operation speed, compared to the model-fitting method. Therefore, in engineering practice, the application of these two methods should depend on the concrete requirement on accuracy and operation speed.
REFERENCES