Algorithms for Estimating Bus Arrival Times Using GPS Data

Li WEIGANG, Marlon W. KOENDJBIHARIE, Ricardo C. de M. JUCÁ, Yaeko YAMASHITA, and Andrew MACIVER

Abstract—This paper presents a method to estimate bus arrival times at bus stops using GPS information, which was implemented in the SITCUO - Information System for Urban Bus Transportation, in Brasilia. The model consists of a main algorithm and two sub algorithms to determine the position and the speed of the bus en route. When using the speed information from the GPS, the arrival time at the stop point will be infinite if the vehicle is stationary. With the improved method and empirical calibration, the results from the developed model were found to be satisfactory in the implementation and experiment.

Keywords— Algorithms, GPS, ITS, Transit.

I. INTRODUCTION

SITCUO, an Information System for Urban Bus Transportation was developed to provide real time information for bus passengers in Brasilia, the capital of Brazil [1, 2]. The system was developed to optimize bus operations and increase the satisfaction of urban transit users by providing estimated bus arrival information at bus stops and via the Internet. The passenger information system involves the convergence of a number of technologies including: the Global Positioning System (GPS), geographic information systems (GIS), database, data mining, telecommunications and Internet. To develop the central operating system and data bank for SITCUO it was necessary to develop a model to estimate the bus arrival time at the bus stop using the information from the

GPS [3] and the data bank. This problem can be divided into two sub problems: determining the position of the bus on the route and estimating the bus speed. The procedure adopted for solving this problem can be summarized as follows:

• Obtain the data from a text file, which contains GPS information.
• Calculate the position of the bus and the estimated time at which the bus will arrive at the next stop point. Some information is extracted from data bank.
• Store the information extracted from the text file and the results of the above step in the data bank.
• Make the results from the previous step available for the worldwide web and to the variable message sign panels.

The algorithm to determine the position of the bus was developed for one bus line, which is curved and has more than one bus in operation. This was implemented using basic mathematical models and the Geographic Information Systems (GIS) software, ArcView (v. 3.2) [4, 5].

In order to estimate the speed of the bus it is not possible to simply use the latest information from the GPS, for example, the velocity, direction and co-ordinates, which describe the current location of the vehicles. When the bus is in congested traffic or at a red traffic signal, the velocity of the vehicle can be zero. In this case, the application of a simple equation to calculate the estimated arrival time is not possible. Therefore it was necessary to develop an algorithm for calculating the vehicle speed using GPS information from a number of previous updates in the database and the average bus speed on this line.

Normally map-matching algorithms are used to estimate vehicle position and travel times [7,8,9,10]. Zhao [9] has reviewed map-matching algorithms and developed a Fuzzy-Logic-Based map matching method. Jarjees and Drane also considered more emphasis on the determinism of the movement of vehicles in their simulation study. Because the deterministic system and limitation of the information update time of this application, the developed algorithms are based on the real-time and definite bus line. The simplified consideration makes the system more efficient.

The paper is organized as follows. Section II describes the algorithms, which were developed for calculating the estimated arrival times. The main algorithm is used to
estimate the bus arrival time, and the two sub algorithms determine the position of the bus and estimate the bus speed. Section III reports on the results of the experiments using the developed model. Finally, some conclusions are discussed in section IV.

II. VEHICLE TRACKING ALGORITHMS

Three algorithms were developed for estimating the arrival time of the bus at the bus stop. The main algorithm calculates the estimated arrival time and the two sub algorithms determine the position of the bus and estimate the speed of the bus along its route. These algorithms are described below.

A. Main algorithm

In order to calculate the estimated arrival time, the bus route is divided into a number of short, straight line, sub-routes, shown in Figure 1. In mathematical terms these lines can be modeled as first-degree equations in a plane. The following steps were performed to obtain the required results:

1) Receive information on the position of the bus;
2) Calculate the distance from this point to all the lines on the route (to determine where the bus is located);
3) The bus is assumed to be located at M, on the line PQ. The line $r$ (NM) is the smallest of all the calculated distances to all lines on the route (see Figure 1);
4) Calculate the equation of the line $s$ (PQ), perpendicular to the line $r$, which passes through the point N, the bus's actual position;
5) Calculate the coordinates of the point M, which is the intersection of the lines $r$ and $s$;
6) For calculation purposes the bus is assumed to be located at point M;
7) Calculate the distance $d(Q,M)$ from point M to the end of the sub-route, which is a bus stop;
8) Knowing the bus's speed ($v$), the time ($t$), which is the time taken to travel to the end of the sub-route, can be calculated using the following equation:

\[ v = \frac{\Delta S}{t} \Rightarrow t = \frac{\Delta S}{v} \]

The methodology used to calculate the bus arrival time is presented below.

B. Algorithm to determine the position of the bus

The GPS equipment in the bus transmits its position, speed and other related information to the control center by some means of data communication. It is unlikely that this position will coincide with any point on the straight-line graph, since the graph is an approximation of the actual curved route. The first step is to map this actual position to a point on the graph. This line is the link between the points P and Q (see Figure 1). The point N is the actual position of the bus and M is the orthogonal projection of N on the link between P and Q.

Let $d(X,Y)$ now represent the function that calculates the distance between two points in the plane. If M is actually located on the link between P and Q, then M must be between these two points.

It can be seen that $(d(P,Q) > d(P,M)$ and $d(P,Q) = d(Q,M))$ or $(d(P,Q) = d(P,M)$ and $d(P,Q) = d(Q,M))$. This will help to determine where the bus is located on the graph.

In Figure 2 we can see what happens when the bus has already passed the sub-route linking P and Q. In this case, the relation $(d(P,Q) > d(P,M)$ and $d(P,Q) = d(Q,M))$ or $(d(P,Q) = d(P,M)$ and $d(P,Q) = d(Q,M))$ is not applicable, which shows that this cannot be the link where the bus is located. If the bus found itself exactly on P or Q, in this case $d(P,M) = 0$ and $d(Q,M) = d(P,Q)$, the relationship is still satisfied.
required. To calculate $s$ we observe the following cases: 1) $r \perp y$-axis; $P \in r$; 2) $r \perp y$-axis; $P \notin r$; 3) $r \perp x$-axis; $P \in r$; 4) $r \perp x$-axis; $P \notin r$; 5) $r \not\perp x$-axis; $P \in r$; 6) $r \not\perp x$-axis; $P \notin r$ [6, 2].

1) Case 1, $r \perp y$-axis, $P \in r$ (Figure 3)

In Figure 3 it can be seen that $P$ and $Q$ are points on line $s$, and that $Q(i,j) = Q(i,0)$, $Q(i,j) = Q(e,j)$, and finally that $Q(i,j) = Q(e,0)$.

Resolving the system we obtain:

\[
\begin{cases}
  e = a + v(a - c) \\
  f = b + v(b - d)
\end{cases} \Rightarrow \begin{cases}
  v = \frac{e - a}{a - c} \\
  v = \frac{f - b}{b - d}
\end{cases}
\]

Because $b - d = 0$, we can say that $v = \frac{e - a}{a - c}$; if $P \in r \Rightarrow f = b$.

Therefore $s : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix} + w \begin{pmatrix} g \\ h \end{pmatrix}$ where $g = e$; $h = 0$;

\[e(\xi) = (\xi) + v(\xi)_{\text{gb}}\]

Fig. 3. Case 1, $r \perp y$-axis, $P \in r$

2) Case 2, $r \perp y$-axis, $P \notin r$ (Figure 4)

\[e(\xi) = (\xi) + v(\xi)_{\text{gb}}\]

Fig. 4. Case 2, $r \perp y$-axis, $P \notin r$

In this case $b - d = 0$, $f \neq b \Rightarrow P \notin r$, and find that $P(e,f)$ and $Q(e,b)$ and finally $s : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix} + w \begin{pmatrix} e \\ b \end{pmatrix}$.

Since $P \in r$ then $r : \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} a \\ v(a - b) \end{pmatrix}$

And since $r \perp y$-axis then: $b - d = 0$; $a - c \neq 0$.

Line $s$ has $P(e,f)$ and $Q(e,0)$ as two of its elements. Algebraically, the equation of a straight line can be determined from two of its elements.

3) Case 3, $r \perp x$-axis, $P \in r$ (Figure 5)

Since $r \perp x$-axis we have, $a - c = 0$; $b - d \neq 0$.

Therefore $s : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ 0 \end{pmatrix}$.

4) Case 4, $r \perp x$-axis, $P \notin r$ (Figure 6)

Since $r \perp x$-axis, $a - c = 0$; $b - d \neq 0$, $\Rightarrow e \neq a$; it can be seen that $P(e,f)$ and $Q(e + 1,f)$. (In this case $e + 1$ was chosen in $Q$ but in fact this could have been any number).

We find that $s : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix} + w \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

5) Case 5, $r \not\perp x$-axis, $P \in r$ (Figure 7)

Here $a - c \neq 0$; $b - d \neq 0$ and $r \not\perp x$-axis $\Rightarrow v = \frac{e - a}{a - c}$ or $v = \frac{f - b}{b - d}$, but these two equations are exactly the same, because $P$ and $r$ can have only one common element which is $P$ itself. Since $r$ and $s$ are perpendicular, $g = b - d$; $h = -(a - c)$.

$P(e,f)$ and $Q(e + (b - d), f-(a - c))$

The result is $s : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix} + w \begin{pmatrix} d - b \\ a - c \end{pmatrix}$.

6) Case 6, $r \not\perp x$-axis, $P \notin r$ (Figure 8)

In this case, $a - c \neq 0$; $b - d \neq 0$, $\Rightarrow v = \frac{e - a}{a - c}$ or $v = \frac{f - b}{b - d}$, as in case 5 it can be seen that $P(e,f)$ and $Q(e + (b - d), f-(a - c))$

Finally $s : \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix} + w \begin{pmatrix} d - b \\ a - c \end{pmatrix}$.

After determining the position of the point in relation to the line which is the approximation of a section of the bus line, the problem is reduced to calculating the distance between two points: $(x_1, y_1)$ and $(x_2, y_2)$ using the following equation:
C. Algorithm to estimate the speed of the bus along its route

In a particular case where the bus transmits a speed, \( v = 0 \), the travel time, \( t \), would be infinity. This may occur for a number of reasons, for example, when the bus arrives at a bus stop or is stopped at traffic signals. Use of the actual speed of the bus in isolation would also result in significant fluctuations in determining the expected arrival time \( t \).

To resolve this problem, two kinds of speeds are used to calculate the speed \( v \). This method considers the average speed \( V_M \) of the bus, which is obtained from historical data collected from buses during similar periods. The other is the actual speed of the bus \( V_R \), transmitted to the Control Center.

Using these two different speeds the speed \( v \) can be calculated. This will be used in further calculations to obtain the time \( t \) at which the bus will arrive at the bus stop.

In order to determine the distance between the vehicle (at point \( P \)) and the bus stop \( F \) (See Figure 9) it can be seen that:

\[
d(P,F) = d(P,B) + d(B,C) + \ldots + d(E,F).
\]

The speed \( v \) can be used to determine \( t \). In Figure 9, \( n \) is the number of nodes on the route and the time \( t \) at which the bus will arrive can be calculated.

The main advantage of this method is that it is possible to provide the passengers with more reliable and accurate information. It can be observed that the longer the distance the bus is from the bus stop, the more the speed \( v \) will depend upon the average speed \( V_M \) and less on its actual speed \( V_R \), and the closer the bus is to the bus stop, the less its speed \( v \) will depend upon the average speed \( V_M \) and the...
more on its actual speed $v_R$. In this way the $v = 0$ problem and the resulting fluctuations for time ($t$) can be overcome.

![Diagram](image)

Fig. 9. Demonstration of the velocity calculation

### III. EXPERIMENTS

#### A. Bus line 0.105

The 0.105 - Grande Circular bus route in the Plano Piloto, Brasilia was selected to test the system. The selection of this route was based on three main criteria:

- **Passenger demand:** according to the Metropolitan Department of Urban Transportation of the Federal District (DMTU-DF), route 0.105 has the highest number of daily passengers for all current bus routes in Plano Piloto, Brasilia. This route also has the greatest potential for future expansion.

- **Scale:** with a total length of approximately 32 Km, this route has a total of 64 bus stops, passing through Plano Piloto in an anti clockwise direction. During peak periods, the operating company schedules a high number of vehicles and this was deemed to provide the ideal conditions for testing and validating the system (performance, reliability, accuracy etc.).

- **Geometry of the route:** to calculate the bus arrival time ($D_t$) at the bus stop, considering the fact that the calculations will be performed concurrently for all the circulating vehicles at that moment, it is necessary to consider two other inputs: the speed ($v$) of the vehicle and the distance ($D_o$) from the vehicle to the bus stop. Having an irregular route (not a straight line), calculating the distance is more difficult. First, the route has to be adapted for this purpose and represented as a graph.

#### B. Test results

The central control system receives data from the vehicle, including its latitude, longitude and speed. For every bus stop the speed is calculated using the algorithm that was presented earlier. The distance to the bus stop is also calculated and using this data the estimated arrival time at the bus stops can be calculated.

It does not make much sense to specify the time of arrival in seconds, because the vehicle does not move at a constant speed. It is widely recognized that better customer satisfaction can be achieved by specifying predicted arrival times in minutes. The arrival time is calculated in minutes and rounded up when it is half a minute or above. For example, a predicted time of 6.876523 minutes would be displayed as 7 minutes, and 6.472448 as 6 minutes.

SITCUO transforms the data from the GPS (Latitude/Longitude) using the Universal Transverse Mercator (UTM) coordinate system. The 0.105 bus route has 64 bus stops. Table 1 shows four of the route sections with the calculated distance and the actual (measured) distance between consecutive bus stops. This is sufficient for the purpose of showing the slight differences between the actual and calculated distances. For all of the 64 lines of data from this table the related error is less than 8% and the standard deviation of the related error is 0.11. This means that the results of the tests are within acceptable limits and the accuracy of the system is not compromised.

<table>
<thead>
<tr>
<th>Initial stop point</th>
<th>Final stop point</th>
<th>Actual distance (m)</th>
<th>Calculated distance (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>610 North</td>
<td>610/611 North</td>
<td>300</td>
<td>309</td>
</tr>
<tr>
<td>610/611 North</td>
<td>612/613 North</td>
<td>400</td>
<td>405</td>
</tr>
<tr>
<td>612/613 North</td>
<td>614/615 North</td>
<td>1100</td>
<td>1085</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>609 North</td>
<td>610 North</td>
<td>350</td>
<td>374</td>
</tr>
</tbody>
</table>

### IV. CONCLUSIONS

As an initial proposal, the developed model is capable of estimating the arrival times of the buses at all the bus stops along the 0.105 bus route in Grande Circular, Brasilia. From the implementation and application of the algorithms, the following conclusions and recommendations are presented.

- During the experiments it was observed that the model provides satisfactory results. The types of algorithms, which are implemented with the support of mathematical concepts make it possible to apply the method for many other purposes, for example, the “shortest path algorithm”.

- The equation to estimate the speed of the vehicle is based on an empirical proposal and the calibration during the practical experiment. Further development of the theory and structure of the model is required.

- The mean relative error between output results from system and the actual position of bus is less than 8%. There are still some opportunities to reduce these errors by increasing the number of lines representing the bus route.

- The proposed SITCUO model has great potential to make the urban public transportation system more attractive to its users. Apart from the arrival times, the system also provides operational reports, which make it possible to monitor vehicle operations with no
addition costs to the bus company. However a suitable communication system must be found for communicating between the central control system and the variable message boards at the bus stops.

REFERENCES


