A Two-Stage Extended Kalman Filter Algorithm for Vehicle Tracking from GPS Enabled Smart Phones through Crowd-Sourcing

Mukesh kiran K¹, Nagenra R. Velaga²*, RAAJ Ramasankaran¹,

¹Remote Sensing division, Civil engineering department, IIT Bombay, Mumbai-400076, India.
²Transportation systems engineering, Civil engineering department, IIT Bombay, Mumbai-400076, India.
*Corresponding author contact details: n.r.velaga@iitb.ac.in Tel: +91-22-2576-7341

Abstract

Crowd-sourcing using smart-phones is one of the emerging areas of Intelligent Transport System (ITS); for instance, tracking public transit passengers’ smartphone GPS locations during their journeys and using it as an Automatic Vehicle Location (AVL) system for public transport vehicles without any positioning systems. However, positioning data obtained from mobile phone GPS are erroneous. With reference to improving the accuracy of vehicle positioning system, many existing research efforts mainly devote to various sensor integration methods (e.g., tight integration of GPS with dead-reckoning sensors). In the existing state-of-the-art literature, enhancement of erroneous GPS locations obtained from smartphone based crowd-sourcing and further integration of individual location information from smartphones for identifying AVL of public transit vehicles was not focused. Accordingly, in this study it is aimed to improve the vehicle positions obtained from multiple mobile GPS receivers (i.e., location data received from public transport passengers (e.g., bus) through Crowd-sourcing) with a two-stage Extended Kalman Filtering (EKF) technique. In the two-stage process, initially, EKF is applied to each GPS epoch separately then further the estimate is combined using the best linear unbiased estimator or Kalman Filter (KF). Total 20 bus routes, covering 341.35 Kms length in Mumbai, India, were selected and crowd-sourcing based GPS positioning was collected. A high accuracy Trimble Juno and GeoXT GPS was considered as reference positioning points on all the bus routes. The data collected was processed through the two-stage EKF. After processing the collected data, it was found that the average error for all bus routes varies between 9.21m to 12.22m (with 95% confidence level).

Keywords: Extended Kalman Filter; GPS; Crowd-Sourcing; Intelligent Transport System.

1. Introduction

Rapid industrialization around the world for urban development has increased the need for improved transportation facilities with ITS applications for efficient transportation planning. Tracking and navigation are important requirements of many location-based ITS applications. Due to recent advances in technology, smartphones with GPS receiver are common and are used by many people travelling on buses. Different smartphones have diverse GPS modules with a wide range of error
covariance. Research on improving the GPS data has been going on from several years using many advanced techniques.

Kalman filter is one of the well-known techniques used for real-time navigation system. Kalman Filters have a long history of accurately predicting the future states of a moving object and has been applied to many different fields (Barrios et al., 2011). Kalman Filters are used for integration of Inertial Measurement Unit (IMU), Inertial Navigation System (INS) or Dead Reckoning (DR) with GPS data for navigation applications (Ge et al., 2001; Li et al., 2002; Quddus 2006; Zhao et al., 2002; Zhang et al., 2005; Krakiwsky et al., 1988). Extended Kalman Filter (EKF) is an extension of KF which integrates the IMU/INS/DR with GPS data and gives the best estimate of the vehicle location when there is positioning data coming from different sensors.

In the present research, effort has been put to formulate, implement and validate an algorithm to track buses using GPS data obtained through crowd-sourcing from passengers’ GPS enabled phones. Telecommunication network enables us to share the positioning data during travel on public transportation. Steinfeld et al. (2011) and Zimmerman et al. (2011) demonstrated a system for collecting user location data during their travel using crowd-sourcing. Crowd-sourcing is a method which can provide GPS data from multiple users. The positioning data from GPS enabled smart-phones of the commuters in public transport can be used as data source for EKF system to track public transport (e.g., buses). EKF system computes the best estimate of the vehicle position from the crowd-sourcing data. Hence, developing an EKF system which can show the vehicle location with less uncertainty and cost effective by collecting data from the public transport users will improve transport planning and support a range of location based ITS applications. By using EKF system, it is possible to assimilate positioning data obtained from multiple smart-phones with different accuracies and estimate vehicle position (for each time frame) with better accuracy and continuous GPS track of public transport.

2. State-of-the-art Literature

For real-time tracking of vehicles, various positioning systems/sensor’s are integrated to obtain vehicle location information; and examples of such sensors are: (1) Deduced Reckoning (DR) system (2) Ground Based Beacon System 3) Global Navigation Satellite System (GNSS) like GPS (Chang et al., 2010; Leung et al., 2011; Shen et al., 2011; Velaga et al., 2012a; Kalpan and Hegarty 2006). Many literatures on GPS and state estimation techniques (such as Kalman Filter, Extended Kalman Filter and Particle Filter) provide detailed overview on the method used for estimating vehicles locations and its application in ITS. GPS data is associated with many errors due to interference caused during the time the signal reaches the GPS receivers. Some of the errors are multipath effect, atmospheric effect, clock error and errors due to satellite geometry. Many literatures like Lahrech et al. (2005), Ge et al. (2001), Salmon et al. (2014), Barrios et al. (2011), Krakiwsky et al. (1988), Zhao et al. (2003), Mosavi (2005), Ding et al. (2007) and Li, et al. (2006) focused on improving the GPS accuracy by integrating external sensors like (DR, IMU or INS) with the GPS data using state estimation techniques.
Mobile phones have become one of the most significant electronic devices in the 21st century. Mobile phones play an important role in day-to-day life; hence it has reached global proliferation. Also, since all the latest smart-phones are GPS enabled, Location Based Services (LBS) using smart-phones are increasing. Recent developments in mobile and information technologies could allow passengers to share information about their journeys and track their location during travel on public transportation (Maclean and Dailey 2001; Politis et al. 2010; Velaga et al. 2012b and Velaga et al. 2014). Zandbergen et al. (2011) determined the performance of the GPS enabled smart-phones under varying conditions like static and dynamic outdoor conditions and compared with regular recreational grade GPS units, which was used to determine positional accuracy. In static outdoors tests, median of horizontal error of position fixes from mobile phones varied between 5 to 8.5m; and the maximum error is below 30m. Whereas, in dynamic outdoor tests, the GPS fix was available 99.8 percent of time with a maximum error of 19.73m (Zandbergen and Barbeau, 2011). Since the GPS signals have errors associated with it, there is a need to reduce the error in the GPS signals. Hence state estimation techniques are used to reduce the error associated with the GPS signals.

Salmon et al. (2014) experimented with Vehicle Model (VM) to replace low cost IMU with GPS results yield the conclusion that navigation using a VM sensor set can drastically improve the position, velocity and attitude solution whether aiding or replacing the IMU. Another goal of their research was the quantitative and qualitative analysis of the benefits of using VM to assist normal GPS/INS EKF and whether the inclusion of VM in either the time update or the measurement update results in a more accurate pose solution. In Du et al. (2014) an EKF was used to estimate the real-time position of the vehicle in the curve sections. The Global Navigation Satellite System (GNSS) is always unavailable near black spot such as the curves of the mountainous regions (particularly during the rainy seasons) which needs to be solved. Therefore, in this literature, a novel Cooperative positioning (CP) scheme is proposed for the curve warning scenario with limited GNSS by utilizing the information of received signal strength and pointer angular of the roadside unit. Here pointer angular and signal strength of the roadside unit are used as observation data in EKF. CP is one of the core features in intelligent transportation systems (ITS) which are used to increase the positioning accuracy via wireless communication between vehicles and infrastructures. Zhao et al. (2003) developed an EKF which was used for integrating GPS with low cost DR sensors for vehicle performance and emission monitoring. Simulation results showed that DR sensors can provide improved heading and velocity information calibrated by Kalman Filter based on the GPS signal whenever available. Field test were carried out in the Grater London for validating the developed EKF. A test vehicle which was equipped with a navigation platform consisting of a 12-channel single frequency GPS receiver, a low-cost MEMS rate gyroscope and the interfaces required to connect to the vehicle speed sensor (odometer) and back-up indicator. The route in London was chosen carefully to have a good mix of important spatial urban characteristics including open spaces, urban canyons, tall buildings, tunnels, bridges, and potential sources of electromagnetic interference.

Barrios et al., (2011) used Kalman Filter to improve estimation of vehicles trajectory using advanced GPS equipment. To obtain accurate prediction of the vehicles future location, four adaptive prediction algorithms were defined to account the possible scenarios like constant location, constant velocity, constant acceleration and constant
jerks. Four models were developed to account for the four scenarios where the vehicle can be found on the road. These independent models were merged to give single prediction by Interacting Multiple Model (IMM) algorithm where the probability of each model is calculated and highest probable model for the location is chosen and used in KF. Further, the estimate was tested whether the estimated location is on the road or not, then the error determined is corrected using the GIS data. Lahrech et al., (2005) explains a new method to accurately estimate the vehicle position in urban areas. This method uses EKF to find the final estimate with multiple data source like GPS, DR and GIS road map. A fusion algorithm was formulated for urban area scenario to find the accurate position of the vehicle. As per the Fusion algorithm, when the filter is initialized, EKF predicts the vehicle position and the predicted position is corrected using GPS data if the GPS data is available for that location. If the GPS position is not available at that location the filter makes use of the DR data and corrects the prediction using the DR data and also further corrects the estimate using the GIS road map data to improve the accuracy of the estimate.

In order to analyze positioning data in real-time, Steinfeld et al. (2011), and Zimmerman et al. (2011) developed a system to obtain positioning data from the users travelling in public transport. Zimmerman et al. (2011) illustrated the potential for crowd-sourcing bus arrival information. Zimmerman et al. (2011) conducted field trials with 28 participants and identified that participants allowed their GPS location to be traced for more than half of their rides and the user-contributed information system could create a crowd-sourcing social computing system (Zimmerman et al. 2011). However, Zimmerman et al. (2011) did not demonstrate how to integrate vehicle location information from different passengers traveling on the same transit vehicle in order to achieve a precise vehicle location on a GIS road map.

3. Methodology

An EKF algorithm was used to estimate the bus location using the positioning data obtained from GPS enabled smart-phones. In this paper, a new method was formulated to estimate the bus location with passengers GPS enabled smart-phones. This has two stage filtering process: (1) EKF is applied individually to the GPS data from users smart-phone and (2) the EKF estimate thus obtained from the individual users smart-phone data is combined using a linear filter in the second stage. By doing the two-stage filtering the error covariance associated with the estimated GPS location is minimized thereby enabling us to achieve better accuracy than the individual GPS accuracy of smart-phones.

3.1. Formulation of EKF for tracking vehicle

State equation:

The dynamics of a system is based on the knowledge of how the system behaves with the time. Here, the dynamics of the vehicle movement is based on the knowledge of how the vehicle is expected to move. Hence the vehicle dynamics was established as state equations with the following states (Zhao et al.2003; Quddus, 2006; and Davidson et al. 2011).
Where

$X$ state vector

$e$ terrestrial easting position in meters

$n$ terrestrial northing position in meters

$v_v$ velocity of the vehicle in m/sec. 

$H_v$ Heading of the vehicle in radians, with north as zero degrees.

$a$ acceleration of the vehicle in m/sec$^2$.

$\omega$ rate of change of heading in radian/sec.

Dynamic equation can be written as

$$
\begin{align*}
\dot{e} &= v_v \cdot \sin H_v + w_1 \\
\dot{n} &= v_v \cdot \cos H_v + w_2 \\
\dot{v}_v &= \dot{a} + w_3 \\
\dot{H}_v &= \dot{\omega} + w_4 \\
\dot{a} &= w_5 \\
\dot{\omega} &= -\beta_\omega \omega + w_6
\end{align*}
$$

Or can be written in vector form as

$$
\begin{align*}
x &= f(x(t), t) + w
\end{align*}
$$

Measurement equation:

From the GPS receiver measurements like position of the vehicle, velocity of the vehicle and heading of the vehicle can be derived. Hence the observation vector is represented as

$$
Z = [e \ n \ v_{GPS} \ H_{GPS}]^T
$$

The measurement equations are defined as

$$
\begin{align*}
e_{GPS} &= e_{GPS} + v_1 \\
n_{GPS} &= n_{GPS} + v_2 \\
v_{GPS} &= v_{GPS} + v_3 \\
H_{GPS} &= H_{GPS} + v_4
\end{align*}
$$

Where,

$e_{GPS}$ is Easting position from GPS receiver in meters

$n_{GPS}$ is Northing position from GPS receiver in meters

$v_{GPS}$ Velocity measurement from GPS receiver in meter/sec

$H_{GPS}$ Heading measurement from GPS receiver in degrees.

$V_1, V_2, V_3$… are the measurement noise

The above equations can be rewritten as

$$
Z = hx + v
$$
2.2 Two Stage Extended Kalman Filter

In this method, the GPS data collected from the smart-phones is passed through the EKF individually and the estimate obtained is further filtered using the best linear unbiased estimator (Kalman Filter) to get an estimate, which has less uncertainty in the position of the vehicle. The proposed methodology is shown in Figure 1. The main aim of developing the two-stage EKF is to reduce the error further in the second stage of filtering. After the first stage of filtering is carried out, the estimates from each of the EKF have an error covariance. In order to further reduce the error covariance all the estimates are further combined using KF. For this study three smart-phones were used to collected the data hence the data from the each smartphone is filtered individually as show in Figure 1, the estimates of the three smartphone positioning data thus obtained is further filtered using the KF which will further reduce the uncertainty or the error covariance associated with the estimates obtained from EKF.

Figure 1: Two Stage Extended Kalman Filter flow chart
3.2 Algorithm Development

Matlab code was developed with the above algorithm to estimate the location of the vehicle and was validated with the simulated location data. Further this code was used to estimate the vehicle location for the real-world data collected from the mobile phones. High end GPS receivers were used to validate the performance of the developed filter in the urban areas with dense roads, bridges, flyovers, junctions and roundabouts. The performance of the developed algorithm is quantified by estimating the Euclidean distance between the two stage EKF position estimate and the reference position collected from the high end GPS receivers like Trimble Juno and Trimble GeoXT. Along-track error and cross-track error of the route are calculated as shown in Figure 2 (Velaga et al, 2009). Let A be the Reference GPS position and B is the two stage EKF estimate, then AO is the euclidean distance measured from the GPS reference position to the EKF estimate along the movement of vehicle direction; hence it is called along-track error. BO is the Euclidean distance between the reference GPS position and the EKF estimate and is called cross-track error. Average error is the Euclidean distance measured from point A to B.

![Figure 2: Performance measurement for the collected data](image)

The GPS data collected from the smart-phones are erroneous and the error is different for different devices based on the GPS chip available in the smart-phone. These error values are represented in a matrix form which is called Measurement covariance matrix (R Matrix). The covariance matrix was calculated using the median of accuracy values obtained from the “GPS logger” android app in all the phones in all routes for covariance in x and y direction.

4. Data Collection

GPS positioning data was collected in 20 bus routes in Mumbai and Navi-Mumbai region (in India) covering 341.35 km road length as shown in Figure 3. The test routes were selected in such way that it is a mix of typical urban scenarios such as flyovers, junctions, intersections, roundabouts and roads near tall buildings. While collecting the data, information such as latitude, longitude, heading and speed of the vehicle is
recorded. The data were collected using multiple GPS enabled smartphones and high end GPS devices like Trimble Juno and Trimble GeoXT which is taken as the reference data for the vehicle location. Data were collected from 25th March, 2015 to 9th April, 2015 using three smart-phones with two people carrying the same (which replicated the crowd sourced data from a public transport vehicle). For this study authors had collected the GPS data from AC and non AC buses to check and validate the signal reception on both. Trimble Juno and Trimble GeoXT GPS receivers were used for collecting the reference data for the corresponding data collected using the smart-phone. GPS Trimble Juno receiver, showed error of more than 5m for 51.72% of the points after differential correction for 470 and AS524 bus routes, same level of accuracy was observed on other routes where Trimble Juno was used. The data accuracy collected from the Trimble GeoXT was between 30cm – 50cm for 44.11% of the GPS points. Table 1 shows the bus route information travelled while collecting data.

Table 1: Mumbai and Navi-Mumbai bus routes in which data was collected

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Route No.</th>
<th>Origin</th>
<th>Destination</th>
<th>Length(Km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65</td>
<td>Cadbury Junction</td>
<td>Borivali</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>A75Exp</td>
<td>Powai</td>
<td>Nehru Planetarium</td>
<td>28.9</td>
</tr>
<tr>
<td>3</td>
<td>AC105</td>
<td>CBD</td>
<td>Vashi</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>138</td>
<td>Borivali</td>
<td>Kharghar</td>
<td>54</td>
</tr>
<tr>
<td>5</td>
<td>86</td>
<td>Nehru Planetarium</td>
<td>Backbay Depot</td>
<td>10.9</td>
</tr>
<tr>
<td>6</td>
<td>470</td>
<td>Ghatkopar</td>
<td>Borivali</td>
<td>23.91</td>
</tr>
<tr>
<td>7</td>
<td>496</td>
<td>TeenathNaka, Thane</td>
<td>IIT main gate</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>AS524</td>
<td>Borivali</td>
<td>IIT main gate</td>
<td>18.37</td>
</tr>
<tr>
<td>9</td>
<td>313</td>
<td>Kurla</td>
<td>Santacruz</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>505</td>
<td>Priyadarshini</td>
<td>CBD</td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td>545LTD</td>
<td>IIT</td>
<td>Andheri</td>
<td>8.79</td>
</tr>
<tr>
<td>12</td>
<td>324</td>
<td>Worli depot</td>
<td>Marol</td>
<td>19.08</td>
</tr>
<tr>
<td>13</td>
<td>225</td>
<td>Bandra</td>
<td>Andheri</td>
<td>8.149</td>
</tr>
<tr>
<td>14</td>
<td>395</td>
<td>santa cruz</td>
<td>Marol</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>AS333</td>
<td>IIT maingate</td>
<td>Ghatkopar</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>185</td>
<td>Sakinaka</td>
<td>Panchkutir</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>2LTD</td>
<td>Andheri</td>
<td>Colaba depot</td>
<td>27.09</td>
</tr>
<tr>
<td>18</td>
<td>444LTD</td>
<td>Ghatkopar</td>
<td>Goregoan</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>AS3</td>
<td>Cadbury Junction</td>
<td>Priyadarshini</td>
<td>21</td>
</tr>
<tr>
<td>20</td>
<td>305</td>
<td>Worli</td>
<td>Ghatkopar</td>
<td>8.161</td>
</tr>
</tbody>
</table>

In order to find position fix, almanac data is needed. In different smartphones almanac data was received at different times from the start of the bus journey depending on the sensors and satellite availability to the sensors in the smartphones. In current generation smart-phones, the position fix is faster because of the assistive GPS technology where the position fix is obtained from the network triangulation method. Hence, in different routes the data logging started at different instants because of the time lag in position fix in different GPS sensors provided in the smart-phones. The mobile data was switched off while collecting the data in the smart-phones. The bus
routes were chosen such that, data covers all types of roads in Mumbai and Navi Mumbai. The location data was collected using multiple android based GPS enabled smartphones like XiaoMi Mi3, Samsung galaxy S3, Samsung galaxy duos, Samsung Galaxy Core I8262, Sony xperia, Redmi Note 3G. These phones have different GPS chips in it; hence the error variance of the phone varies from each other. The positioning data like latitude, longitude, heading, time and speed were collected using a standalone GPS logger app for android platform for every second.

Figure 3: Bus route map in Mumbai city

5. Results and Discussion

The two-stage EKF was applied to a total of 20 BEST (Brihanmumbai Electricity Supply and Transport) and NMMT (Navi Mumbai Municipal Transport) bus routes in Mumbai and Navi Mumbai region (i.e., from Table 1); the summary of the results obtained are shown in Table 2. Along track error, cross track error and average error are the quantities by which the performance of the two-stage EKF is measured. The analysis was carried out for a total length of 341.35 Km. The results show the performance of the developed filter in major parts of the city like intersections, roundabouts, flyovers, crowded areas, bridges, interchanges etc.,
Table 2: Summary of the results obtained from two stage EKF filter

<table>
<thead>
<tr>
<th>S.No</th>
<th>Route No.</th>
<th>Length(Km)</th>
<th>Along Track Error(m)</th>
<th>Cross Track Error(m)</th>
<th>Average Error(m)</th>
<th>Average speed(Kmph)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>65</td>
<td>26</td>
<td>2.582</td>
<td>2.693</td>
<td>4.138</td>
<td>39.5</td>
</tr>
<tr>
<td>2</td>
<td>A75Exp</td>
<td>28.9</td>
<td>3.732</td>
<td>3.639</td>
<td>5.725</td>
<td>25.77</td>
</tr>
<tr>
<td>3</td>
<td>AC105</td>
<td>10</td>
<td>4.292</td>
<td>3.852</td>
<td>6.428</td>
<td>29.616</td>
</tr>
<tr>
<td>4</td>
<td>138</td>
<td>54</td>
<td>4.973</td>
<td>4.804</td>
<td>7.645</td>
<td>28.177</td>
</tr>
<tr>
<td>5</td>
<td>86</td>
<td>10.9</td>
<td>5.07</td>
<td>5.513</td>
<td>8.297</td>
<td>23.707</td>
</tr>
<tr>
<td>6</td>
<td>470</td>
<td>23.91</td>
<td>4.788</td>
<td>6.713</td>
<td>9.05</td>
<td>27.261</td>
</tr>
<tr>
<td>7</td>
<td>496</td>
<td>10</td>
<td>6.06</td>
<td>5.774</td>
<td>9.282</td>
<td>41.204</td>
</tr>
<tr>
<td>8</td>
<td>AS524</td>
<td>18.37</td>
<td>7.102</td>
<td>5.321</td>
<td>9.754</td>
<td>30.351</td>
</tr>
<tr>
<td>10</td>
<td>505</td>
<td>22</td>
<td>6.067</td>
<td>7.311</td>
<td>10.513</td>
<td>31.136</td>
</tr>
<tr>
<td>11</td>
<td>545LTD</td>
<td>8.79</td>
<td>8.753</td>
<td>11.94</td>
<td>16.514</td>
<td>16.723</td>
</tr>
<tr>
<td>12</td>
<td>324</td>
<td>19.08</td>
<td>8.198</td>
<td>6.554</td>
<td>11.565</td>
<td>16.346</td>
</tr>
<tr>
<td>13</td>
<td>225</td>
<td>8.149</td>
<td>7.422</td>
<td>7.492</td>
<td>11.676</td>
<td>17.234</td>
</tr>
<tr>
<td>14</td>
<td>395</td>
<td>10</td>
<td>6.957</td>
<td>8.055</td>
<td>11.726</td>
<td>20.9</td>
</tr>
<tr>
<td>15</td>
<td>AS333</td>
<td>5</td>
<td>8.428</td>
<td>6.858</td>
<td>12.023</td>
<td>19.816</td>
</tr>
<tr>
<td>16</td>
<td>185</td>
<td>5</td>
<td>8.352</td>
<td>7.083</td>
<td>12.137</td>
<td>19.836</td>
</tr>
<tr>
<td>17</td>
<td>2LTD</td>
<td>27.09</td>
<td>9.042</td>
<td>6.853</td>
<td>12.552</td>
<td>17.744</td>
</tr>
<tr>
<td>18</td>
<td>444LTD</td>
<td>18</td>
<td>7.858</td>
<td>8.238</td>
<td>12.925</td>
<td>13.86</td>
</tr>
<tr>
<td>19</td>
<td>AS3</td>
<td>21</td>
<td>10.01</td>
<td>9.416</td>
<td>14.769</td>
<td>31.845</td>
</tr>
<tr>
<td>20</td>
<td>305</td>
<td>8.161</td>
<td>8.89</td>
<td>13.701</td>
<td>17.858</td>
<td>23.041</td>
</tr>
</tbody>
</table>

Figure 4 shows the comparison of average speed of the vehicle with the average error obtained from the EKF algorithm. From the bar chart, authors infer that when the average speed was low the accuracy was low. It is also observed from the chart that vehicle with speed more than 20kmph has accuracy less than 10m with exceptions in the route numbers AS3 and 305.
An attempt to find the correlation between average speed and average error variables was done and following observations were made. The following correlation plot in Figure 5 shows the variation in average error and average speed. The correlation coefficient obtained from the data analysis was found to be -0.495 which shows the average error in final location data is negatively correlated with the speed of the bus.

![Correlation Plot](image)

**Figure 5:** Correlation plot of Average error with Average speed of the vehicle

From the above results one can say that the two-stage EKF algorithm performs well in the places where the vehicle speed is more and continuous GPS track is available in at least one of GPS enabled mobile phone. It does not perform well when the vehicle is moving very slowly. Figure 6 shows the percentage of GPS signals that was available in three smartphones which was used for data collection while travelling in each of the routes. From the graph one can see that there were poor signals while travelling on the routes AS1, 496 and 86. In all other routes at least one of the phones had 90% of the GPS signal availability from the start of GPS logging on the phone.

![Availability of GPS signals in different routes](image)

**Figure 6:** Percentage of GPS signals in all routes
For every second the available location data from passenger’s mobile data was processed using the Matlab code developed for two stage EKF algorithm. The resulting final location was plotted on the map along with the reference data to check the performance of the code. The performance of the filter at different geographic locations is shown in the Figure 7.

![Figure 7: Performance of the two stage EKF filter in Mumbai City](image)

6. **Summary and Conclusion**

This paper talks about the implementation of the two stage extended kalman filter to help tracking a bus using the crowd sourcing of location data from the passengers of the bus. Twenty bus routes were chosen for this study which covered 341.45 KM stretch road in and around Mumbai city covering major different geographic location. A matlab code was developed to filter the multiple GPS data collected from the passengers. The data collected was then processed using the filter developed thereby improving the accuracy and tracking bus every second of the commute. The processed data was compared with the data from high accuracy GPS devices like TRIMBLE GeoXT and Trimble Juno to assess the accuracy of the developed filter. Data analysis was done to interpret the results obtained from the application of filter to the smartphone GPS location data.

From the results one can conclude that the filter performed well at places where the GPS signals were available continuously without any obstructions like bridges, curves in the mountainous area, clover leaf junctions, highways and expressways. One of the advantages of considering two-stage EKF is that it showed continuous location data.
from the start time to end time of the travel which was not seen on individual phones. The average error for all bus routes varied between 9.21m to 12.22m (95% confidence interval). At some places the error increased because of various reasons such as roads with dense trees, when the vehicle was moving slowly near bus stops and traffic signals etc. At highly built-up areas, roundabout, Y junction and complex intersections the algorithm was not consistent as some of the points were shown off roads in some routes. Comparison of average speed and average error showed that the accuracy of final estimates depends on the speed of the vehicle.

Major limitations of the above study are at least one GPS data is needed at any point of time for the algorithm to work since the predictive model is not accurate. The final estimate depends only on the GPS data. Hence development of a working predictive model dynamics, which can work independently, may improve the results. Future scope of the study includes: (a) Determining a technique to identify the optimal system covariance values for each state in the predictive model and also the measurement covariance values for the measurement data; (b) Position estimation of the vehicle by using two-stage EKF can be stopped when the vehicle is at rest and the previous estimate can be taken as the current estimate to improve results near bus stops and traffic signals; (c) Enhancing predictive or system model, which can predict the vehicle location by using additional information like PDOP, GDOP and number of satellites; (d) Particle filter can be implemented for this application to check the feasibility for real-time tracking of vehicle and comparing the results obtained using both EKF and particle filter; and (e) Exploring a way to give GIS data as measurement data thereby assuming that vehicle is always on road, which may improve the results.

References


