Automated lane identification using Precise Point Positioning
An affordable and accurate GPS technique

Victor L. Knoop Peter J. Buist Christiaan C.J.M. Tiberius Bart van Arem

Abstract—Nowadays, many vehicles are equipped with GPS navigation systems, that are accurate to approximately 10 meters. This is insufficient to determine the lane a vehicle is driving in. We introduce a new technique, Precise Point Positioning, which is able to get the accuracy of measurement down to approximately half a meter, without having to resort to expensive high-end GPS receivers. This accuracy is possible with a single measurement of position in real-time. We confirm this accuracy in a real-life test with a vehicle driving at motorway speeds. However, even when the vehicle position is known, the driving lane is not known since there are generally no maps with detailed lane information. Therefore, this paper also presents a self-learning method, using this Precise Point Positioning information, to create maps which include the position of lanes. This method is tested using an artificially created dataset, using the accuracies from the real-world test. This shows that the method can get the position of a lane at an accuracy of 20 cm. The combination of accurate information of the position of a vehicle and information about the position of the lane, can be used to give lane-specific advice for drivers, and can even be a step towards automated driving.

I. INTRODUCTION

Nowadays, positioning through GPS (Global Positioning System) is widely used in the automotive industry. Currently, its main use in vehicles lies in routing applications. There is a wider potential of GPS if the position could be determined more accurately. There is a market need for driver assistance systems such as lane change advice and assist that explicitly use knowledge of the lane the vehicle is driving on. For instance, drivers could be informed about fast and slow lanes, or drivers could even be instructed to perform a lane change to prevent the occurrence of congestion – see [1] for a literature overview of these type of measures. For automated driving, the driving lane is even more important.

We developed a new technique, Precise Point Position (PPP, see the explanation in section III) which is able to provide positions accurate to a decimeter level at a low cost. This can be used to determine the lane a vehicle is driving in, provided that the locations of the lanes are known with sufficient accuracy from a map or in a database. At the moment, this is generally not the case. Interestingly, PPP can also be used to determine the locations of the lanes.

The goal of this paper is twofold: (1) showing the basic principles of the newly developed PPP-GPS technique and (2) showing a self-learning method identifying the positions of the lanes. It does so by using PPP-GPS data from probe vehicles on the motorway. The lane positions could also be determined with very accurate Differential-GPS techniques (D-GPS, see section II-B), but that is too costly. We therefore focus on our newly developed PPP-GPS technique which can be much cheaper, to see whether the position of the lanes can be determined accurately enough to produce suitable a map.

The paper first will describe the existing GPS measuring techniques, regular GPS and Differential GPS, in section II. Next, we explain the basics of our new technique, PPP-GPS, in section III, including a real-world experiment at the motorway to demonstrate the achievable accuracy. Section IV shows the method to map the position of the lanes using measurements which are as accurate as can be expected from PPP-GPS measurements. This method is tested using simulation, of which the set-up is described in section V and the results are shown in section VI. The paper closes with conclusions (section VII).

II. EXISTING GPS TECHNIQUES

This section gives a simplified overview of GPS and Differential-GPS (D-GPS) and the accuracies which can be obtained with each of them. For a full overview, see [2].

A. Regular, stand-alone GPS

In regular GPS data, the position is determined using at least 4 satellites, each continuously transmitting time signals. The position of the receiver is found by measuring travel times to these satellites, and solving for the three dimensional receiver position and its clock offset. The position of the satellites and the satellites clock offset have to be known as well. The US Air Force calculates those, and the predicted trajectories of the satellites are sent with the signal of the satellite. Typically, there is an error of several meters in this predicted trajectory, leading to a similar error in the position estimate.

A second source of errors in the GPS measurement is the disturbance of the signal in the atmosphere. This can cause variable time delays in the signal traveling to the vehicle. The errors as a result of this delay are larger in the vertical direction than in the horizontal direction. Note that the receiver cannot get signals from satellites which are under the horizon, so satellites are usually situated above the vehicle. A slight error will therefore immediately influence the estimated height, but only modestly change the perceived horizontal position.
All combined, the horizontal position of regular GPS positioning is accurate to approximately 5-10 m, and the vertical precision 10-20 m.

B. Differential GPS

Here, we describe Differential GPS, or D-GPS in its simplest form. More advanced techniques are applied in practice, but the basic idea is the same as explained here. It uses the same GPS signal as in regular GPS. However, the mobile device (i.e., vehicle) is also in contact with a base station nearby. The exact location of this base station is known, but the location is also determined using a GPS receiver. At this base station, the difference between the GPS position and the known position is determined.

This difference is due to atmospheric conditions and due the errors in satellite position and clock. These error sources have a similar impact for all GPS receivers nearby. The error in the determined position of the base station is communicated to the mobile device. The mobile device can then compensate for the errors, assuming these are the same as at the base station. The final position accuracy depends on the distance to the GPS base station. In the end, the measurement accuracy is the limiting factor in the eventual position accuracy. This can be down to centimeter-level (using the so called carrier phase cycle ambiguity fixed solution) [2]. For this level of accuracy, high-end, and hence costly, equipment is needed.

III. PRECISE POINT POSITIONING

A. Technique

We developed a new technique, Precise Point Positioning (PPP) as an intermediate technique between stand-alone GPS and D-GPS, but at a cost comparable with the stand-alone GPS. Technical details can be found in [3]. In this paper we present the basic idea, which is as follows. There are several hundreds permanently operating GPS base stations over the world, for which high-accurate positions (<1 cm) are available. At these locations, the GPS signal is measured. Based on these measurements, and orbital mechanics, satellite positions are predicted, as well as atmospheric delay model parameters. This prediction is made for several hours (up to one day) ahead for the whole world. This prediction is essential for real-time positioning. Satellites clock error estimates are available in real-time. All this information is made publicly available on the internet. With this information, the GPS measurements can be corrected. A resulting position error is typically in the order of several decimeters.

In the limit of the vehicle being very close to the base and the prediction of the model being updated continuously, this gives back the D-GPS method. The Wide Area Augmentation System (WAAS) and the European Geostationary Navigation Overlay Service (EGNOS) are other systems to improve GPS accuracy. Errors in the GPS measurement at base stations throughout a continent are shared, thus approaching a D-GPS measurement. However, contrary to PPP, the different sources of errors are not fully separated. Field test show that errors of EGNOS are in the order of meters, e.g.[4].

B. Experiment of position accuracy in motion

Because the accuracy of PPP-GPS in motion might be different from the static accuracy, we carried out a test using a vehicle at speeds up to 100 km/h. A more detailed description of the test can be found in [3]. In particular, we were interested in the quality of a PPP-solution. We tested this in practice in a real-life set-up. We choose here a simple mu-blox TIM LP GPS receiver (cost: tens of euros), connected to a patch antenna (Tri-M Big Brother). This is believed to have the same accuracy as a mobile GPS chip set in modern vehicles. It hence is affordable, and one can assume there will be devices that can determine their position this accurately.

The receiver for which the PPP-GPS signal was retrieved, the mu-blox, was installed exactly in between two high-end geodetic receivers, of which measurements have been processed in precise carrier phase based D-GPS (see figure 1). This set-up allows reconstruction of accurate (ground) truth positions for the mu-blox located exactly in between them at the accuracy of the carrier phase based D-GPS. This ground truth is at least one order of magnitude more accurate than PPP-GPS.

We tested the system on the A13 multi-lane motorway in the Netherlands on a moving van (see figure 1). This is very close (approximately 5 km) to a nearby base station, so therefore also a very accurate D-GPS signal was available.
TABLE I: PPP results expressed in local North, East, and Up components. Given are mean, standard deviation (std) around the mean of the differences between the position estimate from the PPP technique and the very accurate D-GPS technique in m.

<table>
<thead>
<tr>
<th>session (nr. of points)</th>
<th>mean [m]</th>
<th>std [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>E</td>
</tr>
<tr>
<td>session 1 (1835)</td>
<td>0.25</td>
<td>−0.79</td>
</tr>
<tr>
<td>session 2 (1795)</td>
<td>−0.82</td>
<td>0.30</td>
</tr>
<tr>
<td>session 3 (1842)</td>
<td>0.60</td>
<td>0.00</td>
</tr>
<tr>
<td>session 4 (2006)</td>
<td>0.38</td>
<td>−0.20</td>
</tr>
<tr>
<td>Value used in simulation</td>
<td>Std of N&amp;E:</td>
<td>Mean of N&amp;E:</td>
</tr>
<tr>
<td></td>
<td>0.53</td>
<td>0.33</td>
</tr>
</tbody>
</table>

which was used as ground truth in this experiment. Four sessions were carried out at different times of the day, because the error in the atmosphere varies over the day. In each of these sessions, a round trip of approximately 10 km was made 5 times, referred to as 5 runs.

C. Measurement accuracy in field experiment

Figure 2(a) shows the trajectories as recovered by the PPP technique in Google Earth. Different colours show different sessions. In all runs in all sessions, the vehicle was in approximately the same lateral position. Figure 2(b) shows for one run the ground truth (cm precision D-GPS) trajectory. The average of the PPP trajectories is at approximately the same lateral position as this ground truth, suggesting there is no structural offset. The lines are not plotted in the middle of the lane because it is not in the middle of the vehicle and the images of Google Earth may be slightly offset (at the meter-level). Comparing the average deviation in the lateral position with the 3.50 m lane width, we conclude that the spread is typically 50 cm to each side (a lane is 3.50 m wide).

The errors in position of the PPP-GPS measurement compared to the cm-accurate D-GPS measurement are analyzed for all sessions. We differentiate between systematic and random errors (see also section IV-A). Systematic errors show auto-correlation at an hour time scale, and hence are the same for the different runs in one session. The mean quantifies the systematic error in that run, which is most likely due to the disturbances in the atmosphere and the trajectories of the satellites.

The standard deviation shows the variability of the measurements compared to the D-GPS measurements modified with the systematic error. This is therefore the random error, which is 33 cm is the field experiment. It is checked that there is no systematic part in the random noise, i.e. if averaged over a large number of measurements, it will reduce to zero. Table I summarizes the sizes of the systematic and random errors for all sessions.

IV. METHOD TO RECOVER LANEs

In this section, we describe a method to recover the position of lanes in a multi-lane motorway. First, it will be shown which types of errors can be expected in the measurements, and then section IV-B will show how these noisy measurements are used to find the position of the lanes.

A. Errors in the measurement

In the remainder of this paper, we will use the PPP-GPS measurements. This still contains a measurement error, as shown in the previous section. Furthermore, there are other sources of error, limiting the ability to determine the position of the lanes. The list below describes all of these sources, and gives estimates for their magnitudes.

- **Systematic error term.** This error source of the PPP-GPS measurement is mainly the atmosphere and the satellite orbit prediction. So, whereas this error is systematic for time periods up to several hours, it is a random error over data collection periods which are longer [5]. Over these independent measurements, the error will have zero mean, since there is no systematic long-term bias in the atmosphere or the satellite position prediction. The error can be described by a normal distribution with a mean of 0. To find the spread of the distribution, we use the results of the experiment (see table I). The systematic error term cancels for various measurements within a session. Within one session, the mean error is the systematic error. We calculate the spread of the mean errors over all sessions: 0.53 m. We hence model the systematic error as normal distribution with a mean of 0 and a standard deviation of 0.53 m.

- **Random error term.** This is the random variation in the estimated position of the PPP-GPS technique. It is a normally distributed error with the mean zero and the spread is estimated from the measurements. This random error term shows within a session, hence estimate the spread for the model from the average standard deviation within each session. This is 0.33 m, so the random error term is modeled by a normal distribution function with mean 0 and standard deviation of 0.33 m.

- **Lane change.** Leading to a position in-between lanes. A lane change duration of 6 seconds [6] and 0.5 lane changes per km [7] gives on average a lane change duration of 3 seconds per km. This is compared to the travel time for one kilometer. Assuming free flow conditions (to get the highest, and therefore least favorable lane change rate), the speed is 30 m/s, and the travel time 1000/30=33 s. 3/(33-3)=10% of the vehicles is therefore assumed to be somewhere in the process of lane changing. The vehicle is therefore assumed to be anywhere in between the two lanes, with a uniformly distributed uncertainty of 3.5 m.

- **Vehicle width:** the receiver can be anywhere in the vehicle. We assume the position of the receiver in the vehicle, within an assumed vehicle width of 1.80 m. If needed, one could apply correction for this.

- **Position of the vehicle in the lane.** The position of the center of a vehicle within a lane is assumed to follow a normal distribution, centered around the middle of the lane. The lane width is 3.50 m, and the vehicle is 1.80 m.
m wide. Vehicles have therefore 85 cm at each side as a margin. We assume they will generally stay within this margin, and in fact stay within and keep a margin of 40 cm at each side for 95% of the time, based on [8]. Therefore, we assume that the position of the center of the vehicle follows a normal distribution with a standard deviation of 20 cm to each side. For individual vehicles, this will be correlated with time.

All error sources and their assumed distributions are summarized in table II.

### B. Fitting the distribution of lateral passings

The goal is to find the position of the lanes from PPP-GPS traces. The general idea is to check the lateral position of the vehicles passing at a cross section. This ideally would show the exact location of each of the lanes. However, since there is an error on the position, the lateral positions of the vehicles do not exactly match the lateral position of the middle of the lanes. Our method consists of determining a probability density function describing the lateral position of the passing vehicles, represented by \( f(x) \) in which \( x \) is the lateral position on the road. This function depends on several parameters, including the position of the lanes. By optimizing the parameters (i.e., we minimize the difference between \( f(x) \) and the observed distribution of lateral passages, we find the position of the lanes.

The probability density function of the positions of all vehicles is an weighted average of the distribution of the probability density functions of the positions of the vehicles in each of the lanes \( f_l(x) \), weighted with the number of vehicles per lane \( (Q_l) \). Alternatively, we can weight the distributions by the fraction of the flow in each lane \( \eta_l \), i.e. we define \( \eta_l := Q_l/Q \), in which \( Q \) is the total number of vehicles. Note that \( \sum_{l \in \text{lanes}} \eta_l = 1 \). The distribution of the lateral position can now be expressed as follows:

\[
    f(x) = \sum_{l \in \text{lanes}} \eta_l f_l(x) \tag{1}
\]

The magnitude of all errors listed in section IV-A may be estimated given enough data, but that would require several extra parameters of the error terms to be estimated from the data. Moreover, the error terms are not our main interest, but the position of the road is. For simplicity in the procedure, and hence to get a more reliable estimate, we combine all errors in a single error term. We therefore simplify each of the distributions per lane to a normal distribution. In this paper, we will use the notation \( N(x|\mu; \sigma) \) for the normal probability density function with a mean \( \mu \) and a standard deviation \( \sigma \). So, we have for each of the lanes, by assumption,

\[
    f_l(x) = N(x|\mu_l; \sigma_l) \tag{2}
\]

Figure 3(a) shows that a normal distribution fits the simulated data including all errors first reasonably. The expectation value for a lateral position in lane \( l \) is in the middle of the lane \( (\bar{x}_l) \), and the standard deviation of the position is assumed to be unknown \((\sigma)\), but the same for all lanes. So in an equation, this reads:

\[
    f_l(x) = N(x|\bar{x}_l; \sigma) \tag{3}
\]

Note furthermore that once the position of the right end of the right lane of the roadway is determined \( (z) \), the positions of the middle of the lanes can be derived based on the lane width \( (w) \), which is given in the motorway handbook. \( \bar{x}_l = lw - z \) Substituting this and equation 3 into 1 gives:

\[
    f(x) = \sum_{l \in \text{lanes}} (\eta_l N(x|lw - z; \sigma)) \tag{4}
\]

The unknowns in this equation are the number of lanes \( (n_l) \), the lane distribution \( \eta_l \) \((n-1)\) degrees of freedom, since they have to add to one), the offset \( z \) and the spread of the lateral positions \( (\sigma) \).

The parameters values are found by fitting function 4 to the empirically observed distribution using \textit{fnminsearch} in \textit{Matlab}. As a goodness-of-fit the Kolmogorov-Smirnov distance [9] is used, i.e. the distance between the empirically observed distribution and the fitted distribution.

### V. TEST OF THE METHOD USING SIMULATION

To test the method, a simulation is set up which mimics the errors of the position of the vehicles at a cross section. First, it is shown how this artificial data set is created. Then, section VI shows the results of applying the method of finding lanes to this data set.

#### A. Simulation set-up

We artificially create the lateral position of passages of vehicles at a cross section, and introduce errors in that lateral position. In this data creation, the found errors are mimicked as closely as possible, including all of the errors explained in section IV-A; table II summarizes the effect of the errors. These are implemented in a simulation creating an artificial data set. The remainder of the section explains the setup of the simulation. The settings are summarized in table III.

First, the lane of a vehicle is set based on a predefined lane distribution (40-40-20), which reasonably fits the pattern of the passages over a day [10]. Then randomly 10% of the vehicles is marked to perform a lane change. For these vehicles, an adjacent lane is chosen to change into. Randomly, a position between the two lanes (origin lane and destination lane) is taken from a uniform distribution. Furthermore, all vehicles are assumed to not drive very accurately at the middle of the lane, but to have a distribution to their lateral

### TABLE II: Overview of the different sources of errors in the simulated GPS measurement, their assumed distribution and magnitude

<table>
<thead>
<tr>
<th>name</th>
<th>type</th>
<th>width</th>
</tr>
</thead>
<tbody>
<tr>
<td>systematic GPS</td>
<td>normal</td>
<td>0.53m</td>
</tr>
<tr>
<td>random GPS</td>
<td>normal</td>
<td>0.33m</td>
</tr>
<tr>
<td>position in veh</td>
<td>uniform</td>
<td>1.8m</td>
</tr>
<tr>
<td>veh pos in lane</td>
<td>normal</td>
<td>0.2m</td>
</tr>
<tr>
<td>lane change rate</td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>position in lane change</td>
<td>uniform</td>
<td>3.5m</td>
</tr>
</tbody>
</table>
position described by a normal distribution with a standard deviation of 20 cm. Then, the position of the receiver in the vehicle (assumed uniformly distributed over the width of the vehicle) is added to the lateral position. To this assumed position, a measurement error is added. This consists of systematic noise and random noise, each of which is assumed to be normally distributed.

Let us first consider the typical patterns in case the number of observations is large, 10,000. This is a large number of observations to collect with but on an average motorway, the number of passages is more than 10,000 per day, so if users GPS-traces are collected by a company or an open maps initiative, this number is feasible. Using these data created this way, the lateral position function from equation 4 will be fitted. The number of lanes, however, is supposed to be unknown and should be found. To recover the number of lanes, we estimate the parameters of the distribution function in equation 4 for various numbers of lanes: 2-7. The number of lanes which gives the best fit will be considered the actual configuration of the road. The corresponding value for the position of the right lane is then considered to be the position of the rightmost lane.

To test the reliability of the fits, the set is generated 20 times, and the parameters are fitted 20 times. This way, it can be analyzed to which extent the value of the parameters depend on the stochasticity of the data. Since stochasticity will probably play a role, the number of observations will matter. We test the method for 50, 100, 1,000 and 10,000 simulated passages.

VI. DATA PROCESSING AND RESULTS

Figure 3(a) shows the histogram of the lateral position of the passings from the simulation for 100,000 passings. The three driving lanes are clearly visible Figure 3(a) also shows the best fits of the distributions for 2, 3, and 4 lanes are shown. The best fit for 2 lanes is moved towards the middle of the 3 lanes, and clearly does not fit the observed distribution well. The 3-lane estimate can fit the distribution quite well. This was not obvious on beforehand, since not all errors are explicitly modeled in the fit function, since they are all combined in one error term.

The fit worsens if 4 lanes are assumed (figure 3(a)). This is also show in figure 3(b), where the error increases. A fit with no vehicles in the fourth lane, $\eta_4 = 0$, would give the same result as a road of 3 lanes. However, the optimization requires that for each lane the number of vehicles is larger than zero: $\eta_i > 0$

The best fit is achieved with a 3 lane road. The parameters and the spread in those parameters, the latter obtained by iteratively creating a dataset, are shown in table IV. For our goal finding the lateral position of the lanes, only the offset is interesting. This is on average 20 cm off the generated ground truth, with an standard deviation (over the several generated sets) of 12 cm. This is an order of magnitude more accurate than the lane width, and thus sufficiently accurate.

The above estimation procedure works well with enough data. It is tested how much data in fact is needed for the estimation. To this end, we repeat the procedure with less passages. Figures 3(c) to 3(e) show the results of the estimation for 1000, 100, and 50 observations. With 1000 measurements, the shape is still recognizable, and the offset is in the order of two decimeters. Stretching the method to 50 measurements will not give very reliable results - see figure 3(e), and the fit is not reliable. Although the individual fits are not very reliable, the fits still find the lane position with an accuracy of less then 25 cm.(figure 3(f)).

Let us finally compare this with our field experiment. In the section of the field experiment which is shown in (figure 2(a)), the vehicle was always in the same lane. The figure shows that in that case 20 observations already give a good estimate of the lateral position of the GPS receiver in the lane (the ground truth, figure 2(b)). The mean of these 20 observations offset with the position of the receiver in the lane, comes very close to the actual lane position.

VII. DISCUSSION AND CONCLUSIONS

In the paper we presented our new technique, Precise Point Positioning, which gives an accurate position of a PPP-GPS device at a cost of a regular, stand-alone GPS. The field experiment showed that the position accuracy of PPP-GPS is good and lies in the order of several decimeters. These results pertain to real-time, single measurements of positions of a vehicle on the road. It can be expected that vehicles with communication equipment will use Precise Point Positioning in the near future. Precise Point Positioning is a cheap but powerful technique that can be used on large scales. The technique can be used to identify the position of a vehicle an order of magnitude more accurate than the width of a lane. Applications in automated vehicles and lane advice will however only be interesting once the position of the lanes is known this accurately as well.

To get a database or map of all lane positions, we presented a method to create a self-learning map, using the positions collected by the Precise Point Positioning technique as input. In the paper we were very conservative with the assumptions: it is assumed that vehicles send a position based on an instantaneous measurement, i.e., there is no filtering over time of the random error. In practice, this component
can be easily filtered out in the vehicle. Furthermore, the lane change maneuvers can be filtered out, since a PPP-equipped vehicle can detect the movement orthogonal to the axis of the road (known from the map). These improvements are not made to the simulated raw data.

Even then, the proposed method succeeded to reconstruct the position of the lanes within approximately 20 cm, which is sufficient to determine a lane, since a lane is 15 times wider than this uncertainty. This also holds for lower number of measurements. Up to 100 measurements the obtained results are similar: a 20 cm accuracy. For even lower number of observations, the fitting does not give reliable results. Note however, that modern vehicles are able to communicate with internet (required for the PPP-GPS technology), and hence all PPP-GPS they can upload their positions to a database. Even though some drivers may opt out due to privacy reasons, for busy motorways 100 measurements are certainly possible given traffic loads of 10,000 per day or more.

For roads with narrower lanes, the absolute accuracy will be better than the 20 cm found for wide lanes since the vehicles will deviate less from the center of the lane, and the influence of lane changing will be less. Since a lane should be at least 2 m wide to accommodate vehicles, the procedure will also work for these lanes.

**References**


