

**Development of a Micro-Mobility Analysis System
Using Precise GPS Traces**

by

Shubhodeep Mukherji

THESIS

Presented to the Faculty of the Undergraduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

**BACHELOR OF SCIENCE
IN
AEROSPACE ENGINEERING**

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2014

**Development of a Micro-Mobility Analysis System
Using Precise GPS Traces**

APPROVED BY

SUPERVISING COMMITTEE:

Todd Humphreys, Supervisor

Glenn Lightsey, Supervisor

Acknowledgments

I would like to thank Dr. Humphreys and Dr. Lightsey for their support and guidance throughout this project. It was a great experience working with both of you, and I definitely learned a lot in the process. I would also like to thank Ken Pesyna for his initial work on mobility analysis and for his work on the alignment routine. Furthermore, I would like to thank all members of the Radionavigation Laboratory for the valuable discussions. Lastly, I would like to thank Cody Colley and Kellen Wall for their assistance in collecting data and for their constant encouragement throughout this year.

Development of a Micro-Mobility Analysis System Using Precise GPS Traces

Shubhodeep Mukherji, B.S. ASE
The University of Texas at Austin, 2014

Supervisors: Todd Humphreys
Glenn Lightsey

A Micro-Mobility Analysis System (μ MAS) is developed to characterize drivers based on their driving habits. This is achieved by solving a pattern recognition problem, which can be divided into four phases: data collection, data pre-processing, parameter extraction, and classification. These phases are discussed in detail and the habits used to distinguish between drivers are various parameters associated with their general characteristics, turning behavior, and lane change mentality. A cross-validation simulation is implemented to gauge the performance of μ MAS. The results indicate that the system was successful in identifying one of the three drivers, but not the other two. The various factors contributing to this performance are discussed. The techniques developed in this study can be used to measure distance between drivers and place them into clusters, which can then be used to assess whether they drive in a safe or unsafe manner.

Table of Contents

Acknowledgments	iii
Abstract	iv
List of Tables	vii
List of Figures	viii
Chapter 1. Introduction	1
Chapter 2. Precise Point Positioning	4
2.1 Implementation	4
2.1.1 Precise Ephemeris	5
2.2 Results	9
Chapter 3. Mobility Analysis Approach	12
3.1 Data Collection	12
3.2 Data Pre-Processing	14
3.2.1 Alignment	14
3.3 Parameter Extraction	16
3.3.1 General Characteristics	17
3.3.2 Turning Behavior	18
3.3.3 Lane Mentality	18
3.3.4 Post-Extraction Process	19
3.4 Classification Routine	19

Chapter 4. Parameter Extraction Methods	22
4.1 General Parameters	22
4.2 Turning Behavior	22
4.2.1 Identifying Turn Segments	23
4.2.2 Radius of Curvature	26
4.2.3 Lateral Acceleration During Turns	27
4.3 Lane Mentality	27
4.3.1 Determining Lane Centers	28
4.3.2 Parameters	29
Chapter 5. Results	31
5.1 Cross-Validation Simulation	31
5.2 Future Work	33
Chapter 6. Conclusion	35
Bibliography	37

List of Tables

2.1	Average error compared with CSRS-PPP for each method . . .	10
5.1	Performance of μ MAS	32

List of Figures

2.1	Satellite position error between precise and broadcast ephemeris	6
2.2	Vertical TEC map at a given epoch	8
2.3	Geometry of the Single Layer Model	9
2.4	Horizontal offset from true position for each method	10
2.5	Vertical offset from true position for each method	11
3.1	Data collection setup	13
3.2	Effects of applying a moving-average filter on a velocity profile	15
3.3	Effects of applying the position alignment routine	16
4.1	Speed profile with candidate turn segment locations	24
4.2	Angular profile with beginning and end of turn segment identified	25
4.3	Correct identification of a turn segment during a trace	25
4.4	Location of lane centers	30

Chapter 1

Introduction

Quantitatively analyzing humans can be an arduous task, but techniques have been developed to characterize and even predict human mobility [1]. In their study, Sadilek and Krumm were successful in predicting the most likely location for an individual at some specified time in the future. Mobility analysis can also take form in characterizing the driving styles of individuals [2], which would be particularly beneficial to companies operating fleets of vehicles, such as those involved in public transit systems.

While travel time and reliability are of primary importance to riders in public transit systems, safety and comfort are also factors that need to be considered. Even if the transit system is fully optimized for travel time and reliability, the ridership base can be increased by offering safe and comfortable rides. However, due to the low frequency of reported traffic incidents, it would require a long time to develop safety profiles for drivers, thus making it difficult to quantitatively characterize their performance. Furthermore, fleets of vehicles are generally equipped with standard-precision GPS receivers, but a large volume of information is lost due to the sparse temporal resolution and imprecise spatial resolution due to the nature of standard GNSS tracking

systems. For instance, these systems are capable enough to determine the road and heading, but cannot distinguish between lanes. Much of this lost information can be recovered with precise GNSS, which offers decimeter-level accuracy and can be used to distinguish between lanes.

Development of the quantitative tools necessary to construct safety profiles of drivers would allow for the comparison of drivers relative to one another based on their distance from each other. The distance between drivers represents the norm between vectors describing the location of the drivers in a vector space. The coordinates of drivers in this vector space are associated with their driving microbehavior. In order to develop these tools and determine the distance between drivers, a Micro-Mobility Analysis System (μ MAS) was designed to address the related and interesting issue of whether a driver can be identified based on his or her driving habits. It was designed to collect large amounts of driving traces from multiple drivers and construct a database describing the typical driving behaviors of the subjects by extracting certain parameters of interest. If a driving trace of an unknown driver is then obtained, μ MAS can be used to extract the same parameters and compare them to the database in order to determine the identity of the mystery driver.

The setup of μ MAS is similar to that of biometric identification systems, which utilize an individual's unique physiological and behavioral traits such as the face, fingerprint, retina, voice, and electrocardiogram [3]. These systems are commonly used for systems requiring authentication since systems based passwords and secret codes can be breached with a certain amount of effort.

The architectures of these systems are similar since they all attempt to solve a pattern recognition problem.

The issues regarding resolution were accounted for by incorporating the post-processing technique of Precise Point Positioning, which is discussed in Chapter 2. The specific approach to the system design is discussed in Chapter 3, while the methods for extracting the parameters of interest are discussed in Chapter 4. Finally, the performance of μ MAS is analyzed in Chapter 5.

Chapter 2

Precise Point Positioning

In 2000, the United States government ended the practice of Selective Availability (SA), which was an intentional degradation of civil GPS signals that led to position errors of up to 100 meters [4]. Removal of SA reduced errors to within 10 meters and this error can be further reduced by accounting for the sources of these errors. The major causes for errors in the position solution are due to ionospheric delay, tropospheric delay, satellite clock, satellite orbit, receiver noise, and multipath. Precise Point Positioning (PPP) can be used to obtain decimeter-level accurate position solutions by incorporating data products made available by the International GNSS Service (IGS). In this study, a single-frequency pseudorange-based position solution algorithm was augmented with precise ephemeris and ionospheric maps from IGS. Implementation of PPP was used to achieve the lane-level resolution desired for this study.

2.1 Implementation

The pseudorange between a receiver and satellite can be modeled by

$$\rho = \|\mathbf{r}_S - \mathbf{r}_R\| + c(\delta t_R - \delta t_S) + I_\rho + T + \omega_\rho, \quad (2.1)$$

where I_ρ is ionospheric delay, T is tropospheric delay, ω_ρ is noise, and \mathbf{r} and δt are the position and clock offset, respectively, of either the satellite or receiver. In a pseudorange-based position solution, if at least four satellites are tracked, and the pseudorange between each satellite and the receiver is known, then a system of equations can be constructed. The nonlinear least squares estimation method can then be used to solve for \mathbf{r}_R and δt_R .

2.1.1 Precise Ephemeris

The precise ephemeris from IGS contains the satellite orbit and clock information. The sample interval for the data is 15 minutes and interpolation can be used to determine the satellite position and clock state at the desired time. The satellite positions given are accurate to within 2.5 *cm* and are be considered to be the true positions in this study. The difference between the satellite positions obtained from precise and broadcast ephemeris, which is used in Standard Positioning Service (SPS), are depicted in Figure 2.1 for a two hour period. The broadcast ephemeris is updated every two hours, and it can be observed that during this period, the error in the satellite positions increase. Since Equation 2.1 depends on the satellite position, errors in \mathbf{r}_S can significantly affect ρ and therefore, the position solution as well.

The ionosphere is essentially a shell of electrons and electrically charged molecules, and exists primarily due to ultraviolet radiation from the Sun. The interaction between these electrons and electromagnetic signals, such as GPS signals, affects the signal's propagation path through the ionosphere, which in

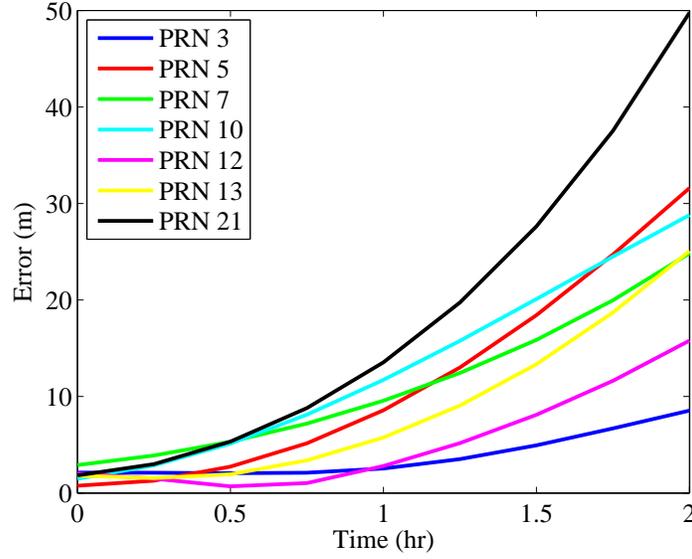


Figure 2.1: Satellite position error between precise and broadcast ephemeris turn causes an error in the measured time of flight. Therefore, accounting for the ionosphere yields a better measurement for the time of flight, thus leading to an improved position solution. The delay is dependent on the total electron content (TEC), which characterizes the electron density in the signal's path. TEC is expressed in terms of TEC Units (*TECU*), where 1 *TECU* corresponds to an electron concentration of $10^{16} e^-/m^2$. For the GPS L1 frequency, f_{L1} of 1575.42 *MHz*, the ionospheric delay experienced is

$$I_{\rho} = \frac{(40.308 \times 10^{16}) TEC}{f_{L1}^2}. \quad (2.2)$$

So, 1 *TECU* represents 16.24 *cm* of delay.

IGS provides a TEC map of the Earth every four hours. In each map,

the Earth is divided into a 71 (latitude) by 73 (longitude) grid and the TEC above each of the grid points is given. Since the data represents TEC directly above a point on Earth, it represents the electron density a signal would face if it was traveling vertically, and is therefore referred to as vertical TEC (VTEC). These maps are low resolution, but with interpolation, high resolution maps can be obtained. A high resolution VTEC map of the Earth taken at a previous time is shown in Figure 2.2. It is unlikely that a signal's path through the ionosphere would be vertical. Thus, it is necessary to properly characterize the electron density faced by GPS signals. This can be achieved by incorporating mapping functions, which transform vertical TEC into slant TEC (STEC). Replacing TEC with STEC in Equation 2.2 can then be used to determine I_p .

The mapping function commonly used is the Single Layer Model (SLM). The ionosphere is composed of multiple layers that affect signals differently, but SLM assumes that the ionosphere is compressed into a thin shell at an altitude of 450 *km*. The point at which the signal crosses this thin shell is referred to as the ionospheric pierce point (IPP). The relationship between vertical and slant TEC is depicted in Figure 2.3 [5]. If the zenith angle at the receiver, z , and VTEC at the receiver are known, then the zenith angle at IPP can be defined as

$$z' = \sin^{-1} \left(\frac{R_e}{R_e + h} \sin z \right). \quad (2.3)$$

Then, STEC can be obtained from

$$STEC = \frac{VTEC}{\cos z'}, \quad (2.4)$$

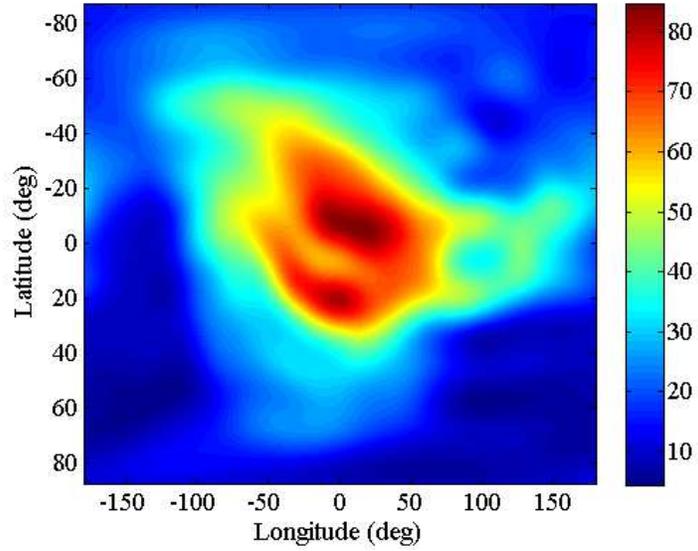


Figure 2.2: Vertical TEC map at a given epoch

As shown above, SLM relates VTEC at the receiver to STEC using the zenith angle at IPP. For low elevation satellites, converting VTEC at the receiver to STEC might not be an accurate representation of the TEC experienced by the signal. A proposed modification to the SLM (modSLM) is to first determine the location of IPP and find VTEC at this location. The procedure for finding I_ρ , would remain the same except for using VTEC at IPP instead of at the receiver. The location of IPP can be found from

$$\mathbf{r}_{IPP} = \mathbf{r}_R + \left(\frac{h}{\cos z} \right) \left(\frac{\mathbf{r}_S - \mathbf{r}_R}{\|\mathbf{r}_S - \mathbf{r}_R\|} \right). \quad (2.5)$$

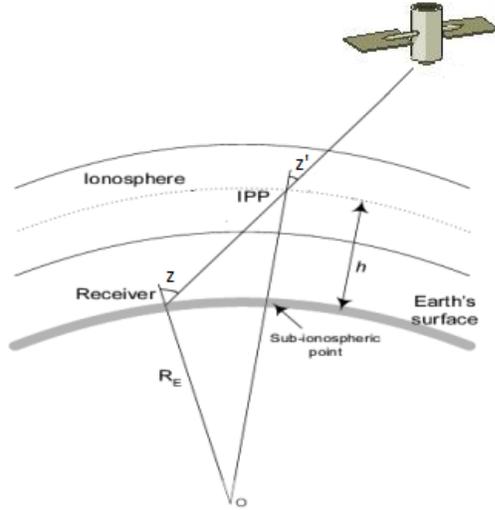


Figure 2.3: Geometry of the Single Layer Model

2.2 Results

Static data from an IGS reference site was used to test the performance of the SPS and PPP algorithms and were compared to post-processed results from National Resources Canada’s online PPP service (CSRS-PPP). CSRS-PPP provides centimeter-level accuracy and for single-frequency recordings, uses both code and carrier phase measurements. It also accounts for satellite antenna offsets, phase wind-up, and the effects of solid earth tide and ocean loading [6]. In this study, the solutions obtained from CSRS-PPP are considered to be the “true” solutions. The horizontal and vertical errors for the SPS, PPP with SLM, and PPP with modSLM methods are shown in Figures 2.4 and 2.5, respectively. The errors of the average position solution for each

method compared to CSRS-PPP are shown in Table 2.1. The data shown was collected with a Javad Delta receiver connected to an Antcom antenna and averaged over 6 hours.

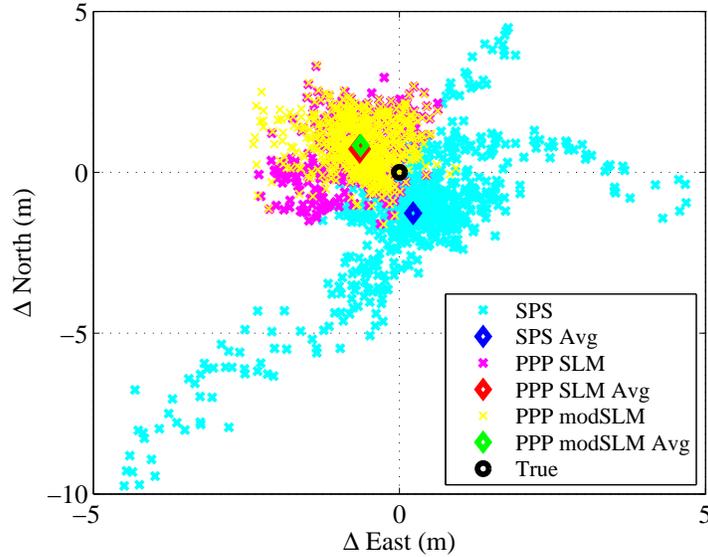


Figure 2.4: Horizontal offset from true position for each method

Table 2.1: Average error compared with CSRS-PPP for each method

Method	Horizontal Error (m)	Vertical Error (m)	Total Error (m)
SPS	1.295	1.561	2.028
PPP with SLM	0.971	0.622	1.153
PPP with modSLM	0.968	0.524	1.101

It can be observed visually that both PPP solutions are more precise than the SPS solution. However, they are still inaccurate compared to CSRS-

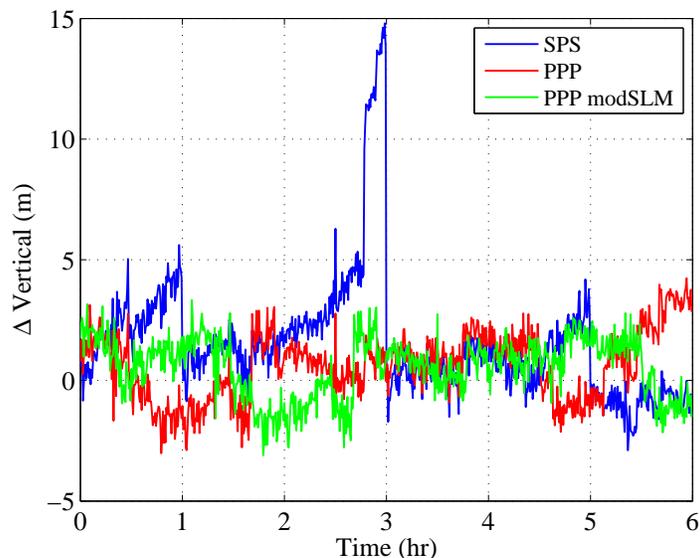


Figure 2.5: Vertical offset from true position for each method

PPP. This is expected since the implementation here was only pseudorange-based and only accounted for errors in the ephemeris and ionospheric delay. Although small, the proposed modification to SLM did reduce the error in the position solution by 5.2 *cm* and increased the precision of the PPP solution based on SLM.

While the PPP algorithm did improve the existing SPS algorithm, it is still not perfect. Instead of implementing all the corrections accounted for by CSRS-PPP, this study utilized their kinematic tool to post-process recorded GPS traces.

Chapter 3

Mobility Analysis Approach

As discussed earlier, μ MAS was setup in a similar manner as biometric identification systems. It was designed to determine the identity of a driver from a test trace by extracting certain parameters from the trace and comparing them to a database consisting of the parameters of various drivers. Since the route affects the driving behaviors of individuals, prior knowledge of the route driven by the mystery driver is required for this system. The parameters from the test trace are then compared with parameters corresponding to the same route. This pattern recognition problem can be divided into the following four phases: data collection, data pre-processing, parameter extraction, and classification. The following sections discuss each of these phases in detail.

3.1 Data Collection

Constructing a database for this pattern recognition problem requires a large volume of data. For μ MAS, recordings of many traces for multiple drivers were desired around various routes to construct an accurate driver profile. However, for the purposes of initial system verification, only sixteen recordings from three drivers were collected around one route in Austin, Texas. This

route was chosen to include residential and highway segments in order to fully characterize the drivers and is shown in Figure 3.1. All traces were recorded with the same vehicle, and used a Javad Delta GPS receiver connected to an Antcom antenna.



Figure 3.1: Data collection setup

3.2 Data Pre-Processing

Once data was collected, it was first processed through CSRS-PPP to obtain the precise traces. A speed and rate of change of speed profile was then developed for each trace. The existing high frequency noise in these profiles were removed by applying a moving-average filter of span 5, which replaced each point in the profile with the average of the 5 neighboring points on both sides. This filter eliminated the effects of thermal and vibration noise, and the result of applying it to a speed profile is shown in Figure 3.2. During the database construction stage, the precise traces were also used to determine the mean trace and velocity profile for each driver, which was subsequently used to determine the mean overall trace for all drivers. However, no additional analysis was conducted after applying the filter in the driver identification stage.

During the trace, the vehicle’s position is expressed as a 3-dimensional curve in the Earth-Centered, Earth-Fixed (ECEF) reference frame. Although there are elevation changes in the route, a vehicle’s movement along the route could be considered to be in a 2-dimensional plane. Thus, the ECEF coordinates of each trace were transformed into a local East, North, Up (ENU) reference frame.

3.2.1 Alignment

It was not possible to directly compare traces along the same route since the time required to complete each trace is different and the drivers would be

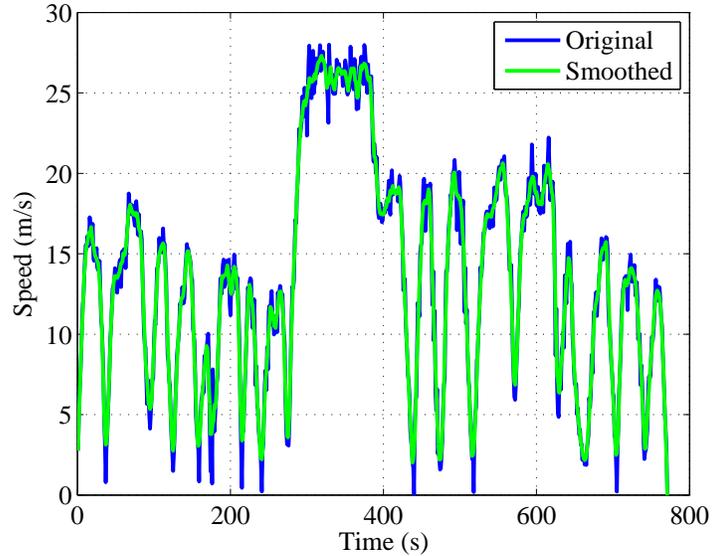


Figure 3.2: Effects of applying a moving-average filter on a velocity profile at different locations at any specified time. A position alignment tool was developed to address this issue. Given a reference trace and time vector, and a candidate trace and time vector, the alignment routine manipulated the candidate time vector so that the reference and candidate traces occupied the same locations at a given time. The results of applying this tool are shown in Figure 3.3.

The positions represented in this plot are the offsets from the mean position for the reference trace in the ECEF reference frame. The blue and green traces represent the actual recordings that were collected. These traces exhibit the same pattern, and aside from the beginning, occupy different positions at any given time. Application of this routine results in the red trace,

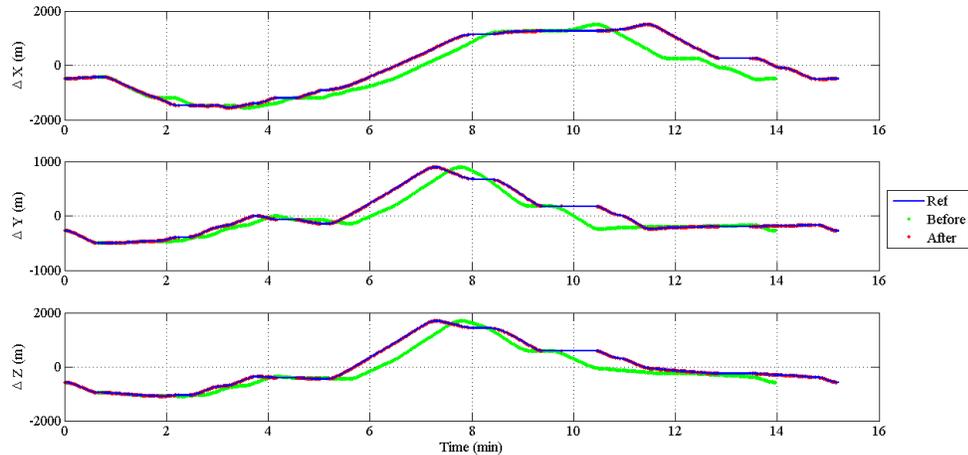


Figure 3.3: Effects of applying the position alignment routine

which “occupies” the same positions as the reference trace, at the same times. The reference trace was randomly chosen and the alignment routine was applied to all other traces, thus allowing for direct comparison between them. Application of this routine allowed for the computation of the mean trace for each driver and the mean overall trace for the route.

3.3 Parameter Extraction

The majority of the work in developing μ MAS was devoted to this phase. In order to properly characterize drivers, it is necessary that each parameter chosen be invariant with driver and conditions. These parameters were chosen because it is believed that they play an important role in distinguishing between drivers and include the following:

- Mean and standard deviation of speed

- Mean and standard deviation of the rate of change of speed when speeding up and when slowing down
- Trace repeatability
- Radius of curvature for each turn
- Average lateral acceleration for each turn
- Maximum lateral acceleration for each turn
- Frequency of lane changes
- Turning angle during lane maneuvers
- Lane preference in 2-lane environment
- Lane preference in 3-lane environment

These parameters can be decomposed into three classes and are discussed in further detail below.

3.3.1 General Characteristics

The general characteristics that describe drivers include the mean and standard deviation of speed, rate of change of speed when speeding up, and rate of change of speed when slowing down. Since drivers speed up and down differently, and interpret the speed limits in different ways, these characteristics can be used to describe the aggressiveness and the general mental state of the driver. Lastly, the trace repeatability of a driver was also computed. In general, drivers tend to drive in a similar manner when traversing the same route. The deviation of trace with respect to the mean overall trace, found during the data pre-processing phase, was used to define the trace repeatability

parameter.

3.3.2 Turning Behavior

It is assumed that the manner in which individuals make turns can be used to distinguish them. Thus, the radius of curvature during turns were found for each trace, in addition to the average and maximum lateral accelerations experienced during the turns. These parameters indicate the path driven by the driver, as well as their aggressiveness during turn maneuvers. Since the geometry of the turn affects driving behavior, these three parameters were extracted for each turn on the route.

3.3.3 Lane Mentality

Naturally, the lane changing habits of drivers also vary. While some drivers are passive and are satisfied remaining in a lane for long periods of time, others are active and tend to change lanes frequently. Thus, the frequency of lane changes during a trace can be found and expressed as the number of lane changes per kilometer traversed. Furthermore, the actual lane change maneuver varies from driver to driver. While some individuals tend to prefer gradual changes, which generally implies open roads, others are prone to sudden lane changes, which usually occur during congestion. Therefore, the turning angle during each lane change maneuver was used to define the aggressiveness of the maneuver. Lastly, when driving in multi-lane roads, the lane preference of the driver was also extracted.

3.3.4 Post-Extraction Process

After the parameters were extracted, using methods described in the next chapter, they were compiled into a parameter vector. Since $p = 11 + 3t$ parameters were extracted, where t is the number of turn segments in the route, the parameter vectors are elements of \mathbb{R}^p . The next steps depended on whether the system was in the database construction or driver identification stage. During the database construction stage, the parameter vectors for each trace used to construct the database were used to determine the mean parameter vector for each driver. On the other hand, during the driver identification stage, no further actions were taken and the system proceeded to the next phase.

3.4 Classification Routine

Suppose that data was recorded for m drivers on a route and that n_i traces were recorded for the i^{th} driver. Let \mathbf{v} be the parameter vector for the test trace and \mathbf{u}_{ij} be the parameter vector for the j^{th} trace of the i^{th} driver. Then, the mean parameter vector for i^{th} driver can be denoted as $\bar{\mathbf{u}}_i$. The M -ary hypothesis test were used to determine the identity of the mystery driver.

In binary hypothesis testing, a parameter takes on one of two discrete values and each hypothesis is mapped onto a point in the observation space [7]. This can be extended to the general case to contain m different hypotheses.

Let H_i denote the hypothesis that \mathbf{v} is distributed as p_i :

$$\begin{aligned}
H_1 &: \mathbf{v} \sim p_1 \\
H_2 &: \mathbf{v} \sim p_2 \\
&\vdots \\
H_m &: \mathbf{v} \sim p_m,
\end{aligned} \tag{3.1}$$

where $p_i(\mathbf{v}) = \mathcal{N}(\mathbf{v}; \bar{\mathbf{u}}_i, S_i)$. Let S_i be the sample covariance matrix for the data set $\mathbf{u}_{i1}, \dots, \mathbf{u}_{in_i}$ and is defined as

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{u}_{ij} - \bar{\mathbf{u}}_i)(\mathbf{u}_{ij} - \bar{\mathbf{u}}_i)^T. \tag{3.2}$$

Then,

$$p_i(\mathbf{v}) = \frac{1}{\sqrt{(2\pi)^{n_p} |S_i|}} \exp \left[-\frac{1}{2} (\mathbf{v} - \bar{\mathbf{u}}_i)^T S_i^{-1} (\mathbf{v} - \bar{\mathbf{u}}_i) \right]. \tag{3.3}$$

In this case, $\mathbf{v} \in \mathbb{R}^p$ is mapped onto a point in an observation space Z through the probabilistic mechanism described above. The observation space is divided into m regions, indicating the regions where the various hypotheses are declared. So, if the mapped point resides in Z_i , then H_i is declared correct and the mystery driver is identified as Driver i . The decision criterion for determining the driver is to maximize the probability of proper identification [8], which occurs when $p_i(\mathbf{v})$ is maximized. Therefore, if $p_i(\mathbf{v}) \in Z_i$ and H_i is declared correct, then the mystery driver will be

$$i = \arg \max_i p_i(\mathbf{v}), \tag{3.4}$$

Besides serving as a classification routine that can choose optimally between different hypotheses, this framework can also be used to measure the distance between two parameter vectors in \mathbb{R}^p by using the weighted Euclidean distance. For the vectors \mathbf{v} and \mathbf{u}_i , this distance is defined as

$$d_i = \left\| (\mathbf{v} - \bar{\mathbf{u}}_i)^T S_i^{-1} (\mathbf{v} - \bar{\mathbf{u}}_i) \right\|_2. \quad (3.5)$$

Chapter 4

Parameter Extraction Methods

A major focus of this study was to develop the techniques necessary to extract the parameters discussed in the previous chapters. The extraction process for each class is discussed below.

4.1 General Parameters

The coarse parameters of mean and standard deviation of velocity, acceleration, and deceleration were computationally simple and extracted directly from the smoothed speed and rate of change of speed profiles formed in the data pre-processing phase. The trace repeatability was defined as deviation of the trace from the overall mean trace. The position alignment routine was applied to the trace, with the overall mean trace being the reference. The repeatability parameter was then obtained by determining the average offset between the two traces.

4.2 Turning Behavior

Given a trace, it was first necessary to identify when a turn occurred. This process was automated, but did require prior knowledge of the number of turns

in the trace. The radius of curvature and lateral accelerations for each turn were computed after determining the turn segments for each trace.

4.2.1 Identifying Turn Segments

In μ MAS, a turn is defined to be a conscientious decision by the driver to change the vehicle's heading, not simply a change of heading due to the curvature of the road. If a decision is being made to change directions, it is likely that a reduction in speed occurred prior to the initiation of the turning maneuver. The first step in identifying turn segments was to identify all local minima from the velocity profile of the given trace, and treat the immediate neighborhood around each minima as a candidate turn segment. The smoothed speed profile for the first 100 seconds of a trace is shown in Figure 4.1, along with the candidate locations for a turn segment.

Suppose that the vehicle has been traveling in a straight line for some time. Then, the angle between successive velocity vectors on the trace would be close to 0. Now, suppose that the vehicle has made a turn. Then, the angle between a position vector right before and right after the turn would fairly large. This property motivated the construction of an angular profile of the trace, which represented the rate of change of the vehicle's heading. The value at any given time represented the angle between the velocity vector at the time and the velocity vector 10 seconds into the future. Thus, if a turn occurred within any 10 second interval, then a spike in the angular profile would be observed. However, if the heading remained fairly constant during this time

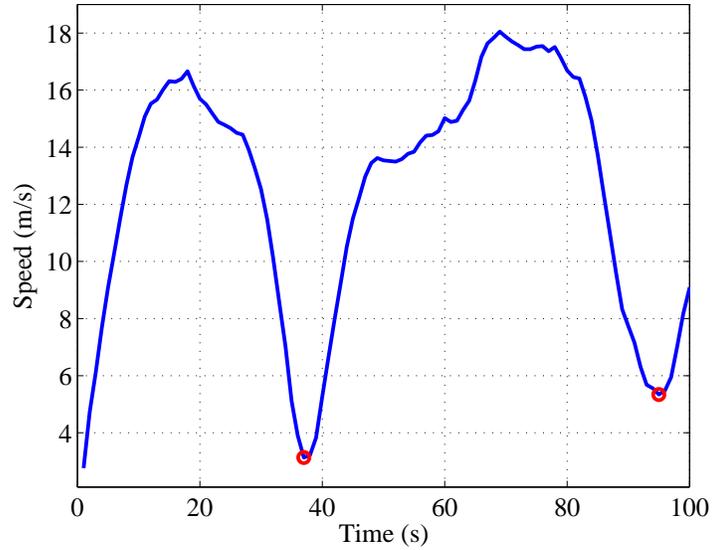


Figure 4.1: Speed profile with candidate turn segment locations

interval, then the angular profile would have no spikes in this interval. If the route was said to have t turn segments, then the neighborhoods around the t largest peaks in the angular profile coinciding with the candidate locations from the speed profile would be considered to be neighborhoods containing turn segments. Once the neighborhood was identified, a shape was fit to the peak in order to estimate the spread of the peak, which was then used to define the beginning and the end of the turn segment. This process is illustrated in Figure 4.2 and the first 100 seconds of this trace is shown in a local East-North plane in Figure 4.3. It can be observed that this automated process correctly identified the turn segment.

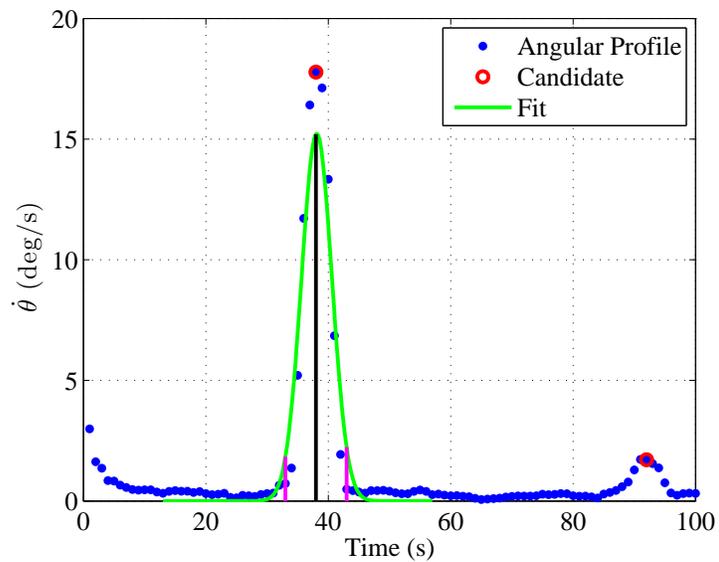


Figure 4.2: Angular profile with beginning and end of turn segment identified

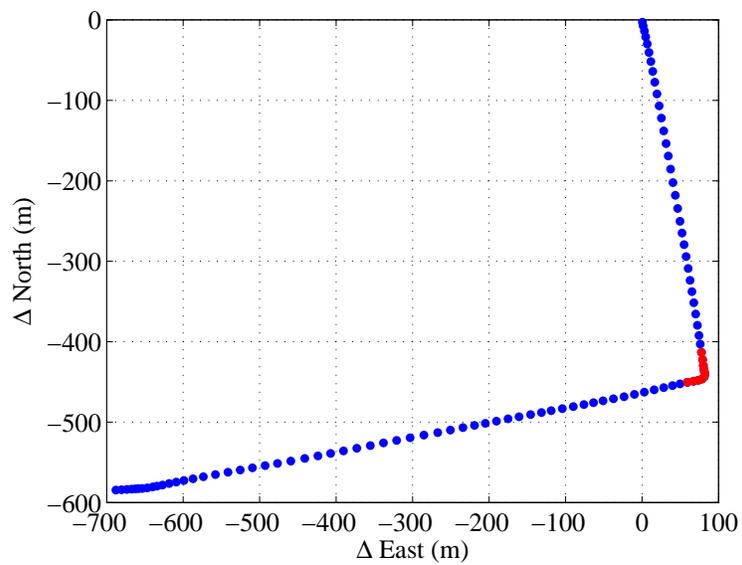


Figure 4.3: Correct identification of a turn segment during a trace

4.2.2 Radius of Curvature

Identification of the turn segment yielded measurements of the vehicle's position in the horizontal plane during the turn. Let the i^{th} position of the vehicle be denoted by (x_i, y_i) and the state vector \mathbf{x} , which needs to be estimated, be defined as $[x_c, y_c, r]^T$. The i^{th} position on the curve can be modeled as

$$0 = h'_i(\mathbf{x}, x_i, y_i) + w'_i = (x_i - x_c)^2 + (y_i - y_c)^2 - r^2 + w'_i, \quad (4.1)$$

where the noise w'_i is modeled as a Gaussian distribution with zero mean and variance of σ_p^2 . It is assumed that all measurements are equally noisy with $E[w'_i] = \sigma_p^2 = 0.1 \text{ m}^2$. Multiple measurements can be stacked to form

$$\mathbf{0} = \mathbf{h}'(\mathbf{x}, x, y) + \mathbf{w}', \quad (4.2)$$

where \mathbf{w}' is a multivariate Gaussian distribution with zero mean and covariance matrix $P_{w'} = \sigma_p^2 I$. The state vector can be approximated by minimizing (4.2) using the nonlinear least squares estimation method, which requires the Taylor series expansion

$$\mathbf{h}'(\mathbf{x}, x, y) \approx \mathbf{h}'(\bar{\mathbf{x}}, x, y) + H'(\mathbf{x} - \bar{\mathbf{x}}). \quad (4.3)$$

In this equation, $\bar{\mathbf{x}}$ is an approximate guess of \mathbf{x} and H' is the Jacobian matrix, whose i^{th} row is given by

$$H'_i = \left. \frac{\partial h'_i}{\partial \mathbf{x}} \right|_{\mathbf{x}=\bar{\mathbf{x}}} = [-2(x_i - \bar{x}_c), -2(y_i - \bar{y}_c), -2\bar{r}]. \quad (4.4)$$

The state vector was obtained by minimizing the cost function

$$J_{NL} = \|\mathbf{h}'(\bar{\mathbf{x}}, x, y) + H'(\mathbf{x} - \bar{\mathbf{x}})\|^2. \quad (4.5)$$

A Cholesky factorization was then applied to $P_{w'}$ to obtain $P_{w'} = R_a^T R_a$, which was then used to define the following variables:

$$\mathbf{h}(\bar{\mathbf{x}}, x, y) = R_a^{-T} \mathbf{h}'(\bar{\mathbf{x}}, x, y), \quad H = R_a^{-T} H', \quad \mathbf{n} = \mathbf{h}(\bar{\mathbf{x}}, x, y) + H\bar{\mathbf{x}} \quad (4.6)$$

The resulting cost function,

$$J_{NL} \approx J_L = \|H\mathbf{x} + \mathbf{n}\|^2, \quad (4.7)$$

was then used to solve for \mathbf{x} using the standard least squares estimation method. This yielded the center and radius of curvature for the given turn.

4.2.3 Lateral Acceleration During Turns

The lateral acceleration experienced during a turn, a_c is defined as

$$a_c = \frac{v^2}{r}, \quad (4.8)$$

where v is velocity and r is the radius of curvature, which can be found using the method described above. In this study, the average and maximum lateral accelerations were computed by replacing the velocity in the above equation with either the average or maximum velocity experienced during the turn, respectively.

4.3 Lane Mentality

Before extracting the parameters associated with lane mentality, it is necessary to determine when lane change maneuvers are actually occurring. However,

in order to determine when lane change maneuvers occur, it is necessary to determine what lane a driver is traversing at any given location on the route. This was achieved by first determining the center of each lane at all locations around the route. This information was subsequently used to extract the relevant parameters.

4.3.1 Determining Lane Centers

Over short periods of time, the drivers path when not executing turning maneuvers is approximately linear. Therefore, each trace was divided into 10 second non-turning segments and a second-order polynomial was fit through the segments for each trace in the local East-North reference frame. A similar polynomial fit was also applied to the corresponding segment on the mean overall trace.

Consider a particular segment on the route. At each location on this segment, a line perpendicular to the polynomial fit of the mean overall trace was computed. The intersection of this perpendicular line and the polynomial fits for all traces was found. This resulted in a set of collinear points in \mathbb{R}^2 , which was then transformed onto \mathbb{R} . A k -means clustering algorithm was applied, which determined which cluster each point was a member of. The number of clusters, k , formed depended on whether the number of lanes the segment contained. After application of this clustering algorithm, the center of each lane for each location on the segment was determined.

The results of this process is shown in Figure 4.4 for a segment along

the route. It can be observed that this segment is a two-lane road and that Driver 2 executed a lane change maneuver during one of his traces during this segment. Furthermore, the process correctly identified the center of each lane in this segment at various locations.

4.3.2 Parameters

In addition to determining the centers of each lane, the k -means clustering algorithm also determined which lane the vehicle was for each trace. By knowing what lane a vehicle occupied throughout the trace, the segments when a lane change occurred could be easily extracted. Once the number of lane changes was known, the frequency was found, which was defined to be the number of lane change maneuvers per kilometer traversed. Furthermore, the time spent in each lane during two and three-lane segments was also determined. The weighted average of the lane and the time spent in each lane was used to compute the two lane preference parameters. The turning angle was determined by computing angle between the velocity vector at the beginning of the maneuver and the velocity vector in the middle of the maneuver. Since there were multiple lane change maneuvers during a route, the average turning angle for all maneuvers was used as the turning angle parameter.

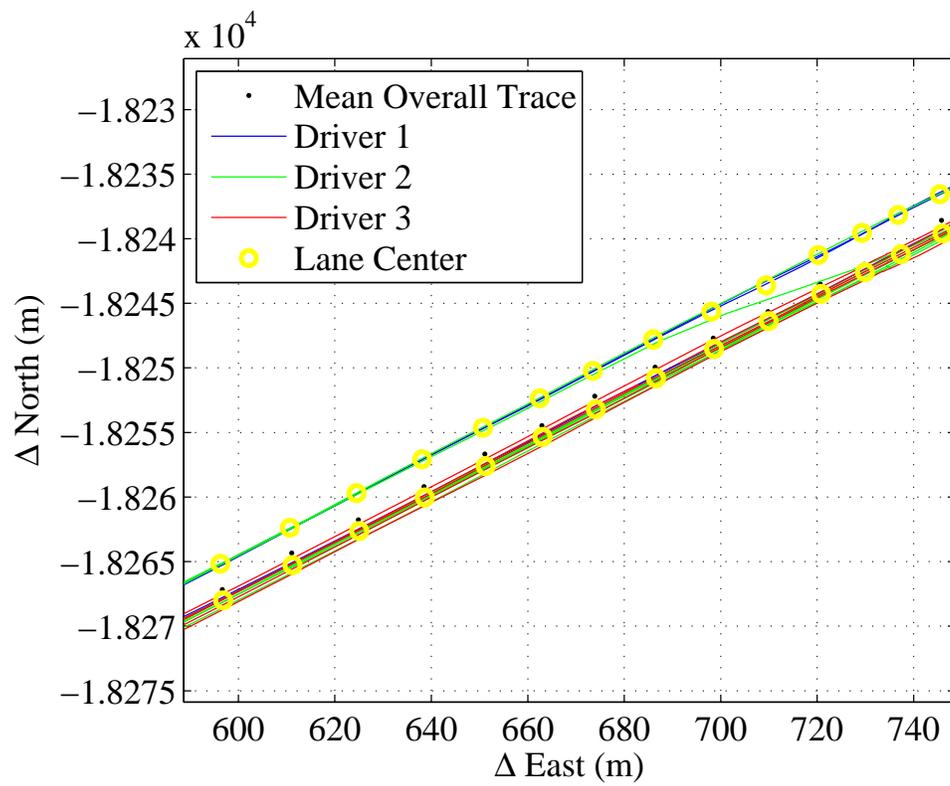


Figure 4.4: Location of lane centers

Chapter 5

Results

As stated before, sixteen traces for three drivers were collected around one route. A simulation was set up to analyze the performance of μ MAS.

5.1 Cross-Validation Simulation

In order to test the performance of μ MAS, a cross-validation simulation was set up. In this scenario, each of the sixteen recorded traces took turns being the test trace while the remaining ones were used to construct the database. This data was then run through μ MAS to determine if the mystery driver was correctly identified. The performance of this system is summarized in Table 5.1 and indicates the frequency of correct identification for each driver using the M -ary hypothesis test framework.

It can be observed that this classification routine was not fully successful in correctly identifying the driver. A probable cause can be attributed to the manner in which the drivers drove during the data collection phase. While Driver 1 drove consistently throughout and did not have significant changes in his style, Drivers 2 and 3 drove in a very cautious manner during their first few traces, but when instructed to drive normally, their style changed. This

Table 5.1: Performance of μ MAS

Driver	Correct Identification	Traces Collected
Driver 1	5	5
Driver 2	3	7
Driver 3	1	4
Total	9	16

change was more pronounced for Driver 3 than it was for Driver 2. Since only a few traces were collected, a change in the driving style for a driver had a significant impact on the performance of μ MAS. When the driver was consistent, as in the case of Driver 1, the resulting parameters for each trace were similar and had a low variability. However, when there was a style change during data collection, as in the case of Driver 3, the resulting parameters for each trace had a high variability, which made the system believe that the data came from different drivers, instead of a single one. This sample size is too small to properly assess μ MAS. More traces are required to properly characterize its performance and account for the variability in the driver’s style.

Further reasons for the low performance could be due to the inclusion of non-informative parameters, which leads to noise [9], measurement error caused during parameter extraction, or unfamiliarity with the vehicle. Data was collected for all three drivers using the vehicle for Driver 1. Since Drivers 2 and 3 had never driven the vehicle before, they were not accustomed to its handling characteristics and nuances. The unfamiliarity of Drivers 2 and

3 with the vehicle served as a confounding variable and the evolving driving styles of these drivers could be due to their increasing familiarity with the vehicle as more time was spent driving it.

5.2 Future Work

As discussed before, more data is required to improve the performance of μ MAS. Priority will be placed on collected a large volume of traces for as many drivers as possible. In this study, the drivers changed their behavior during the data collection process. Ideally, a recording device would be placed on top of the vehicle and the driver would not know that they were being recorded, thus ensuring that their behavior does not change. However, this is not possible at the moment. Instead, the drivers need to be better informed about what is expected from them during the data collection process.

The parameters extracted in this study were chosen based on intuition. It was believed that these parameters would be able to distinguish between drivers. However, there are more parameters that can be extracted. It would be beneficial to understand which of the parameters play a significant role in distinguishing between drivers and isolate non-informative parameters. The Principal Component Analysis (PCA) method can be used as a data reduction mechanism to yield this information. This method incorporates an orthogonal transformation to transform a set of potentially correlated parameters into a set of linearly uncorrelated ones.

In this study, M -ary hypothesis testing was used for the classification

routine. However, the data mining field is rich with methods that are commonly used for classification purposes. Incorporation of techniques such as Linear Discriminant Analysis (LDA) and Decision Trees would allow for the comparison of multiple methods in this application [10].

Hypothesis testing is a statistical process and has the potential for false driver identification. After more data is collected, the probability of false identification should be computed for this system to further characterize its performance and reliability.

With the tools developed in this study, various questions regarding the driving behaviors of individuals can be answered. Some of them include:

- Is there truth behind the stereotypes regarding the driving behaviors of individuals from various groups?
- Can a commercial driver be distinguished from a non-commercial driver when driving the same, non-commercial vehicle?
- Does a person's driving style depend on the time of day?
- Can drivers driving for different companies with similar vehicles be distinguished from each other?
- How does a driver's behavior change if they are driving under the influence?

Chapter 6

Conclusion

In order to develop quantitative tools to construct safety and comfort profiles of drivers, the Micro-Mobility Analysis System was designed to distinguish between drivers based on their driving habits. This involved solving a pattern recognition problem, in which certain parameters were extracted from a test trace, whose driver was unknown, and compared to an existing database of parameters for various drivers in order to determine the identity of the mystery driver. Since the resolution of standard GPS tracking devices were inadequate for this application, Precise Point Positioning was incorporated as a post-processing tool to yield the desired resolution.

An existing pseudorange-based, single-frequency position solution algorithm was augmented with accurate ionospheric data, in addition to precise satellite positions and clock states. In this PPP algorithm, a modification to Single Layer Model was proposed and results indicate that the modification led to an improvement of 5.2 *cm*. An online PPP engine was utilized to post-process recorded traces.

This post-processed trace was used to extract various parameters related to a driver's general characteristics, turning behavior, and lane change

mentality. Various methods were developed for this extraction process and resulted in parameter vectors for each trace recorded. In this study, M -ary hypothesis testing was used as a classification routine to determine the identity of the mystery driver.

A cross-validation simulation was constructed to measure the performance of this system. The results indicated that the system was successful in identifying one of the three drivers, but not the other ones. Various reasons were proposed to account for this result and a larger volume of data is required to fully understand the performance of this system. Due to the small sample size, no conclusions were made regarding its performance.

The classification routine utilized in μ MAS was also used to define distance between drivers as a weighted Euclidean distance. This distance can be used to place drivers into clusters. It is conjectured that a correlation exists between a driver's safety record and the cluster they belong. If the safety records of drivers are made available, then this conjecture can be tested. If a correlation does exist, then μ MAS can be used to classify drivers based on their safety records. Furthermore, μ MAS can also be used to answer a wide range of questions regarding the driving behaviors of individuals.

Bibliography

- [1] A. Sadilek and J. Krumm, “Far out: Predicting long-term human mobility,” in *Association for the Advancement of Artificial Intelligence*, July 2012.
- [2] Z. Constantinescu, C. Marinoiu, and M. Vladoiu, “Driving style analysis using data mining techniques,” *International Journal of Computers, Communications & Control*, vol. 5, pp. 654–663, 2010.
- [3] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 4–20, Jan. 2004.
- [4] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*. Lincoln, Massachusetts: Ganga-Jumana Press, revised second ed., 2012.
- [5] N. Ya’acob, M. Abdullah, and M. Ismail, “Determination of gps total electron content using single layer model (SLM) ionospheric mapping function,” *International Journal of Computer Science and Network Security*, vol. 8, pp. 154–160, 2008.
- [6] Y. Gao, “Precise point positioning and its challenges,” *GNSS Solutions Column, Inside GNSS*, vol. 1, no. 8, pp. 16–18, 2006.

- [7] H. Van Trees *Detection, Estimation, and Modulation Theory, Part I*. New York: John Wiley and Sons, April 2004.
- [8] R. G. Gallager, *Stochastic Processes, Theory for Applications*. Cambridge, UK: Cambridge University Press, 2013.
- [9] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013.
- [10] A. M. Martinez and A. C. Kak, "PCA versus LDA," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 228–233, Feb. 2001.