GISTMARG:

GPS and GIS for Traffic Monitoring and Route Guidance

by

Young-Ji Byon

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Civil Engineering
University of Toronto

© Copyright by Young-Ji Byon 2011
Abstract

Cell phone providers have started fitting assisted global positioning system (AGPS) chips in new cell phones. The phone users travel on the roads voluntarily and if the phones can be queried anonymously at a reasonable cost, the phones can replace costly dedicated probe vehicles. The major challenge in using AGPS chips is that the phones may not always be on the roads. Even if they were on the roads, they could be in different modes of transportation. Since traffic conditions are usually monitored in terms of private automobiles, the modes of transportation the phones are in first needs to be determined in order for them to be used as traffic probes. Once the traffic data is correctly identified as coming from the private automobile mode, it is another challenge to combine the information with different information sources. This thesis develops a method to fuse the multiple sources of traffic data for more reliable estimation of traffic conditions. Traditionally, route guidance systems generally have focused on the shortest or the fastest route for a particular pair of origin and destination. However, by utilizing available three dimensional (3-D) and geospatial data, it is possible to aid route guidance systems that are optimizing for other objectives such as finding the most scenic, the most level or the safest route.
The objective of this thesis is to develop a framework of traffic monitoring and route guidance system named GISTMARG (GPS and GIS for Traffic Monitoring and Route Guidance) that deals with incorporating cell phones or mobile GPS devices as probes, fusing the traffic data from multiple sources and incorporating 3-D and geospatial data into route guidance methods.
Acknowledgments

“The fear of the LORD is the beginning of knowledge” (Proverbs 1:7)

I thank God for helping me throughout the research and using me for plans He has for me.

I would like to acknowledge all those who contributed either directly or indirectly to this research. First of all, I would like to express my deep sense of gratitude and indebtedness to my supervisors Prof. Baher Abdulhai and Prof. Amer Shalaby who gave me the opportunity to conduct this research. They both guided me with specific research directions and financially supported me throughout the research period. Without them, it would have been impossible to complete this Ph.D. thesis work. I would like to thank Prof. Eric J Miller and Prof. Matthew Roorda for equipping me with the fundamental knowledge in transportation through their courses and providing me with career related advices throughout my thesis work.

I would like to acknowledge Samah El-Tantawy and Hossam Abd El-Gawad for helping me with the QQ network micro-simulation in the model verification stage in this thesis.

I would like to thank all my family members Sung-Sang Byron, An-Soon Byron and Ji-Hyun Byron for their support and prayers. I also thank Hak-Chul Choi, Young Ja Choi, Hyo Sang Byron and Bok Ae Byron for being my mentors.

Finally, I would like to thank all volunteers and survey participants who participated in this research including Hyo-Jung Choo, Joo-Hyun Song and Ho Kyung Moon.

Chapter 4 and Chapter 6 are reproduced with permissions of Taylor & Francis and Transportation Research Board respectively.
Table of Contents

Contents

Abstract ........................................................................................................................................... ii
Acknowledgments .......................................................................................................................... iv
Table of Contents .............................................................................................................................. v
List of Tables ..................................................................................................................................... xi
List of Figures .................................................................................................................................... xii
List of Appendices ............................................................................................................................. xvi
Abbreviations ..................................................................................................................................... xvii
Chapter 1 ........................................................................................................................................... 1
  1 Introduction ................................................................................................................................... 1
    1.1 Introductory Research Background ......................................................................................... 1
    1.2 Objective ................................................................................................................................... 6
    1.3 GISTMARG Overview .............................................................................................................. 7
    1.4 Research Challenges and Issues ............................................................................................... 8
      1.4.1 Detecting Mode from Relatively Small Amount of Data ..................................................... 8
      1.4.2 High Cost Associated with Pinging ..................................................................................... 9
      1.4.3 Lack of Location-based Services Available in Toronto ...................................................... 10
      1.4.4 Urban Canyon ..................................................................................................................... 11
      1.4.5 Information Fusion of Multiple Information Sources .................................................... 13
      1.4.6 Lack of Detailed Elevation Models ..................................................................................... 14
      1.4.7 Subjective Measure of Safety and Scenic Quality .......................................................... 14
    1.5 Overview of Proposed Approach ............................................................................................. 15
      1.5.1 Neural Network (NN) for Mode Detection ......................................................................... 15
1.5.2 Kalman Filters for Data Fusion and Prediction ............................................. 15
1.5.3 3-D GIS for Route Guidance ..................................................................... 16
1.6 Thesis Outline ................................................................................................. 16

Chapter 2 ............................................................................................................ 18

2 Background ....................................................................................................... 18

2.1 Moving Forward From GISTT ...................................................................... 18

2.1.1 GISTT (GPS-GIS Integrated System for Travel Time Surveys) .............. 18
2.1.2 GISTT vs. GISTMARG .......................................................................... 20

2.2 Mode Detection ............................................................................................ 21

2.2.1 Neural and Adaptive Systems (NAS) ..................................................... 22
2.2.2 Linear Adaptive System ......................................................................... 23
2.2.3 Regression for Multiple Variables .......................................................... 27
2.2.4 Pattern Recognizer ................................................................................. 30
2.2.5 Artificial Intelligence ............................................................................. 34
2.2.6 Artificial Neural Networks (ANNs) ....................................................... 34
2.2.7 Neurosolutions ....................................................................................... 44
2.2.8 Neural Network vs. Fuzzy Logic ........................................................... 46

2.3 Data Fusion and Kalman Filtering ............................................................... 48

2.3.1 Introduction to Data Fusion .................................................................... 48
2.3.2 Introduction to Kalman Filters ............................................................... 50
2.3.3 Formulation of a Kalman Filter ............................................................. 51
2.3.4 SCAAT Kalman Filters ......................................................................... 52
2.3.5 Quantifying Subjective or Non-numerical Information ......................... 54

2.4 Global Positioning System (GPS) ............................................................... 57

2.4.1 Background of GPS ............................................................................... 57
2.4.2 The 3 Segments of GPS ....................................................................... 57
2.5 Geographic Information Systems (GIS) ........................................................................................................60
  2.5.1 Vector vs. Raster Layers ..........................................................................................................................61
  2.5.2 Digital Elevation Model (DEM) .............................................................................................................62
2.6 Different Levels of Route Guidance Systems .................................................................................................64
2.7 Paramics Micro-simulation .............................................................................................................................64

Chapter 3 ...........................................................................................................................................................65

3 State of the Art and Literature Review ..........................................................................................................65
  3.1 Traffic Monitoring .........................................................................................................................................65
     3.1.1 Vehicle Information and Communication System (VICS) (Kyobashi, 2008) .........................65
     3.1.2 Traffic Monitoring Using Digital Sound Field Mapping (Chen et al., 2001) .................66
     3.1.3 Efficient Vehicular Traffic Monitoring Using Mobility Management in Cellular Networks (Saraydar et al., 2004) ......................................................................................................................66
     3.1.4 Microscopic Traffic Data Collection by Remote Sensing (Hoogendoorn et al., 2003) .................67
     3.1.5 An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys (Tsui and Shalaby, 2006) ......................................................................................................................67
     3.1.6 Using Information Technology to Evaluate Non-traditional Traffic Monitoring Systems (Guo et al., 2007) .................................................................................................................................................68
     3.1.7 Kalman Filtering Applied to Network-based Cellular Probe Traffic Monitoring (Qiu and Ran, 2008) .................................................................................................................................................68
     3.1.8 Evaluation of a Cellular Phone-based System for Measurements of Traffic Speeds and Travel Times: A Case Study from Israel (Bar-Gera, 2007) .................................................................68
     3.1.9 Transit Vehicles as Traffic Probe Sensors (Cathey and Dailey, 2002) .................................69
     3.1.10 Cellphone Probes as an ATMS Tool (Smith et al., 2003) .........................................................69
     3.1.11 Reconsideration of Sample Size Requirements for Field Traffic Data Collection with Global Positioning System Devices (Li et al., 2002) .................................................................70
  3.2 Data Fusion ....................................................................................................................................................70
     3.2.1 Probe Vehicle Runs or Loop Detectors? Effect of Detector Spacing and Sample Size on the Accuracy of Freeway Congestion Monitoring (Kwon et al., 2007) .................................................................71
3.2.2 Traffic Flow Reconstruction Using Mobile Sensors and Loop Detector Data (Herrera and Bayen, 2008) ................................................................. 71
3.2.3 SCAAT: Incremental Tracking with Incomplete Information (Welch, 1996) ...... 72

3.3 Route Guidance ........................................................................................................... 72

3.3.1 Vehicle Route Guidance Systems: Classification and Comparison (Schmitt and Jula, 2006) .................................................................................. 72
3.3.2 Evaluation of Walkability of Routes (Scholossberg, 2006) and (Scholossberg et al., 2007) ..................................................................................... 73
3.3.3 Enhancing In-car Navigation Systems with Personal Experience (Bederson, 2008) ............................................................. 73
3.3.4 Comparative Study of Decentralized Feedback Control Strategies for Route Guidance Purposes (Zuurbier, 2007) ................................................. 74
3.3.5 A Behavioral Component Analysis of Route Guidance Systems Using Neural Networks (Hamad, 2003) ........................................................................ 74
3.3.6 Park and Bike: A New Multi-modal Concept for Congested Areas (Bos and Vrugt, 2008) ................................................................. 75

Chapter 4 .............................................................................................................................. 76

4 Transportation Mode Detection ....................................................................................... 76

4.1 Introduction .................................................................................................................... 77
4.2 Study Objective ............................................................................................................. 78
4.3 Data Collection ............................................................................................................. 78
4.3.1 Study Region ......................................................................................................... 79
4.3.2 Data Post-Processing to Compute Acceleration .................................................. 83
4.4 Classification of Process Inputs and Outputs ............................................................. 83
4.5 Design of Neural Network Classifier for Mode Detection .......................................... 85
4.6 Experimental Scenarios and Results ......................................................................... 88
4.6.1 Universal Mode Detection Neural Network for 4 Different Modes ..................... 89
4.6.2 Impact of Varying Pinging Frequency and Monitoring Duration ...................... 90
4.6.3 Route-Specific Auto Mode Detection Neural Network ........................................ 93
4.6.4 Auto Mode Detection Neural Network during Peak and Non-Peak Period .................94
4.6.5 Specification of NNs in Neurosolutions .........................................................................95
4.6.6 Sensitivity Analysis ........................................................................................................95
4.6.7 Discussion on Low Satellite Visibility Regions .................................................................99
4.6.8 Effect of Inclusion of Data from Incorrectly Identified Mode on Travel Time Estimation .........................................................................................................................100
4.7 Summary, Conclusions and Future Research .....................................................................103

Chapter 5 .................................................................................................................................106
5 Data Fusion of Different Sensors ..........................................................................................106
5.1 Introduction ..........................................................................................................................107
5.1.1 Robust Traffic Information with Multiple Information Sources .................................109
5.1.2 Calibration of SCAAT Filter ............................................................................................111
5.1.3 Used Parameter Values .....................................................................................................111
5.2 First Phase: Data Fusion for Highway 401 in Toronto .......................................................111
5.2.1 Data Collection ................................................................................................................112
5.2.2 Description of Available Data Sources .........................................................................112
5.2.3 Design of Experiment .......................................................................................................116
5.2.4 Results of Data Fusion ......................................................................................................118
5.2.5 Traffic Monitoring GIS Database .....................................................................................121
5.3 Second Phase: Data Fusion in Micro-simulation .................................................................123
5.3.1 Network Modeling in Paramics .......................................................................................124
5.3.2 Used Physical Parameters in Paramics ............................................................................126
5.3.3 Data Collection from Micro-simulation ...........................................................................128
5.3.4 Evaluation of Estimation Error .........................................................................................130
5.3.5 Second Phase Results of Data Fusion ..............................................................................132
5.4 Conclusion ............................................................................................................................139
Chapter 6 .......................................................... 140

6 Multi-criteria Route Guidance .......................................................... 140

   6.1 Introduction .................................................................................. 141

6.2 Multi-criteria Data Collection .......................................................... 144

   6.2.1 Scenic View and Attraction Zone Survey ................................. 144

6.2.2 Digital Elevation Model ......................................................... 147

6.2.3 Bike Paths ............................................................................... 149

6.3 Construction of Various GIS Layers .............................................. 149

   6.3.1 Scenic View Layer (SVL) ..................................................... 149

6.3.2 Road Slope Layer (RSL) ....................................................... 154

6.3.3 Crime Rate Layer (CRL) ....................................................... 155

6.3.4 Determination of Minimum Cost Path .................................... 157

6.4 Illustrative Case Studies in Toronto .............................................. 157

   6.4.1 Across Toronto Scenario ..................................................... 157

6.4.2 Across Downtown Scenario for Safest Route Computation ........ 160

6.4.3 Further Potential Application of 3-D GIS Methods ................... 161

6.5 Conclusion .................................................................................. 162

7 Conclusions and Recommendations .............................................. 163

   7.1 Summary .................................................................................. 163

7.2 Recommendations ..................................................................... 165

7.3 Future Research ....................................................................... 166

References ..................................................................................... 168

Appendices ..................................................................................... 183
List of Tables

Table 2.1 Comparison of GISTT and GISTMARG ................................................................. 21
Table 2.2 Common components of Neurosolutions ............................................................... 45
Table 4.1 4 Mode detection rates (%).................................................................................. 90
Table 4.2 Auto-auto and non-auto-non-auto detection rates (%) .......................................... 91
Table 4.3 Sensitivity analysis on speed and acceleration ....................................................... 96
Table 4.4 % of Bike data inclusion vs. travel time estimation error ....................................... 102
Table 5.1 Categorization and quantification of verbal information from radio broadcasts ..... 116
Table 5.2 Travel time estimation deviation from the truth with varying sampling interval ..... 122
Table 5.3 Network-wide parameters used in Paramics ......................................................... 127
Table 5.4 Vehicle characteristics parameters ...................................................................... 128
Table 5.5 Conventional performance measures ..................................................................... 131
Table 5.6 Effect of varying number of probe vehicles .......................................................... 133
Table 5.7 Effect of varying sampling strategy ....................................................................... 136
Table 5.8 Varying sampling interval vs. number of probes .................................................... 137
Table 5.9 Effect of varying sampling interval ....................................................................... 138
List of Figures

Figure 1.1 Probe vehicle ........................................................................................................... 2

Figure 1.2 Assisted Global Positioning System (AGPS) .......................................................... 3

Figure 1.3 Triangulation of cell phone towers for location determination ............................... 4

Figure 1.4 GISTMARG overview .......................................................................................... 8

Figure 1.5 Alternative methods of receiving GPS data ............................................................ 10

Figure 1.6 Urban canyon effect ............................................................................................. 12

Figure 1.7 Multipath effect ...................................................................................................... 13

Figure 2.1 Adaptive system's design (Principe et al. 2000) .................................................... 22

Figure 2.2 Linear processing element .................................................................................... 24

Figure 2.3 Linear adaptive system .......................................................................................... 25

Figure 2.4 Performance graph (Principe et al. 2000) ............................................................ 26

Figure 2.5 Search method on performance surface with gradient information (Principe et al.
2000) ........................................................................................................................................ 27

Figure 2.6 The 2-D performance surface (Principe et al. 2000) ............................................. 28

Figure 2.7 Regression for multiple variables .......................................................................... 29

Figure 2.8 Classification boundary ........................................................................................ 32

Figure 2.9 Discriminant functions and the decision surface (Principe et al. 2000) .................. 33

Figure 2.10 Parametric classifier for c classes ....................................................................... 34

Figure 2.11 Biological neuron in a human brain ..................................................................... 35
Figure 2.12 McCulloch-Pitts processing element (artificial neuron) .................................. 37

Figure 2.13 Partial derivatives in an M-P processing element ........................................ 38

Figure 2.14 Perceptron ........................................................................................................ 40

Figure 2.15 One hidden layer perceptrons ......................................................................... 42

Figure 2.16 Detail of a hidden-layer .................................................................................... 43

Figure 2.17 Screenshot of one-hidden-layer perceptrons in Neurosolutions ....................... 46

Figure 2.18 Temperature membership function example ................................................... 47

Figure 2.19 General set-up of a Kalman filter (Welch and Bishop, 2007) ............................ 50

Figure 2.20 Screenshot of colour coded traffic monitoring web sites for Hwy 401 and arterial roads ............................................................................................................................ 56

Figure 2.21 6 sigma covering 99.7% of population in a normal distribution ....................... 56

Figure 2.22 The 3 segments of GPS ..................................................................................... 58

Figure 2.23 The orientation of the 24 GPS satellites (Chung, 2003) .................................. 59

Figure 2.24 Calculation of the current location of the GPS device ..................................... 60

Figure 2.25 Vector vs. raster formats .................................................................................. 62

Figure 2.26 Digital elevation model example ...................................................................... 63

Figure 4.1 Arterial routes for the data collection in the Greater Toronto Area (GTA) .......... 81

Figure 4.2 Highway data collection routes in the GTA ....................................................... 82

Figure 4.3 Highway 427 route with auto mode and bus mode only .................................... 82

Figure 4.4 Distribution of number of satellites in view for different modes of transportation.... 85

Figure 4.5 Multi-layered feed-forward neural network with 4 available modes .............. 86
Figure 4.6  Auto-auto and non-auto-non-auto mode detection rate surfaces................................. 92

Figure 4.7  Average mode detection rate vs. varying % change in both speed and acceleration values .................................................................................................................................................. 97

Figure 4.8  Sensitivity of input factors (speed and acceleration values) on mode detection rates 99

Figure 4.9  Study region for the data collection of bike and auto modes ........................................ 101

Figure 4.10  Absolute % deviation on travel time estimations vs. % of bike data inclusion..... 103

Figure 5.1  Illustration of various sources of traffic monitoring...................................................... 108

Figure 5.2  An example of missing information of loop detectors ................................................. 110

Figure 5.3  Data fusion of various sensors and comparison of SCAAT filters with varying GPS sampling frequencies ........................................................................................................................................ 119

Figure 5.4  Relative EUT of SCAAT filters with varying sampling intervals with respect to SCAAT with GPS sampled at 1 Hz ........................................................................................................ 120

Figure 5.5  A prototype of real-time traffic monitoring GIS database ........................................... 121

Figure 5.6  Ground truth based on all vehicles on the road from Paramics................................. 124

Figure 5.7  Queen's Quay network.................................................................................................. 126

Figure 5.8  Data collection area for the data fusion using Paramics................................................ 129

Figure 6.1  3-D building models in Google EarthTM for Tokyo, Japan (Top Left) and Toronto, Canada (Bottom Right). (Accessed June 2009) ........................................................................................................... 143

Figure 6.2  Comparison of 2-D and 3-D road maps in Toronto .................................................... 144

Figure 6.3  Zonal division of the GTA ............................................................................................ 145

Figure 6.4  Scenic route survey form used ...................................................................................... 146

Figure 6.5  Canadian Digital Elevation Model (CDEM) of Toronto............................................. 148
Figure 6.6  The 9 controllable characteristics of the viewshed tool ........................................... 150

Figure 6.7  The visibility from Highway 401 based on the 10 observation points (top) and the visibility from the Gardiner Expressway based on 10 observation points (bottom) ...................... 152

Figure 6.8  Computation of the scenic value of a particular link .................................................. 153

Figure 6.9  Slope variations in the GTA ......................................................................................... 154

Figure 6.10  Homicides in Toronto from 2005 to 2008 ................................................................. 156

Figure 6.11  Computation of the measure of the safety risk of a link .......................................... 157

Figure 6.12  Test case scenario routes for the evaluation of 3-D GIS route guidance ............... 158

Figure 6.13  Micro-level crime rate layer ...................................................................................... 161
List of Appendices

Appendix A. Weights of the trained auto mode classifier (10 pings per 10 minutes scenario)

Appendix B. Traffic count example of the intersection at Front Street and Church Street used for calibration of Paramics simulation
Abbreviations

2-D: 2 Dimensional
3-D: 3 Dimensional
a-a: Auto-Auto (Auto mode is correctly identified as auto mode)
AI: Artificial Intelligence
ANN: Artificial Neural Network
BOB: Bike-On-Bus
CBD: Central Business District
CDEM: Canadian Digital Elevation Model
CRL: Crime Rate Layer
DEM: Digital Elevation Model
DGPS: Differential GPS
DOT: Department of Transportation
GIS: Geographic Information Systems
GISTT: GPS and GIS for Travel Time Surveys
GPS: Global Positioning System
GTA: Greater Toronto Area
KF: Kalman Filter
LBS: Location Based Services
n-n: Non-auto-Non-auto (Non-auto mode is correctly identified as non-auto mode)
NN: Neural Network
NNs: Neural Networks
OD: Origin-Destination
PDA: Personal Digital Assistant (device)
RSL: Road Slope Layer
SA: Selective Availability of GPS
SCAAT: Single-Constraint-At-A-Time
SVL: Scenic View Layer
TAZ: Transportation Analysis Zone
TTS: Transportation Tomorrow Survey
VBA: Visual Basic for Application
VL: Visibility Layer
WAAS: Wide Area Augmentation System
WLT: Wireless location technology
ZSWRL: Zonal Scenic Worth Raster Layer
1.1 Introductory Research Background

Loop detectors and video cameras have been used for traffic monitoring in recent decades. Both loop detectors and video cameras are capable of monitoring traffic conditions at a limited number of fixed points on the roads. The loop detectors provide continuous traffic information at fixed points that are easily converted into digital formats and numerical values. Video cameras can provide visual live information but it usually requires subjective interpretation since processing digital video stream into meaningful traffic information requires a generally expensive conversion process. In conventional traffic monitoring, loop detectors provide numerical values such as speed and volume while video cameras need traffic personnel who are trained to interpret the live video and take appropriate actions in a traffic control centre. With emerging wireless communication and Global Positioning System (GPS) related technologies, the current trend is to incorporate real-time traffic information from probe vehicles (Figure 1.1) as moving sensors on the roads.
Probe vehicles provide more realistic experience of a traffic flow in the sense that they actually move with the traffic while providing information on traffic conditions throughout the road links as if they were moving sensors. As of now, probe vehicles are considered as an extra traffic monitoring technique that aids currently existing loop detector based traffic monitoring systems. Traffic departments send out probes to major congested sections of their road network to closely monitor those links that matter the most where loop detectors alone cannot provide desired level of information. Traffic departments can choose to use the probing technique in real-time or off-line. In the case of real-time probing, they send out vehicles with dedicated wireless communication equipments to provide the live traffic information. In the off-line, or passive probing, they equip the vehicles with GPS data loggers that are capable of archiving detailed second-by-second trip information that are to be post-processed afterwards to find out the trends.
in traffic conditions. Generally, a probe vehicle requires a dedicated driver, location-determining technology such as a mobile GPS device, and a wireless communication device to send real-time traffic information to the real-time central database. It is costly to operate even one probe vehicle with such equipment and a dedicated driver. Therefore, operating multiple probe vehicles of this type would not be economically feasible, or would be very expensive at best.

From early 2000, cell phone service providers in the United States of America have started embedding assisted GPS (AGPS) chips in their mobile devices to enhance their location-based services (LBS). Figure 1.2 illustrates the AGPS technology.

![Diagram of Assisted Global Positioning System (AGPS)](image)

**Figure 1.2 Assisted Global Positioning System (AGPS)**

This trend was accelerated by the U.S. Federal Communications Commission (FCC)’s 911-mandate making the location of a cell phone available to emergency call dispatchers. Canadian service providers have joined the movement and also started fitting AGPS chips in their new cell phones. The AGPS chips in mobile phones enable the service providers to determine the location of the phones within 15 meters. (Byon et al., 2006) The traditional method of locating a cell phone involves triangulating nearby cell phone towers that are typically a few kilometers
apart and the method is accurate to within 100 meters. Figure 1.3 illustrates how a cell phone can be located by the traditional triangulation method.

![Figure 1.3 Triangulation of cell phone towers for location determination](image)

The use of AGPS chips improves the accuracy of the location determination method and this has prompted transportation engineers to test the feasibility of using such GPS devices for traffic probes. However, there are some challenges for using cell phones as traffic probes. Unlike dedicated traffic probes with dedicated drivers, the cell phone users are not necessarily on the roads when they are being monitored. Even if they are on the road, it is not clear which mode of transportation they are in. Considering that we typically monitor driving conditions for private automobiles (also known as “auto” mode), if the cell phone owner is in a bus, a street car, on a bike or on foot, the data from such phones cannot directly be used as probes.

Since traffic monitoring generally refers to real-time monitoring of traffic conditions of private automobiles, determining the mode of transportation in a short duration of time (as opposed to having full GPS information throughout the entire trip over days) is a challenge. In addition, querying multiple phones can be costly depending on where you are. For example, wireless
communication tends to be cheaper and economically more feasible when the density of its users is high, as in parts of Asia. However, in North America, including Canada, it is more expensive to implement and operate similar services. It is a challenge to keep the number of querying or “pinging” the phones to a minimum while achieving the desired transportation mode detection rate. There are also political challenges in using cell phones as traffic probes. Some persons may not want to be tracked by traffic agencies and may demand privacy in this regard. A solution for this could be tracking only the phones of users who have agreed to be tracked, or tracking random phones anonymously. In fact, ironically, all cell phones are being and have been tracked at all times in the past because cell phones must be connected to nearby cell phone towers to be operational even for voice communications. If one knows the location of a tower a cell phone is connected, with a certain degree of error, the position of the phone can approximately be determined even without having an AGPS unit. Therefore, in summary, all cell phone users have been monitored whether they wanted or not and privacy issues existed from the very beginning of the use of cell phones. The only difference now from before is that the introduction of the AGPS units can provide more accurate location information. In addition, there is an issue regarding who gets the location information. For example, if a cell phone user signs up for the cell phone service with a service provider, the user would expect only the service provider knows the location information for the technical operation of the phone. In other words, the user is giving consent to the provider to access his location. If a government agency accesses such data that the user did not expect, there is an issue regarding implications of signing up for a cell phone service.

This thesis does not consider privacy issues and only focuses on the technical side of the potential feasibility of using mobile GPS devices, not limited to cell phones. In one scenario, traffic departments could at least replace their currently operational dedicated probes with mobile GPS devices on public service vehicles to avoid public complaints.

Even if it turns out that the use of mobile GPS devices as traffic probes is technically possible, there are still other traffic data sources that are operational already. At any one point in time during the rush hour, it is common to have multiple sensors monitoring the same section of the road network. The multiple sensors include loop detectors, probes, radio/TV broadcasts and
traffic department's web site information. Some of them provide quantitative measures while others provide qualitative and subjective opinions. For example, if a probe vehicle is traveling on the same link where a loop detector is already operational while listening to a radio broadcast that verbally comments on the traffic condition of the road that contains the same link, there are multiple sources of information that represent the traffic condition of the same section of the road. It is a challenge to combine or fuse the multiple data from multiple sources in a reasonable way to provide more reliable and accurate real-time traffic information.

In addition to the actual GPS data from mobile devices, with advancements in computing, traffic simulation packages can provide simulated GPS data on which the data fusion techniques can be applied. The actual GPS data always contain errors. Therefore, it is impossible to know the true speed information. However, from the simulated GPS data, it is at least conceptually possible to know the true traffic condition of the simulated environment by averaging the speed of all vehicles on the desired section of the road. With the true traffic condition, referred to as the "ground truth" throughout this thesis, it is possible to compare the performances of different multi-sensor-data-fusion scenarios against the ground truth.

Provided that the multiple information sources including loop detectors, probes, radio broadcasts and the traffic department's website are operational and the fused traffic data are ready and available online, it is possible to develop innovative route guidance methods with the aid of Geographic Information Systems (GIS). In the past, route guidance methods focused attention primarily on computing the shortest or the fastest routes. We live in a three-dimensional world where road gradients and the visibility on the roads may affect our route choices. By considering these new factors, one can develop a multi-criteria route guidance method that can compute the most scenic or the most level (the least hilly) route. In addition, by incorporating crime rates in regions (geospatial data), it is also possible to find the safest route.

1.2 Objective

The objective of this research is to develop a framework of a traffic monitoring and route
guidance system named as, GPS and GIS for Traffic Monitoring and Route Guidance (GISTMARG). GISTMARG consists of 3 main components and is proposed to be able to:

- Detect the mode of transportation of mobile GPS devices and filter for the devices in the “auto” mode in order to use them as traffic probes

- Fuse the mobile probe data with existing conventional information sources such as loop detectors, radio broadcasts and a traffic department’s website.

- Provide methods of incorporating 3 dimensional (3-D) and geospatial data for route guidance systems that find the most scenic, the most level or the safest route.

### 1.3 GISTMARG Overview

GISTMARG is the traffic monitoring and route guidance system and this thesis develops its framework by incorporating 3 main modules. They are mode detection, data fusion and route guidance modules. Mobile GPS devices, not limited to AGPS cell phones, travel with cell phone owners on various modes of transportation. Those mobile devices are queried through a central server or internet (if the mobile device is connected to the road-side internet) for their GPS-based location information. After the GPS data are received from the mobile devices, the mode detection module would try to estimate the most probable mode of transportation. Then, only the GPS data from auto mode are filtered for the traffic monitoring purposes as probes. The probe data are fed into the data fusion module along with data from other types of sensors. The data fusion module attempts to fuse the multiple sourced data in order to provide reasonable and reliable traffic information. The fused traffic information can be used as a basis on which the route guidance methods can find the fastest route based on the real time traffic information. With GIS, it is possible to consider non-traditional factors, in addition to distance and travel time, into the route guidance algorithms. The route guidance module of GISTMARG, with GIS, computes the most scenic, the most level and the safest routes when provided with a digital elevation model (DEM) layer and a safety related geospatial data layer. Figure 1.4 illustrates GISTMARG.
1.4 Research Challenges and Issues

There are different types of challenges in different stages of this research. Some of the more significant issues are identified in this section.

1.4.1 Detecting Mode from Relatively Small Amount of Data

In order to use mobile GPS traffic probes for real-time traffic monitoring, it is important to use the data from the mobile GPS devices that are in the “auto” mode selectively to provide more realistic information on traffic conditions. It is relatively easier to detect the mode of transportation when multiple days of GPS trip data are logged in a GPS data logger. In this case,
there is a sufficient amount of time to process the data off-line since it is not needed as an input to a real time application. In addition, the entire data set of multiple days can be accessed which aids the mode detection process. However, this method is not suitable for real-time traffic monitoring because the very reason for such a system is to harvest real-time traffic information and the mode of transportation needs to be determined as soon as possible. With access to relatively short traces of GPS trip data, it is a challenge to determine the mode of transportation correctly. Especially when the number of queries to the mobile devices can incur certain service fees associated with it (wireless access fees), it is essential to keep the number of queries (referred to as "pings" throughout this thesis) to the mobile device to the minimum while ensuring the desired mode detection rate is maintained.

### 1.4.2 High Cost Associated with Pinging

One can argue that the mode detection module is not feasible if the required pinging frequency of the cell phones is too high and expensive. However, this concern is simply not true. There is a cheaper alternative method of pinging the phones which is using the internet connection instead of using the central server of the local cell phone service provider. By subscribing to an unlimited internet access package of cell phone service providers, at a relatively low fixed cost ($20 ~ $50 per month as of May 2009 in Ontario, Canada), it is possible to implement the mode detection module without needing to query the phones via the wireless location technology (WLT) servers of the cell phone service providers. In this case, independent web server can query the phones at no additional pinging cost. One can also install a small application on the phone so that the phone is configured to automatically send its GPS information to the internet server via the internet connection at a certain set interval with no additional cost. One suggestion may be that traffic departments can subsidize the cost of the unlimited internet service for cell phone users who agree to install the location reporting application on their phones and participate in the traffic monitoring system as probes themselves. In return, they may surf the web for free. Even though this would solve the high pinging cost associated with the WLT of the service providers, optimizing the number of pinging is still needed because frequent pinging will reduce the battery life of the phone greatly which can be considered as one form of associated costs.
Figure 1.5 shows two alternative methods of retrieving GPS information from the phone.

Figure 1.5 Alternative methods of receiving GPS data

1.4.3 Lack of Location-based Services Available in Toronto

At the time of writing, a cell phone service provider in Toronto, Bell Mobility, provides LBS that have been enhanced with AGPS technologies embedded in their phones. However, the maximum querying frequency is arbitrarily set by the provider at once every 5 minutes. This is mainly due to the business plan of the company rather than the lack of the required technologies and hardware not being available. The company is currently focusing on providing location determining services for the application of asset management services. Construction companies who have multi-million dollar worth of heavy equipments dispatched to construction sites may
want to know at least where they are occasionally. In that case, typical sampling frequency would be once per day for example. If we are interested in detecting the mode of transportation hopefully by capturing the patterns in the vehicle's movement, we need a lot more frequent sampling. It is not possible at this point to query the GPS phones more often than once every 5 minutes if we were to use the WLT services of the service provider. In order to test the feasibility of using mobile GPS devices as probes, the ability to query phones at varying frequencies is essential.

This thesis approaches this challenge by using off-line GPS data loggers. The archived trip information from the GPS data loggers is sampled as if they are the mobile GPS devices. In this thesis, the raw GPS data produced by the GPS data loggers are assumed to be identical to the data produced by mobile AGPS cell phones.

1.4.4 Urban Canyon

Whenever GPS devices are used for research, it is important to recognize the weaknesses of using GPS devices in urban settings. High-rise buildings often block the GPS signals transmitted from the satellites, which can lead to poor accuracy in the observed location information and other derived values, such as speed and acceleration. Figure 1.6 shows how high-rise buildings can block the GPS signals. In addition, the buildings can act as mirrors and reflect the GPS signals before they reach the GPS receiver. This problem is known as "multipath effect". Figure 1.7 illustrates the effect. The resulting location can contain a few meters of errors. During congestion, vehicles often do not move much more than a few meters in 1 second. GPS devices are usually designed to compute speed values within the hardware based on the current location and the previous location 1 second ago. The mode detection module of GISTMARG depends heavily on the speed and acceleration values of the vehicle and the urban canyon effect can deteriorate the accuracy of the mode detection processes. This research intentionally includes the central business district (CBD) of Toronto as one of the study areas in order to test the feasibility of the mode detection module in the real-world like environment.
Figure 1.6 Urban canyon effect
1.4.5 Information Fusion of Multiple Information Sources

Different traffic sensors or information sources produce different kinds of information with different frequencies and variances. It is a challenge to fuse the different types of information into meaningful information. For example, even though both GPS device and loop detector produce quantitative speed values, they generally operate at different sampling frequencies. In addition, other information sources provide qualitative information such as colour-coded information (traffic departments’ websites) and verbal information (radio broadcasts). Qualitative information can be transformed into quantitative information in order to fuse them with other quantitative data. It is an academic challenge to design the conversion process.
1.4.6 Lack of Detailed Elevation Models

The newly introduced route guidance method in this thesis recognizes the fact that the roads are 3 dimensional in nature. In the past, we were accustomed to think that the roads are flat on maps or we were forced to think so due to the lack of computing power with which computer applications had processing difficulties even with 2-D maps. However, in real-life, drivers are always affected by the 3 dimensional factors. Currently, the available elevation model for Toronto (Fortin, 2008) only has elevation information for the ground surfaces. The buildings on top of the ground surfaces are not considered in building DEM models, which obviously affects visibility issues. In order to use the visibility factor in transportation applications, one needs to have detailed elevation models of the buildings in the city in addition to the ground surface models. Many leading software companies including Google Inc. are starting to realize the need for 3-D maps and have started collecting surrounding 3-D images of the road links. 3-D models of monuments and other major buildings are created and are available in their 3-D maps as of September, 2009. If the location and height information of road networks and all buildings are digitized into GIS database, transportation engineers can conduct extensive research with visibility related issues. Some of those research areas include analyzing the effect of the visibility on the route choice, mode choice and car following models.

1.4.7 Subjective Measure of Safety and Scenic Quality

The new route guidance method introduced in this thesis recognizes that people do not always choose the shortest or the fastest routes for their trips. For example, some tourists may prefer the most scenic route while others may prefer the safest route in the new cities they visit. However, the measure of the scenic quality or the safety of certain regions is usually subjective in nature. It is a challenge to quantify and integrate measures of the scenic ratings and the safety ratings into the route guidance algorithm.
1.5 Overview of Proposed Approach

GISTMARG consists of 3 main components; mode detection, data fusion and route guidance modules. This thesis applies unique approaches to each module and the general overview of those approaches is presented in this section.

1.5.1 Neural Network (NN) for Mode Detection

Estimating the mode of transportation from relatively short traces of GPS trip data is challenging especially when the results are needed as soon as possible. Artificial intelligence (AI) is one method for solving the classification problem. In this research, neural networks (NNs) are used for their superior non-linear relationship handling abilities. Some of the important factors such as speed, acceleration and their accuracies related factors are fed into the NNs as input. Then the NNs attempt to classify the mode of transportation in efforts to filter for the “auto” mode to use them as traffic probes.

1.5.2 Kalman Filters for Data Fusion and Prediction

Standard KFs (KF) are capable of capturing the data trend and they can estimate future values based on the readings from multiple sensors. In other words, they can simultaneously fuse multiple sensors and also estimate future trends. The main output of any traffic monitoring system is the prediction of near-future traffic conditions. KFs can be applied to a traffic monitoring system to fuse multiple traffic information sources and estimate near-future traffic conditions. KFs generally update their predictions based on multiple sensors at each time step. However, in the case of traffic monitoring, often, only some of the sensors are active and others are waiting for their next measurements with their unique sampling frequency. A modified type of standard KFs known as SCAAT: Single-Constraint-At-A-Time (Welch, 1996) KF is used in this research. SCAAT filters update their estimations as soon as any one of multiple sensors are active instead of waiting until all sensors are active as in the case of standard KFs.
1.5.3 3-D GIS for Route Guidance

Until now, most route guidance algorithms have used 2-dimensional road networks with link costs (weights) based on the distance. By replacing the costs with real-time travel times, it is also possible to find the fastest route. With the application of advanced 3-D features of the GIS, it is possible to integrate the 3-dimensional nature of the roads into the route guidance algorithms. Recognizing the fact that roads are in 3-D space, opens the door to new research opportunities. A DEM of the Greater Toronto Area (GTA) is a GIS layer that contains the elevation model of the study regions in Toronto. With the DEM layer processed in parallel with the Toronto streets layer in the GIS, visible areas from particular points on the roads can be determined. In addition, road gradients can also be determined. With the 3-D GIS layers, the most scenic or the most level (the least hilly) route can be found for route guidance applications. In addition to the 3-D GIS layers, as an illustrative example of incorporating new geospatial factors, this thesis converts crimes rates into quantifiable costs for the route guidance method.

1.6 Thesis Outline

This thesis is organized as follows. Chapter 1 briefly introduces the research background and states the research objectives. Major research challenges are identified and briefly discussed, and main approaches used throughout the research are overviewed.

Chapter 2 introduces various background information that is essential for understanding this research.

Chapter 3 reviews some of the related work in the past on the topics of traffic monitoring, data fusion and route guidance systems.

Chapter 4 focuses on the transportation mode detection module of GISTMARG.
Chapter 5 discusses the data fusion module of GISTMARG.

Chapter 6 develops methods of incorporating scenic view, road slope and crime rates into route guidance systems.

Chapter 7 summarizes this thesis and presents final recommendations.
Chapter 2
Background

2 Background

In Chapter 2, required background information for understanding this thesis are presented for audiences who are not familiar with some of the topics discussed. The introduction to each topic is kept at minimum whenever possible so that there is just enough technical information for understanding this thesis. More detailed descriptions, extensions to the information provided in this chapter, are presented in later chapters as needed.

2.1 Moving Forward From GISTT

GISTMARG is inspired by GISTT (GPS and GIS for Travel Time Surveys) (Byon, 2005) which was my master’s thesis work. In this section, GISTT is briefly reviewed in comparison to GISTMARG.

2.1.1 GISTT (GPS-GIS Integrated System for Travel Time Surveys)

Most transportation agencies collect travel times along various sections of the transportation network on a periodic basis. Travel time is an important performance indicator because it is easily understood by the public and politicians, easily applicable to all modes of transportation, and can effectively be utilized for transportation modeling and forecasting. In addition, travel time variability provides an accurate measure of the reliability of a transportation system and can be utilized for emergency planning and estimating response time.
Conventional methods for auto travel time collection typically require 2 technicians in a probe vehicle: one driving and the other manually recording the location and time of the vehicle as it passes predetermined checkpoints. Such techniques are labour intensive, costly and prone to errors. Recent advances in GPS and GIS technologies present opportunities to implement methods that can replace the conventional travel time collection method.

The GISTT that combines the GPS and GIS technologies is a proposed solution. GISTT collects travel time data along predetermined routes in static and dynamic modes. The static mode refers to the case where a previously dispatched vehicle equipped with a GPS data logger is used to collect travel times along a pre-determined route. The dynamic mode refers to the real time monitoring of travel times on road links using a GPS equipped probe vehicle that communicates with a main server using the wireless internet.

Initially, both static and dynamic GISTT codes were implemented by customizing existing GIS software, ArcGIS via Visual Basic for Applications (VBA) language. The main drawback of the customization method is that the user interface is unnecessarily crowded and the loading of trip data is not user friendly and is time consuming. In order to address these problems, stand-alone versions of static and dynamic GISTT codes were developed. The purpose of developing the stand-alone versions was to create simpler user interfaces with GIS functions that are relevant to travel time collection processes only.

It was found that the GISTT produces similar travel time predictions when compared with classical travel time collection methods in all areas (arterial roads, highways etc.) of the GTA except the downtown CBD region. The poorer performance in the CBD region was due to the GPS signal blockage and multipath effects caused by high-rise buildings.

The development of GISTT makes the travel time collection process less labour intensive and reduces associated costs. GISTT estimates not only the link travel times but also provides the second-by-second speed information of the vehicle throughout the whole trip. Frequent uses of static GISTT in travel time collection processes would enhance the wealth of travel time information and can help better estimate link travel times at particular time slots. The dynamic
GISTT can be used for monitoring current real-time traffic conditions by providing more accurate route guidance for drivers.

The dynamic GISTT was tested with a single probe vehicle (even though it can theoretically handle multiple probes). Multiple probe vehicles can monitor the traffic conditions on multiple links more effectively. GISTMARG is inspired by the needs of operating multiple probes in real-time.

2.1.2 GISTT vs. GISTMARG

This PhD thesis work, GISTMARG, is inspired by GISTT. GISTT is a travel time analysis tool that analyzes raw GPS trip data. It outputs travel times on pre-determined routes at desired time slots. All GPS trip data are assumed to be in the “auto” mode and are treated as probes by default. GISTT can theoretically handle any number of multiple probes as long as the required data communication bandwidth is available. Both real-time and off-line trip data can be used to estimate current and historic travel times on desired routes.

GISTMARG enhances and extends the capabilities of GISTT. The mode of transportation is not assumed to be the “auto” mode by default any more. Instead, GISTMARG attempts to detect the mode of transportation from the relatively short trace of the most recent real-time GPS data. It also deals with the fusion of data from multiple sources such as loop detectors in addition to probe vehicles alone. GISTMARG also suggests new approaches for the route guidance systems that are based on various criteria that fully utilize the powerful capabilities of the 3-D GIS platform.
Table 2.1 Comparison of GISTT and GISTMARG

<table>
<thead>
<tr>
<th></th>
<th>Static GISTT</th>
<th>GISTMARG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS Technology</td>
<td>GPS data logger</td>
<td>AGPS cell phones</td>
</tr>
<tr>
<td>Sampling Frequency</td>
<td>Once every 1 second</td>
<td>varies</td>
</tr>
<tr>
<td>On-line or Off-line</td>
<td>Off-line. Post-processing</td>
<td>On-line. Real-time</td>
</tr>
<tr>
<td>Number of Probes Tested</td>
<td>One</td>
<td>Multiple</td>
</tr>
<tr>
<td>Mode Detection</td>
<td>None (Assumes &quot;auto&quot; mode)</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Fusion</td>
<td>None</td>
<td>Yes</td>
</tr>
<tr>
<td>Monitoring Duration</td>
<td>Multiple days</td>
<td>A few minutes</td>
</tr>
</tbody>
</table>

Components of GISTMARG contribute to the following 3 research categories; traffic monitoring, data fusion of multiple sources and route guidance.

2.2 Mode Detection

Detecting the mode of transportation from relatively short traces of GPS data requires mapping the GPS data trail or string onto the corresponding mode of transportation. This task can be viewed as a classification task. Neural networks, from artificial intelligence studies, as described in Section 2.2.8, are superior to classical classification methods such as the fuzzy logic, and hence are chosen for the mode detection task in this research.
2.2.1 Neural and Adaptive Systems (NAS)

“Neural and adaptive systems (NAS) are used in many important engineering applications, such as signal enhancement, noise cancellation, classification of input patterns, system identification, prediction and control.” (Principe et al. 2000) NAS finds input-output relationship by adapting to a dataset in order to set their parameter values that minimize its cost functions. Figure 2.1 shows the general set-up of an adaptive system. Input $x$ is fed into the adaptive system that consists of weight values, $w$. With initial weight $w_0$, the system estimates the current state, $y_0$. Then the measured current state is compared with the initial estimation. The difference between the two, $d$, is then carried into a training module where the weight values of the adaptive system are adjusted in efforts to produce less error. Over time, the adaptive system would estimate with higher precision and also track the changes in the system state. In traffic monitoring, such system can first try to estimate the current or near-future traffic conditions and can detect changes in the traffic conditions over time as well.

Figure 2.1 Adaptive system's design (Principe et al. 2000)
2.2.2 Linear Adaptive System

As a starting point for illustrating an adaptive system, a linear adaptive system is considered first here. Linear models represent relationships between input and output variables by fitting a straight line that minimizes the deviation of data from the fitted line. Assuming \( x \) is the independent variable and \( d \) is the dependent variable, if it is suspected that there is a linear relationship between the two variables, the relationship between the two variables can be written as,

\[
d \approx y_i = wx + b \quad \text{(Equation 2.1)}
\]

where \( y_i \) is the estimation of \( d \),
the slope \( w \) and
the bias \( b \) are the parameters of the relationship.

In general, linear models have a form of the following for one sample data pair:

\[
d_i = y_i + \varepsilon_i = (wx_i + b) + \varepsilon_i \quad \text{(Equation 2.2)}
\]

where \( i \) is the sample number,
\( d_i \) is the dependent variable,
\( x_i \) is the independent variable and \( w \) is the slope of the line,
\( b \) is the \( d \) axis intercept or bias and
\( \varepsilon_i \) is the error that is added to \( y_i \).

The linear model can be represented as a linear processing element (LPE). Figure 2.2 shows the LPE, and it is built from two multipliers and one adder. The multiplier \( w \) scales the input, and the multiplier \( b \) is a bias.
Figure 2.2 Linear processing element

When fitting a line to the data set or finding the values of $w$ and $b$, the least square method can be used as a criterion to minimize the difference between the estimated and observed values. The average sum of square errors $J$, is also known as the mean square error (MSE).

$$J = \frac{1}{2N} \sum_{i=1}^{N} \varepsilon_i^2 = \frac{1}{2N} \sum_{i=1}^{N} (d_i - w x_i)^2 \quad \text{(Equation 2.3)}$$

where $N$ is the number of observations and

$\varepsilon_i$ is the difference between the estimation and observed value of the dependent variable.

In order to find the $w$ and $b$ that minimizes $J$, the partial derivatives of $J$ with respect to $b$ and $w$ can be set to zero simultaneously and it yields analytical formulae for $b$ and $w$.

$$\frac{\partial J}{\partial b} = 0 \quad \text{(Equation 2.4)}$$
\[
\frac{\partial J}{\partial w} = 0 \quad \text{(Equation 2.5)}
\]

\[
b = \frac{\Sigma x_i^2 \Sigma d_i - \Sigma x_i \Sigma x_i d_i}{N \Sigma (x_i - \bar{x})^2} \quad \text{(Equation 2.6)}
\]

\[
w = \frac{\Sigma (x_i - \bar{x})(d_i - \bar{d})}{\Sigma (x_i - \bar{x})^2} \quad \text{(Equation 2.7)}
\]

where \(\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}\) and \(N\) is the number of observations.

Linear models can be converted into an adaptive system shown in Figure 2.3 where the system adjusts its parameters as input and output data pairs are fed into the system while the error value is fed back to adjust parameter values similar to the illustration in Figure 2.1.

**Figure 2.3 Linear adaptive system**

The error criterion \(J\) varies with varying values of \(w\) and \(b\) of the linear system. Assuming the bias \(b = 0\), which does not affect the process of finding the \(w\) that gives the minimum \(J\), the optimal value of \(w^*\) results in the minimum value of \(J, J_{\text{min}}\). The resulting graph in Figure 2.4 is also known as a performance graph.
The gradient of the performance graph is a vector that always points toward the direction of maximum J change and with a magnitude equal to the slope of the tangent of the performance graph. If b was not set to zero, the performance surface will be 3-D with additional dimension of b orthogonal to this page resulting in 3-D surface that resembles a valley with a local minimum point where the slope is zero both in directions of w and b dimensions.

The gradient information can aid the adaptive system in finding the coefficient value w that results in the minimum value of J. The procedure can search for the minimum J by moving in the direction opposite to the gradient. The procedure can start with arbitrary initial weight w(0) and computes the performance surface values near w(0). It then, modifies the w(0) proportional to the gradient value in the negative direction of the gradient and updates w to w(1).

\[ w(k + 1) = w(k) - \eta \nabla J(k) \quad \text{(Equation 2.11)} \]

Figure 2.4 Performance graph (Principe et al. 2000)
Widrow proposed a faster algorithm to estimate the gradient that uses the instantaneous value as the estimator for the true quantity. (Principe et al., 2002)

\[
\nabla J(k) = \frac{\partial}{\partial w(k)} J = \frac{\partial}{\partial w(k)} \frac{1}{2N} \sum \varepsilon_i^2 \approx \frac{1}{2 \frac{\partial}{\partial w(k)}} (\varepsilon^2(k)) = -\varepsilon(k)x(k) \tag{Equation 2.12}
\]

The algorithm is known as the least mean square (LMS) algorithm and it enables the gradient finding procedure faster and more practical for implementation.

### 2.2.3 Regression for Multiple Variables

If there are two weight values in a particular adaptive system, the gradient is formally defined in terms of partial derivatives of a function \(f(w_1, w_2)\).
\[ \text{grad } f(w_1, w_2) = \nabla f(w_1, w_2) = f_{w_1}(w_1, w_2)\mu_{w_1} + f_{w_2}(w_1, w_2)\mu_{w_2} \quad \text{(Equation 2.8)} \]

where \( \mu_{w_1} \) and \( \mu_{w_2} \) are the unit vectors along \( w_1 \) and \( w_2 \).

\( f_{w_1} \) and \( f_{w_2} \) are the partial derivatives of \( f \) along the \( w_1 \) and \( w_2 \) directions,

\[ f_{w_1} = \frac{\partial f(w_1, w_2)}{\partial w_1}, \quad \text{(Equation 2.9)} \]

\[ f_{w_2} = \frac{\partial f(w_1, w_2)}{\partial w_2}. \quad \text{(Equation 2.10)} \]

Figure 2.6 shows the linear processing element for the regression system for two-input weight variables.

---

**Figure 2.6 The 2-D performance surface (Principe et al. 2000)**

In the case of having more than two input variables, for a particular input-output pair of observed data, the error can be defined as in the following.

\[ \varepsilon_i = d_i - (b + \sum_{k=1}^{D} w_k x_{ik}) = d_i - \sum_{k=0}^{D} w_k x_{ik} \quad \text{(Equation 2.13)} \]
where \( i = 1 \ldots N \) with \( w_0 = b \) and \( x_{i0} = 1 \)

The mean square method yields,

\[
J = \frac{1}{2N} \sum_i (d_i - \sum_{k=0}^{D} w_k x_{ik})^2 \quad \text{(Equation 2.14)}
\]

By taking the derivatives of \( J \) with respect to the unknowns and equating them to zeros, \( D + 1 \) equations in \( D + 1 \) unknowns are generated.

\[
\sum_i x_{ij} d_i = \sum_{k=0}^{D} w_k \sum_i x_{ik} x_{ij} \quad \text{(Equation 2.15)}
\]

where \( j = 0,1,\ldots, D \)

Figure 2.7 shows the linear processing element for the regression system for multiple input variables.

\[
\begin{align*}
x_{1i} & \quad w_1 \quad d_i \\
x_{2i} & \quad w_2 \\
x_D i & \quad w_D \\
+1 & \quad b \\
\end{align*}
\]

\[
\sum \quad y_i \quad + \quad \varepsilon_i
\]

**Figure 2.7** Regression for multiple variables
The similar gradient techniques can be used by defining $\nabla J$, which is a vector with $D + 1$ components.

$$\nabla J = \left[ \frac{\partial J}{\partial w_0}, ..., \frac{\partial J}{\partial w_D} \right]^T \quad \text{(Equation 2.16)}$$

and, the vector of weights,

$$w(k+1) = w(k) - \eta \nabla J(k) \quad \text{(Equation 2.17)}$$

where $w(k) = (w_0(k), ..., w_D(k))^T$

NAS is basically an iterative regression technique that searches for the optimal set of parameters that minimizes the error between the output and input pairs. NAS navigates the n-dimensional valley of performance surface of $J$ with the vector of $W^*(W_0, ..., W_n)$ that results in the local minimal value of $J$. The $\eta$ is a step size of the search method using the gradient. Higher value of $\eta$ results in faster search but has higher chance of not converging at all.

2.2.4 Pattern Recognizer

The regression techniques can be applied to pattern recognition applications with some modifications. Unlike the regression method that estimates function values, pattern recognition method instead, classifies the output into a finite number of discrete classes. All output in the same class shares the identical label. Class assignments are mutually exclusive and therefore the classifier needs a nonlinear mechanism such as an all-or-nothing switch. Even though both regression and classifier maps input values with output values, the nature of the objective in each case is fundamentally different. “We can nevertheless use the machinery utilized in linear regression … as pieces to build pattern classifiers” (Principe et al. 2000). LPE can be extended with a threshold to act as a decision device and organization of such devices in proper manner.
can act as a classifier in a similar fashion to biological classifier humans have which is further explained in the following sections.

One of the major challenges in designing a classifier is the determination of boundaries above and below which different classes are assigned. Optimal classifier can be thought of as a classifier that chooses the class $c_i$ that maximizes the a posteriori probability $P(c_i | x)$ that the given sample $x$ belongs to the class.

$$P(c_i | x) > P(c_j | x) \text{ for all } j \neq i \quad \text{(Equation 2.18)}$$

Bayes’ rule suggests that the posteriori probability can be computed from $P(c_i)$, the prior probability of the classes, multiplied by $p(x | c_i)$, the likelihood that the data $x$ was produced by class $c_i$ and normalized by the probability of $P(x)$.

$$P(c_i | x) = \frac{p(x | c_i)P(c_i)}{P(x)} \quad \text{(Equation 2.19)}$$

$P(c_i)$ can be estimated from the data set. With an assumption that the a posteriori probability is normally distributed with input axis in the case of 1 dimensional input space, the classification boundary can be determined by equating the two posteriori probabilities for two classes.

$$\frac{p(x | c_1)P(c_1)}{P(x)} = \frac{p(x | c_2)P(c_2)}{P(x)} \quad \text{(Equation 2.20)}$$

Figure 2.8 shows how the classification boundary for two different classes of $c_1$ and $c_2$ are found.
In the case of multiple input variables, there will be a intersecting contour lines between “mountains” where there are equal chances that a given input variables belong to both mountains sharing the same contour line. Assuming there are \( N \) measurements \( x_1, x_2, \ldots, x_n \), where each measurement \( x_k \) is a vector with \( D \) components, \( X_k = [x_{k1}, x_{k2}, \ldots, x_{kD}] \) and each measurement can be thought of as a point in the \( D \)-dimensional pattern space. With the class assignment based on the Bayes’ rule which involves the comparison of likelihoods scaled by the corresponding a priori probability, the measurements or input data can be classified into a class.

Each scaled likelihood can be thought of as a "discriminent" function \( g(x) \) that assigns a score to every point in the input space. Each class has its own scoring function and the intersections of those functions are called decision surfaces and they can be thought of as the contours mentioned earlier. Therefore, the measurement \( x_k \) will be assigned to a class \( i \) if \( g_i(x_k) > g_j(x_k) \) for all \( j \neq i \).
Figure 2.9 Discriminent functions and the decision surface (Principe et al. 2000)

Figure 2.10 shows conceptual overview of classification set-up with discriminant functions. It is important to note that, hidden layers of NNs described later implicitly include the discriminent functions in them.
2.2.5 Artificial Intelligence

AI is the intelligence of machines and is a branch in computer science. It is often defined as the study and design of intelligent agents. An agent is a system that perceives the environment and makes decisions to maximize the chances of success. In the case of transportation mode detection, the AI is supposed to observe given GPS data inputs and attempt to correctly classify the mode of transportation the GPS device is in. There is no established theory that defines exactly what AI is. Some of the more common AI techniques include artificial neural networks (ANN)s (also known as NNs) and methods based on fuzzy logic.

2.2.6 Artificial Neural Networks (ANNs)

One branch of artificial intelligence is known as an artificial NN (ANN). ANNs are inspired by the way biological nervous system such as a human brain processes information. The human brain is known to be a superior pattern recognizer. By imitating the human brain, ANNs try to detect patterns rather than memorizing the input and output pairs. An ANN is composed of a number of interconnected processing elements also known as neurons. When neurons receive certain triggering signals in unique patterns, they send signals to the next interconnected neurons. There can be multiple layers of neurons inside the NNs. Similar to how the human brain works, an ANN needs to learn first to be operational in applications. Human brains learn new information by biologically adjusting synaptic connections between the neurons. Figure 2.11 shows the biological neuron. An ANN instead adjusts the connection weight values between the processing elements as cycles of input-output pairs of training data are passed through the ANN during the training session.
NNs are capable of harvesting meaningful information from imprecise and complicated input data. NNs can detect patterns or trends that other computer algorithms cannot easily find. Even though NNs mimic human brains, they can detect patterns that most human brains cannot. Once an NN is exposed to a certain category of input-output pairs of information, it trains itself and becomes an expert in the sense that it can now handle similar categories of input information and acts as an expert and predicts the unseen output based on its previous training or learning sessions. This feature of learning based on given initial data is also known as an “adaptive learning”.

Figure 2.11 Biological neuron in a human brain
NNs tackle given problems quite differently from conventional computers. Conventional computers follow a set of instructions in order to solve a problem. If any one of the steps is not clearly specified in the form of a programming code, a conventional computer fails at solving the problem. This means that the conventional computers can only solve problems that have been solved before and fully understood by humans beforehand. NNs instead approach problems by learning example scenarios. NNs have more freedom than conventional programming codes in the sense that they can freely adjust their connection weight values as they sense the example data sets in efforts to blindly match the input and output data pairs numerically. Because of this non-instructed or “blind” way of matching input-output pairs, NNs can result in unexpected and non-intuitive outcomes. However, these non-intuitive outcomes ironically are the strongest feature of NNs because they can detect patterns that human brains or human-instructed machines cannot.

As labeled in Figure 2.11, a neuron collects signals from other neurons through a structure called dendrites. Then the neuron transmits electrical signal through a pipe known as an axon, which splits into multiple branches. At the end of each branch, the electrical signal from the axon excites synapse that in turn sends the signal to other neurons. When a neuron receives a set of input signals from multiple other neurons that is sufficiently large compared to its neutral state, it generates or “fires” electrical signal down the axon.

ANNs are derived from the design of neurons and their interconnections in the human brain. Typically, an artificial neuron has input connections arriving from other neurons and has output connection(s) to other neurons. There are two modes of operation for the neuron; the training mode and operation mode. In the training mode, the neuron is trained to fire or excite for certain input-output pairs. In the operation mode, when input patterns are similar to the input patterns of the trained input-output pairs, the neuron fires or excites and sends out signals to other neurons through the output connection(s). When the products of input variables and weights are summed and fed into a threshold nonlinearity function known as "sigmoidal" functions, the neuron is called the McCulloch-Pitts (M-P) processing element(PE). A special type of sigmoidal function
known as a hyperbolic tangent function can produce smooth ranges of output values near the transition region. The smoother transition provides better basis for computing the instantaneous rate of change of the output value for determining the optimal weight values. The two types of sigmoidal functions are shown in the following and Figure 2.12 shows the M-P PE with the hyperbolic tangent sigmoidal function.

Threshold Function: \[ f(\text{net}) = \begin{cases} 1 & \text{if } \sum_i w_i x_i > 0 \\ -1 & \text{if } \sum_i w_i x_i < 0 \end{cases} \] (Equation 2.21)

Hyperbolic Tangent Function: \[ f(\text{net}) = \tanh(\alpha \text{net}) \] (Equation 2.22)

**Figure 2.12** McCulloch-Pitts processing element (artificial neuron)

The mathematical formulation of M-P PE is presented in the following. Let us define:

\[ \sum_{i=1}^{n} x_i w_i = \text{net} \text{ and } \frac{\partial y}{\partial \text{net}} = f'(\text{net}), \] (Equation 2.23)

And \[ \frac{\partial \text{net}}{\partial w_i} = \frac{\partial}{\partial w_i} (\sum_k w_k x_k) = 0 + \cdots + \frac{\partial w_i x_i}{\partial w_i} + 0 + \cdots = x_i. \] (Equation 2.24)
By substituting the equations into the following chain rule based equation results in a rather simple instantaneous relationship between the output of the M-P PE and one of the weights.

\[
\frac{\partial y}{\partial w_i} = \frac{\partial y}{\partial \text{net}} \frac{\partial \text{net}}{\partial w_i} = f'(\text{net})x_i \quad \text{(Equation 2.25)}
\]

![Diagram of partial derivatives in an M-P processing element.](image)

**Figure 2.13 Partial derivatives in an M-P processing element**

The weight updating mechanism can be presented in the following format.

\[
w(n+1) = w(n) + \eta(d(n) - y(n))x(n) \quad \text{(Equation 2.26)}
\]

\[
w(n+1) = w(n) + \eta \varepsilon(n)x(n) \quad \text{(Equation 2.27)}
\]
where \( w(n+1) \) is the newly adjusted weight, \( w(n) \) is the old weight, \( \eta \) is the learning rate, \( d(n) \) is the desired output, \( y(n) \) is the actual estimated output, and \( x(n) \) is the input value.

The chain rule from calculus can be used to calculate partial derivatives of a variable with respect to other variables. If \( y = f(x) \), then \( \frac{\partial y}{\partial x} = \frac{\partial y}{\partial f} \cdot \frac{\partial f}{\partial x} \). (Equation 2.28)

The LMS algorithm based cost function \( J \) is presented in Section 2.2.2. By taking the derivative of \( J \) with respect to \( w \) (the weight variable) and equating it to zero, the minimum \( J \) can be computed.

\[
J = \frac{1}{2} \sum_p (d_p - y_p)^2 = \sum_p J_p \quad \text{(Equation 2.29)}
\]

Where \( J_p \) is \( J \) of the \( p \)th sample and \( y_p = wx_p \).

\[
\frac{\partial J_p}{\partial w} = \frac{\partial J}{\partial y_p} \frac{\partial y_p}{\partial w} = -(d_p - y_p)x_p = -\varepsilon_p x_p \quad \text{(Equation 2.30)}
\]

Based on the illustration shown in Figure 2.5, combining Equation 2.11 and Equation 2.12 results in,

\[
\Delta w_p = -\eta \frac{\partial J}{\partial w} = \eta \varepsilon_p x_p. \quad \text{(Equation 2.31)}
\]

This LMS concept that uses the chain rule to update the weight of one neuron can be extended to the M-P processing element which is a nonlinear system.

\[
\frac{\partial y}{\partial w_i} = \frac{\partial y}{\partial \text{net}} \frac{\partial \text{net}}{\partial w_i} = f'(\text{net})x_i \quad \text{(Equation 2.32)}
\]

As long as the nonlinearity function is differentiable, the change in actual output with respect to the weight can be computed from the input value. The rate of change of \( J \) with respect to \( w \) can now be rewritten as in the following.

\[
\frac{\partial J}{\partial w_i} = \frac{\partial J}{\partial y_p} \frac{\partial y_p}{\partial \text{net}_p} \frac{\partial \text{net}_p}{\partial w_i} = -(d_p - y_p)f'(\text{net}_p)x_{ip} = -\varepsilon_p f'(\text{net}_p)x_{ip} \quad \text{(Equation 2.33)}
\]

\[
w_i(n + 1) = w_i(n) + \eta \varepsilon_p(n)x_{ip}(n)f'(\text{net}_p(n)) \quad \text{(Equation 2.34)}
\]
Therefore, the mechanism for how M-P processing element can update its weights is shown. The weight updating mechanism can further be improved by adapting what is known as “momentum learning”. The magnitude of the weight update in the previous step affects the weight updating process in the current step. This learning method can help the algorithm avoid local minima.

\[ w_i(n + 1) = w_i(n) + \eta \varepsilon_p(n)x_{ip}(n)f'(\text{net}_p(n)) + \alpha(w_{ij}(n) - w_{ij}(n - 1)) \]  
(Equation 2.35)

A perceptron is a pattern recognition machine that uses M-P PEs. Multiple inputs to the perceptron are fully connected to an output layer with multiple M-P PEs. Each input \( x_j \) is multiplied by an adjustable weight \( w_{ij} \) before being fed into the \( i^{th} \) processing element in the output layer.

\[ y_i = f(\text{net}_i) = f(\sum_j w_{ij}x_j + b_i) \]  
(Equation 2.36)
By having m number of M-P PE in the output layer, each class can now have its own unique discriminant function in a D-dimensional space. The formulation of J is modified to include multiple output classes.

\[
J = \frac{1}{2N} \sum_{p=1}^{N} \sum_{i=1}^{M} \varepsilon_{pi}^2 \quad \text{(Equation 2.37)}
\]

where p is the index for patterns and 

i is a counter for the output PEs.

\[
\frac{\partial J}{\partial w_{ij}} = \frac{\partial J}{\partial y_{ip}} \frac{\partial y_{ip}}{\partial \text{net}_{ip}} \frac{\partial \text{net}_{ip}}{\partial w_{ij}} = -(d_{ip} - y_{ip})f'(\text{net}_{ip})x_{ip} = -\varepsilon_{ip}f'(\text{net}_{ip})x_{ip} \quad \text{(Equation 2.38)}
\]

By defining a local error \( \delta_i \) for the ith PE as

\[
\delta_{ip} = \frac{\partial J}{\partial y_{ip}} f'(\text{net}_{ip}), \quad \text{(Equation 2.39)}
\]

the weight updating equation including the delta rule and the momentum learning becomes

\[
w_{ij}(n + 1) = w_{ij}(n) - \eta \frac{\partial J}{\partial w_{ij}} = w_{ij}(n) + \eta \delta_{ip}x_{ip} + \alpha(w_{ij}(n) - w_{ij}(n - 1)). \quad \text{(Equation 2.40)}
\]

The mechanism presented in the perceptron can be extended with an additional hidden layer with PEs. The extended system is known as one-hidden-layer multilayer perceptrons (MLP). By introducing an extra layer of mapping, the relationship between the final classification and inputs are of the form

\[
y = f(\sum f(\Sigma \cdot)) \quad \text{(Equation 2.41)}
\]

The new mapping is a nested composition of nonlinearities. The resulting map provides
additional flexibility by introducing more adjustable weights. One-hidden-layer MLP with M-P PEs is a universal map. (Principe et al. 2000) Provided there are enough number of hidden-layer PEs, it can classify the input patterns into appropriate classes. Theoretically speaking, if there are infinite number of hidden PEs, the system can directly memorize the entire set of data as if it was a database of the training data resulting in 100% of accuracy when tested with the training data. However, the main purpose of such system is to perform well against unseen new data, it is a challenge to use the right number of PEs in order to allow sufficient flexibility and avoid data memorization.

![Diagram of one hidden layer perceptrons]

**Figure 2.15 One hidden layer perceptrons**

The training of the one-hidden-layer MLP uses an algorithm known as the back-propagation algorithm. The method is a form of error-correction based learning which requires the desired target to be known. In other words, it is a supervised learning method that uses known input-output training data. The following figure shows the detail of a hidden-layer network.
The weight updating using the back-propagation algorithm, basically derives the error from the output layer and propagates the error backwards while the sensitivity is computed by using the chain rule. The formulation is

\[ w_{ij}(n + 1) = w_{ij}(n) + \eta f'(net_i(n)) \left( \sum_k e_k(n)f'(net_k(n))w_{ki}(n) \right)y_j(n) \]  

(Equation 2.42)

One-hidden-layer MLP is one form of a network of neurons known as a Feed-Forward (FF) ANN. Feed-forward ANNs are designed so that input signals travel one way only from input to output. There are no feedback connections so that outputs of any layer do not affect the same layer. In this research, feed-forward networks are used by default because a part of the objective of this research is to test the feasibility of using NNs for detecting transportation modes. If the task can be achieved by one of the simplest architecture of NNs, there is always a room for improvements with more complicated organizations of networks.

The most common One-hidden-layer MLP consists of 3 layers; input layer, hidden layer and output layer. The input layer has only input neurons that represent the input data. The hidden layer has processing neurons that are fired or excited when certain patterns of input data are detected. The output layer contains output neuron(s) that either estimate the numerical output value(s) or classify the output class. In this research, the output layer classifies the input data into modes of transportation based on GPS data inputs.
In this research, a supervised learning method is used. Supervised learning refers to the case where the training data includes both input and known target values. The ANN is forced to learn certain combinations of GPS data resulting in certain transportation modes. In order to train the ANN, a back-propagation algorithm is used. Basically, the algorithm looks for differences between the actual output and the estimated output throughout the training process and propagates the errors backward from output layer to hidden layer to input layer. As the errors are propagated backward, the weights on the connections are modified so that the connections between the neurons that are more significant are given more weights.

2.2.7 Neurosolutions

Neurosolutions is a software package that specializes in building adaptive NN systems. Its object-oriented approach provides user friendly interfaces. M-P PEs of multiple layers can be built on a “breadboard” to form a “circuit” with inputs and outputs. Then the built NN is tested with training data. The number of neurons per each layer, number of input/output variables can be modified in its graphical user interface. The one-hidden-layer perceptrons are implemented in this research. Table 2.2 shows the most common components used in this thesis. Figure 2.17 shows the breadboard layout of the one-hidden-layer perceptrons.
Table 2.2 Common components of Neurosolutions

(Neurodimension Inc., 2008)

<table>
<thead>
<tr>
<th>Icon</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Axon</td>
<td>Layer of processing elements.</td>
</tr>
<tr>
<td><img src="image1.png" alt="Icon" /></td>
<td>TanhAxon</td>
<td>Hyperbolic nonlinear transfer function. M-P PE can be built by attaching TanhAxon after Axon.</td>
</tr>
<tr>
<td><img src="image2.png" alt="Icon" /></td>
<td>FullSynapse</td>
<td>Connects Axons from the left to the right.</td>
</tr>
<tr>
<td><img src="image3.png" alt="Icon" /></td>
<td>L2Criterion</td>
<td>Cost function (J).</td>
</tr>
<tr>
<td><img src="image4.png" alt="Icon" /></td>
<td>BackAxon</td>
<td>Placed together with Axon and manages the back-propagation algorithm.</td>
</tr>
<tr>
<td><img src="image5.png" alt="Icon" /></td>
<td>BackFullSynapse</td>
<td>Connects BackAxons from the right to the left.</td>
</tr>
<tr>
<td><img src="image6.png" alt="Icon" /></td>
<td>Momentum</td>
<td>Updates weights with momentum learning algorithm.</td>
</tr>
</tbody>
</table>
In the field of AI, NN and fuzzy logic based methods are two of the most common approaches for classification problems. In classifying the transportation mode, fuzzy-logic based algorithms could also be an alternative method. This method assumes that the observed data set has certain membership towards different attributes. For example, let us say a house needs a fuzzy-logic based fire alarm that can call 911 if a fire is detected. If there are two temperature sensors, where one is located outside and one is located inside the house, the fuzzy-logic based decision algorithm would require a set of membership functions and decision rules shown in the following.

**Figure 2.17 Screenshot of one-hidden-layer perceptrons in Neurosolutions**

2.2.8 Neural Network vs. Fuzzy Logic
Membership Function Example (Figure 2.18)

Figure 2.18 Temperature membership function example

Decision Rules Example

- If inside is both warm and hot and outside is cold, fire is detected
- If inside is warm and outside is cold, there is a chance of fire
- If outside is warm and inside is cold, no fire is detected
- If inside is cold and outside is warm, no fire is detected

The major challenge in using a fuzzy-logic classifier is intuitively determining the decision rules and membership functions. For example, if the temperature is 35°C, fuzzy logic may assign 80% membership to the “warm” category and 20% membership to the “hot” category. The determination of the temperature value above which it is classified as “hot” is a subjective process. Similarly, constructing decision rules is as subjective as building membership functions. For example, the abovementioned decision rules can now be specified further with specific membership values as decision conditions, as in the following.
- If the inside temperature belongs to the warm and hot categories with membership values greater than 10% and less than 90% respectively, and if the outside temperature belongs to the cold category with greater than 80% membership value, the house is not on fire.

In the real-time transportation mode detection applications, it is very difficult to determine highly subjective membership functions and decision rules. Experts can help deciding the parameters of a series of membership functions and decision rules. However, they will only make real-time application slower with many “if” statements in the codes. The more objective approaches to the determination of parameters involve an optimization tool such as the genetic algorithm that tries different combinations of values of parameters and look for the optimal set of parameters that minimizes the error. However, the genetic algorithm has its own flaws in that it cannot guarantee a global optimal result from its random nature.

Detecting the mode of transportation from a short trace of GPS trip data is a relatively new field and there are not yet many “experts” in this field. NNs do not require any subjective determination of parameters; instead, they adjust their connection weights based on given example data sets. One might say that NNs are not reliable because they blindly match input-output pairs. However, that is the whole purpose of moving away from human instructions and giving the NN the opportunity to educate itself freely to match the correct output classes from unseen input data better.

In this research, NNs are used in preference to methods based on fuzzy logic for the reasons set out in the previous paragraph.

2.3 Data Fusion and Kalman Filtering

2.3.1 Introduction to Data Fusion

Data fusion is a process of combining different sources of information to increase the
performance of a system. The different sources of information come from different traffic
sensors such as loop detectors, video cameras, traffic websites and radio broadcasts. The data
fusion process can fill the information gaps of any one sensor by utilizing multiple sensors in
unison covering a wider range of regions. In the case of multiple sensors monitoring the same
section of a road, the different sensors can complement each other to provide more reliable
information.

Data fusion processes are generally categorized as low, intermediate or high level fusion,
(Dasarathy, 1994) depending on the processing stage at which the fusion takes place. A low
level fusion refers to the case where multiple sources of raw data are used to produce new raw
data that is considered to be more informative than the raw input data. For example, some
consumer-level video cameras are equipped with 3 different sensors that record red, green and
blue respectively. Once all sensors record video streams in their respective colours, each colour
stream is transferred individually through data pipelines. At this stage, there are 3 different raw
data streams. Once the 3 streams reach a display device such as a colour computer monitor, all
three data streams converge at a point known as a “pixel” on the screen reproducing a colour
which is the newly created one raw data stream made from three original raw data streams.

An intermediate level fusion generally combines the common features from different sensors.
Even though a loop detector and a mobile GPS device operate on different technologies, they all
can provide a common feature such as a speed value. If the speed values from different types of
sensors are combined, the process can be categorized as an intermediate level fusion.

A high level fusion combines decisions from different “experts”. In this case, the quantitative
speed values from different sensors are not of interest. Instead, the final decisions from different
experts are generally qualitatively considered together to draw the final decision.

GISTMARG accepts both intermediate level data and high level data. The high level data is
converted into quantitative data first. Then, GISTMARG fuses both types of data in the
intermediate level. GISTMARG uses speed values from different types of sensors such as loop
detectors and mobile GPS devices. However, it also combines radio broadcasting verbal
information and colour-coded on-line website information supplied by the City of Toronto. In order to fuse the high level decisions (colour-coded decisions such as “green = moving well” or “red = moving slow”) from radio broadcasts and on-line traffic websites, GISTMARG downgrades those experts’ decisions into feature-based intermediate level speed data by quantifying the qualitative decisions from the experts.

2.3.2 Introduction to Kalman Filters

A KF is a recursive filter that can estimate the state of a dynamic system in a noisy environment. It can be used to estimate traffic conditions based on observed measurements from multiple sensors. The general set-up is presented in Figure 2.19. The “time update” module predicts the traffic condition of the next time step. The “measurement update” step collects the actual measurements and computes the Kalman gain which is used to “correct” the previous prediction made in the time update module. In addition, KFs also keep monitoring the error covariance values of its estimates, providing some insights into the confidence of the estimation at each time step. Because of this statistical feature, KFs are preferred over other data smoothing or forecasting techniques.

Figure 2.19 General set-up of a Kalman filter (Welch and Bishop, 2007)
2.3.3 Formulation of a Kalman Filter

In general, a KF estimates the state of a system which is represented in a form of a state vector variable, $x \in \mathbb{R}^n$ of size $n$. The filter uses measurements from various sensors, in a form of a measurements vector variable, $z \in \mathbb{R}^m$ of size $m$. In the case of traffic monitoring, $n = 1$ because the current traffic condition can be represented with a single speed value for a particular section of the road. The size of the measurement vector, $m$, can be greater than 1 for traffic monitoring because there can be multiple sensors that are monitoring the same section of the road.

In the time update (or prediction) step, the previous estimation about the previous system state is projected for the future.

$$\hat{x}^\sim_k = A\hat{x}_{k-1} + Bu_k$$  \hspace{1cm} (Equation 2.43)

$$P^\sim_k = AP_{k-1}A^T + Q$$ \hspace{1cm} (Equation 2.44)

where $\hat{x}^\sim_k$ is the current state estimation before sensor measurements,

$\hat{x}_{k-1}$ is the state estimation in the previous step after the measurement updating,

$u_k$ is an optional control input vector $u \in \mathbb{R}^l$ (where $l$ is the size of the dimension of $u$) which can be thought of as a trend in the state of the system that prematurely updates the current state before the actual measurements are made,

$P^\sim_k$ is the estimation of the current error covariance,

$P_{k-1}$ is the estimation of the error covariance in the previous step after the measurement updating,

$A$ (n $\times$ n) is a matrix that relates the state at the previous time step $k-1$ in the current step $k$. 
B (n \times 1) is a matrix that relates the optional control input vector \( \mu \) in the current step \( k \), and 

\( Q \) is the variance of the process noise.

In the measurement update (or the correction) step, the immature projections made in the time update step for the system state and the error covariance are corrected with sensor measurements via a variable \( K_k \) known as a Kalman gain.

\[
K_k = P_k^{-}H^T(HP_k^{-}H^T + R)^{-1} \quad \text{(Equation 2.45)}
\]

\[
\hat{x}_k = \hat{x}_k^{-} + K_k(z_k - H\hat{x}_k^{-}) \quad \text{(Equation 2.46)}
\]

\[
P_k = (I - K_kH)P_k^{-} \quad \text{(Equation 2.47)}
\]

where \( K_k \) is the Kalman gain,

\( H (m \times n) \) is a matrix that relates the state to the measurement vector \( z_k \),

\( R \) is the measurement error covariance,

\( \hat{x}_k \) is the current state estimation after the current step measurements are made,

\( z_k \) is the measurement vector of size \( m \),

\( P_k \) is the estimation of error covariance in the current step after the measurement updating and,

\( I \) is the identity matrix.

### 2.3.4 SCAAT Kalman Filters

A KF is a set of mathematical equations that implement a predictor-corrector type estimator that
is optimal in minimizing the estimated error covariance (Welch, 1996). The traditional application of the KF described in Section 2.3.3, collects a group of sensor measurements and attempts to simultaneously solve a system of equations for the future state estimations. This multiple constraint method has one major disadvantage: it requires measurements from all the sensors for each estimate, resulting in significantly high updating time intervals because the algorithm waits until all the sensor measurements are input. In traffic monitoring, different sensors provide measurements at different updating frequencies and occasionally, the measurements are unavailable. Therefore, a special type of KF is used for traffic monitoring in this thesis. Welch (1996) developed a modified version of KF known as the Single-Constraint-At-A-Time (SCAAT) filter. The method uses the single most recent measurement from any available sensors to update its estimations based on the characteristics of the observed sensor (i.e., the variance of the measurement) and the accumulated (resulting in faster processing time) state estimation from the previous step. The formulation is almost similar to the conventional KFs except that the SCAAT filter reads the single most recent sensor in the measurement update step. SCAAT filtering is useful in traffic monitoring because not all sensors are always accessible simultaneously.

Because the SCAAT filter updates its estimation based on one sensor at a time for one monitoring state variable (speed), the formulation of the conventional KF is greatly simplified. Instead of using multi-dimensional vectors of parameters, the SCAAT uses scalar parameter values. For the case of speed monitoring, the formulation of the time update step of SCAAT filter is shown in the following.

\[
\hat{x}_k^- = a\hat{x}_{k-1} + b\mu_k \quad \text{(Equation 2.48)}
\]

\[
p_k^- = a p_{k-1} + q \quad \text{(Equation 2.49)}
\]

where \(\hat{x}_k^-\) is the speed estimation of current step before the most recent single sensor measurement,

\(\hat{x}_{k-1}\) is the speed estimation in the previous step,

\(p_k^-\) is the variance estimation of current step before the single most recent measurement,
and

\[ p_{k-1} \] is the variance estimation of the previous step.

The measurement update step of the SCAAT filter for the speed monitoring is shown in the following.

\[
k_k = p_k^- (p_k^- + r)^{-1} \tag{Equation 2.50}
\]

\[
\hat{x}_k = \hat{x}_k^- + k_k (z_k - \hat{x}_k^-) \tag{Equation 2.51}
\]

\[
p_k = (1 - K_k)p_k^- \tag{Equation 2.52}
\]

where \( k_k \) is the scalar Kalman gain,

\( r \) is the device variance,

\( \hat{x}_k \) is the speed estimation in the current step,

\( z_k \) is the single most recent measurement in the current step, and

\( p_k \) is the variance of estimated speed value \( \hat{x}_k \) after the measurement updating.

2.3.5 Quantifying Subjective or Non-numerical Information

In order to fuse various types of data from different sensors, the SCAAT filter needs to know how to handle subjective and qualitative information and convert them into quantitative data. For example, radio broadcasts and traffic departments’ traffic monitoring web sites provide us with highly subjective information. A radio broadcaster for example, uses terms such as "moving well" or "congested", and the web sites colour the roads on their maps green (as shown in Figure 2.20) for moving-well condition. SCAAT filter requires numerical sensor data with sensor-specific error information to be operational. In this research, the sensor inputs from
qualitative sources are assumed to be the mean value of that category and the sensor readings are assumed to be distributed with a normal distribution. This is rather a strong assumption and the philosophy behind this approach is to reuse widely available online information from “credible” sources even when the exact decision processes behind those qualitative expert-based 3rd party information are unknown. There are many examples of similar practices in internet web services nowadays. Information originally prepared for Google Earth™ are ported out to other commercial web services such as the location finder of retail stores in certain locations. Instead of ignoring the 3rd party information completely, this thesis attempts to utilize them as an illustrative example of treating such information sources as less accurate sensors. If the accuracy and the data distribution of those information sources are carefully analyzed in the future, they are not the “less accurate sensors” anymore and they then can be treated as if they are normal quantitative sensors. However, this is not the goal of this thesis regarding the use of the expert-based information sources.

It is a well-known fact that 3 standard deviations to the left and right of the mean value covers 99.7% of total samples in the case of a normal distribution as shown in Figure 2.21. Using this fact, the range of any given category of qualitative sensor can be divided by 6 to roughly estimate the corresponding standard deviation of that sensor in that category.
Figure 2.20  Screenshot of colour coded traffic monitoring web sites for Hwy 401 and arterial roads

Figure 2.21  6 sigma covering 99.7% of population in a normal distribution
2.4 Global Positioning System (GPS)

2.4.1 Background of GPS

This research uses GPS technology in order to trace a vehicle trip and estimate travel times on desired links. GPS is a satellite-based positioning system planned by the U.S. Department of Defense in the early 1970s and has been operational since 1995 (Chung, 2003). The system enables the positioning of objects near the surface of the Earth, 24 hours a day and 7 days a week. It does not require any subscription fee in order to use the system as long as a GPS receiver is purchased before using the service. Prior to May 1, 2000, GPS signals were intentionally embedded with errors known as the selective availability (SA). After that date, the GPS signals have been free of intentional errors. The GPS receiver devices are associated with typical location errors ranging from 10 to 15 meters. (Byon et al. 2006) There are hybrids and improved versions of GPS implementations such as DGPS (Differential GPS) or WAAS (Wide Area Augmentation System) enabled GPS. They both require ground-based radio stations that send corrective signals to the GPS devices in addition to the general GPS signals from the satellites. The hybrid systems result in higher positioning accuracy because the corrective signals are transmitted from radio stations with accurately known fixed geographical locations. In Canada, DGPS radio stations are available only in the coastal regions. WAAS-enabled GPS devices are available for purchase while the correctional signal is not available in Ontario, Canada as of December 2004.

2.4.2 The 3 Segments of GPS

The GPS consists of 3 segments; the space segment, the control segment, and the user segment, as shown in Figure 2.22.
The space segment includes 24 satellites that broadcast navigation signals to receivers through carrier waves. Figure 2.23 shows the orientation of the 24 satellites that are distributed in 6 circular orbits with an inclination angle of 55° in groups of 4 satellites each. Each satellite is located 26,565.5 km away from the Earth’s center, orbiting the globe every 12 hours. The shown arrangement guarantees that 6 to 11 satellites are always in view at any point on the globe’s surface provided that they are not blocked by physical interruptions (Chung, 2003).
The control segment traces the satellites through ground control stations to monitor their locations and check their status. The need for the control segment is that the satellites can go out of their planned positions and they may need to be adjusted to maintain the designed ideal spacing in the GPS satellite network.

The user segment refers to the user-end GPS receiver that calculates the time the radio signals travel from the satellites to the GPS receiver. The communication is “one-way” in the sense that the satellites only emit the radio signals while the user-end GPS receiver only receives the signals. By calculating the travel times of the signals between the satellites and the GPS receiver, the distances between the current position of the receiver and the satellites can be found. As long as there are at least 3 satellites, the 3 spheres constructed from the 3 distances between the receiver and the satellites can be used to pinpoint the current location. 2 spheres intersect in the shape of a circle and the third sphere intersecting with the circle will produce 2 points that are the 2 possible choices of the current location. In general, one of the 2 points is near the surface of the Earth and the other one is out in space. The one near the surface is the theoretical current location of the receiver. One additional satellite can validate the chosen point.
further. As a rule of thumb, a higher number of satellites in view results in more accurate location determination. The location determining concept is illustrated in Figure 2.24.

Figure 2.24 Calculation of the current location of the GPS device

2.5 Geographic Information Systems (GIS)

This section presents background information about the Geographic Information Systems (GIS). This research mainly uses a particular GIS software package known as ArcGIS Desktop from ESRI Inc.
2.5.1 Vector vs. Raster Layers

In GIS, features of maps are usually represented in a vector format. The vector format uses geometrical primitives such as points, lines, curves, and polygons. The primitives are basically mathematical equations that are used to represent shapes for computers to understand. Generally, the vector format results in sharp images with relatively small file sizes because not all points on lines are stored but only the important corners or line-ends are actually stored in the file format and they are simply connected with equations whenever the files are loaded to computers. Sometimes, it is more beneficial to use another representation method known as a raster format. A raster format is a data structure representing a generally rectangular grid of cells where each cell has a value associated with it. The vector format is useful when a map is trying to represent shapes and locations of features, while the raster format is more suitable when there is a continuous spread of data values over the entire map. For example, if one wants to represent the elevation values over the entire Toronto region, the raster format is naturally more suitable. By definition, the raster format covers the entire region with a grid of cells. Graphically speaking, when raster files are zoomed in, eventually, a grid of cells will be shown on screen where each cell contains a uniform information (or colour if the file is an image). However, when vector files are zoomed in, the quality of the lines between the important points remain sharp and clear because the file contains equations to connect the points rather than the pixel-by-pixel information. Figure 2.25 illustrates the concept.
2.5.2 Digital Elevation Model (DEM)

A digital elevation model (DEM) is a digital representation of a ground surface topography or terrain. It is sometimes also referred to as a digital terrain model. A DEM can be represented as a raster or as a triangular irregular network. Figure 2.12 shows an example of DEM models. For Toronto, recent DEM models are in the raster format.
The accuracy of a DEM can be evaluated from the perspectives of absolute values and relative values. The first perspective focuses on the accuracy of the elevation data with respect to the actual terrain. Data collection is generally carried on at some of important points on the terrain such as the highest and lowest points. Then the surface between those important points are estimated using terrain interpolation algorithms. The second perspective focuses on the relative accuracies of the interpolated values with respect to the important points with which the interpolation is carried on. If the elevation data are collected at as many locations as possible with the most accurate elevation measuring device available, resulting in finer grid of cells with less amount of data interpolated between the data collection points, this would produce an ideal DEM.
2.6 Different Levels of Route Guidance Systems

Different route guidance systems find different routes based on the objective that they are optimizing for. They can be categorized into de-centralized and centralized systems. In de-centralized systems, the algorithm finds the optimal path for the driver to minimize the driver's cost such as travel times. In centralized systems, the algorithm finds the routing strategies for all drivers to achieve the network-wide optimal cost such as a minimum average travel time for all drivers. The results from the 2 different systems produce different routing strategies. Depending on whether the system has a feedback mechanism or not, the route guidance algorithms can also be categorized into open-loop or closed-loop systems. Open-loop algorithms do not include real-time traffic conditions into their calculations while closed-loop algorithms monitor current traffic information and feed them back into computing their current optimal paths.

2.7 Paramics Micro-simulation

In transportation engineering, there are various micro simulation software packages. They simulate movements of vehicles on a pre-determined road network on a micro level. Micro simulation models are mainly used to assess the changes in traffic patterns as changes are made to the physical environment of the road. As one of the most popular packages, Paramics (Quadstone, 2007) is available in the ITS Lab at the University of Toronto. Paramics is expected to be able to produce second by second trajectories of massive number of vehicles in the GTA with associated mix of different types of vehicles with different behaviours. In this research, the outputs from Paramics are processed with the developed fusion module and the results are compared with real-life GPS based results. One major advantage of using Paramics based simulated data is that, at least, in its own simulation environment, there are no errors. Therefore, the true road condition of a particular road link can be computed second by second by averaging speed values of all vehicles on the road.
3 State of the Art and Literature Review

This thesis work is related to 3 main research topics: traffic monitoring, data fusion and route guidance. This chapter reviews related past research works in the 3 topics.

3.1 Traffic Monitoring

3.1.1 Vehicle Information and Communication System (VICS) (Kyobashi, 2008)

VICS is a digital data communication system that provides the latest road traffic information for drivers, via car navigation systems. It has been in operation in Tokyo, Japan from April, 1996. It uses various technologies such as loop detectors, roadside radio-wave beacons, infrared beacons and data broadcasting via FM radio signals. It uses sufficiently many fixed point detectors throughout the city. The system promotes safe driving by sending real time traffic information via FM signals. It is noted that such system may only be feasible in cities with high density of population and sufficient funding sources. Implementing high cost equipments in high density at fixed locations will however provide reliable and continuous traffic information even without incorporating probing methods. However, the approach of implementing expensive equipments in high density seems less feasible for typical North American cities where population density is not as high as in Japanese cities.
3.1.2 Traffic Monitoring Using Digital Sound Field Mapping (Chen et al., 2001)

A new traffic sensing technique is described by Chen et al. (2001) that utilizes a microphone array to detect the sound waves generated by the vehicles on the road. The detected signals are then digitized and processed by an on-site computer to determine the traffic condition. Ultimately, the method is a fixed-position sensor based method that may replace loop detectors. The authors are credited for contributing to the field by applying innovative sound mapping techniques for traffic monitoring. As far as the nature of the operation is concerned, the microphone based monitoring is in the same category with loop detectors because they all use location-fixed sensors.

3.1.3 Efficient Vehicular Traffic Monitoring Using Mobility Management in Cellular Networks (Saraydar et al., 2004)

The paper attempts to apply the know-how in wireless communications research known as “mobility management” to gather real-time vehicular traffic information. The main goal of the mobility management is to track where the cell phones are in order to keep them connected to the service provider’s network. Mobility management makes it possible for the cellular system to track a mobile user as it roams around within the wireless network using occasional location updates. The cellular network collects simulated GPS-based location information from mobile units and maps the current location to a particular link on the GIS transportation network and finally estimates the link occupancy. As of 2006, most cell phones are equipped with AGPS chips in them. This enables future other researchers with opportunity to work with real GPS data from the cell phones.
3.1.4 Microscopic Traffic Data Collection by Remote Sensing
(Hoogendoorn et al., 2003)

This paper describes a prototype of a new data collection system for determining individual vehicle trajectories from sequences of digital aerial images. The new method involves taking a series of aerial photos from a helicopter over a highway and was able to determine trajectories of one vehicle by analyzing the digital pictures. The performance was influenced by weather conditions and it would be an expensive traffic monitoring system both in taking pictures and transferring high resolution pictures over wireless data communication system if it was to be processed in real time. The newly introduced method is not suitable for a real-time traffic monitoring application due to the high cost for uncertain level of improvements (if there are any) over already sufficiently accurate GPS devices. However, the authors still contribute to the field by opening the possibilities of using remote sensing techniques for collecting microscopic-level traffic data.

3.1.5 An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys (Tsui and Shalaby, 2006)

In order to use GPS data for traffic monitoring purposes, one needs to identify the mode of transportation first and use the data only if they are from the “auto” mode that refers to private automobiles. Tsui and Shalaby (2006) post-processed GPS trip data and they were able to identify the transportation mode the GPS data logger was in. They applied fuzzy-logic based mode identification algorithm against multiple days’ amount of GPS trip data. The challenge in real-time traffic monitoring is to achieve similar results from relatively short trace of GPS data in shorter amount of time.
3.1.6 Using Information Technology to Evaluate Non-traditional Traffic Monitoring Systems (Guo et al., 2007)

The paper recognizes that there are new approaches of traffic data collection (monitoring) methods emerging that can be grouped into one category known as “link monitoring systems”. Such systems “mine” location data from existing infrastructure (such as cellular telephone networks or fleet management systems) to estimate link speeds and travel times. However, they are generally developed and operated by private firms, who then seek to sell the data to transportation agencies. The paper presents a link monitoring system evaluation procedure in order to help transportation agencies make sound purchasing decisions. Authors contribute to the field by developing an evaluation scheme for traffic monitoring systems mainly focusing on the output data regardless of the specific technologies the data are produced from.

3.1.7 Kalman Filtering Applied to Network-based Cellular Probe Traffic Monitoring (Qiu and Ran, 2008)

The paper attempts to provide real-time traffic states by analyzing information about hand-off or hand-over times of a cell phone and tries to estimate the speed of the mobile device by applying KF techniques. The method assumes the cell phones are already in “auto” mode and uses location determination method based on the locations of cell towers which generally produces unreliable speed information in its raw format. Instead of applying filtering techniques to the inaccurate information in order to improve the quality, the cell phone service providers can remotely activate the embedded GPS chips in cell phones and query the speed information directly to the chips instead of querying hand-off times.

3.1.8 Evaluation of a Cellular Phone-based System for Measurements of Traffic Speeds and Travel Times: A Case Study from Israel (Bar-Gera, 2007)

The paper evaluates the traffic monitoring system that is based on triangulating cell phone towers
with multiple cell phones on a particular route in Israel. In comparison to the results from loop detectors, the cellular tower based method produced similar results. The traffic monitoring technique that uses triangulation of cell phone towers can produce reliable information on major highways where there are sufficiently many vehicles travel and many cell towers are installed nearby. In order to build a framework of a city-wide traffic monitoring system, different technologies are necessary. AGPS based cell phones can provide the city-wide monitoring as long as their modes of transportation they are in are correctly identified.

3.1.9 Transit Vehicles as Traffic Probe Sensors (Cathey and Dailey, 2002)

The paper presents a method of using transit vehicles as probes for determining traffic speeds and travel times along freeways. The correlation between probe data and inductance loop detector data is found. The presented algorithm is innovative in a sense that transit vehicles can indeed be the probe vehicles themselves as long as there is a verified correlation between the transit vehicles and private automobiles. Using voluntarily moving transit vehicles as probes, traffic conditions throughout a city where bus routes are can be achieved at a fractional cost of operating the actual probe vehicles. However, there are potential limitations with this method. Transit vehicles operate only on major roads with a fixed route limiting the monitoring range of coverage. In addition, the speed correlation between the transit vehicles and other vehicles are not always the same temporarily and spatially even in a same city. Building and managing multiple correlation database maybe more expensive and less accurate than the traditional probing techniques.

3.1.10 Cellphone Probes as an ATMS Tool (Smith et al., 2003)

Authors assess wireless location technology (WLT)-based traffic monitoring along with recent operational tests. The recent tests are from;

- The Virginia Department of Transportation (VDOT)
- Maryland State Highway Administration (MSHA)

- US Wireless Corporation (USWC)

The paper concludes that WLT-based traffic monitoring is an extremely appealing conceptual approach for collecting traffic information. The authors agree that there is a need for concerted efforts to address the significant sampling challenges raised by the WLT-based traffic monitoring systems. The paper contributes to the field by drawing a conclusion, based on actual case studies, that WLT-based traffic monitoring is indeed an appealing approach.

3.1.11 Reconsideration of Sample Size Requirements for Field Traffic Data Collection with Global Positioning System Devices (Li et al., 2002)

Li, S. et al. (2002) investigates the general sample size requirements for establishing a consistent method for GPS-based network monitoring. Their work focuses on determining how many probes on a given link are needed to estimate the travel conditions on that link. This work can be considered as an effort for resolving the “significant sampling challenges” that Smith et al. (2003) refer to. Authors find that sufficient monitoring accuracy usually requires 5 to 10 readings to estimate travel time, delay and work zone conditions. Based on their findings, in an AGPS cell phone traffic monitoring system, for each desired road link, at least 5 to 10 AGPS cell phones need to be queried at once. The study suggests that in the case of less than 5 pings per road link, the results may be unreliable and biased due to the small number of sampled vehicles.

3.2 Data Fusion
3.2.1 Probe Vehicle Runs or Loop Detectors? Effect of Detector Spacing and Sample Size on the Accuracy of Freeway Congestion Monitoring (Kwon et al., 2007)

Freeway congestion monitoring can be based on either sampling-based methods such as probe vehicle runs, or continuous data from loop detector infrastructure. The paper recognizes the sample size, in terms of the number of days sampled, affects the accuracy of sampling-based method, and the detector spacing affects the accuracy of detector-based method. The paper concludes that in order to achieve within 10% of errors on congestion parameters, 4 to 6 days worth of probe vehicle data are needed for the sampling-based method and half-mile spacing is needed for the detector-based method. The main focus of their paper is to compare the effect of varying parameters on 2 different monitoring methods. How to handle other forms of sensors or information sources is outside the scope of the paper.

3.2.2 Traffic Flow Reconstruction Using Mobile Sensors and Loop Detector Data (Herrera and Bayen, 2008)

Herrera and Bayen proposed and investigated new algorithms that make use of data provided by mobile sensors in addition to stationary detectors, to estimate traffic flow. The authors recognize that there are multiple sources of data that are monitoring the same section of the road and try to combine the 2 different types of data. They approached the data fusion problem with multiple methods including KF method. However, their proposed KF-based method disregards the mode detection analysis step. Although their KF method performed better than other methods they developed, for practical applications, they recommended one of the lower performing methods due to the missing mode detection capability of the Kalman filtering method. The results in all cases showed that data provided by the mobile sensors significantly improves the accuracy of the estimates made by loop detectors.
3.2.3 SCAAT: Incremental Tracking with Incomplete Information (Welch, 1996)

As introduced in Section 2.3.4, Welch (1996), in his doctoral thesis paper, suggests a new type of KF technique that updates its next estimate based on the most recent single sourced data that is known as “single-constraint-at-a-time” (SCAAT) KF. In his method, the filter updates its estimate on the state of the system as soon as any sensor delivers information. In conventional KF method, the filter would wait for all sensors to provide information and estimates the next state by using all data from all sensors. The use of such filter is beneficial in traffic monitoring applications where it utilizes various types of sensors for monitoring same sections of the road. Traffic sensors generally provide information with certain sampling intervals which varies depending on the type of the sensor. SCAAT filter can be used to update traffic information as soon as the most recent data is received and at the same time it can simultaneously fuse the data from different sensors.

3.3 Route Guidance

3.3.1 Vehicle Route Guidance Systems: Classification and Comparison (Schmitt and Jula, 2006)

The authors classify and compare different branches of route guidance systems. They classify the route guidance algorithms into “Static vs. Dynamic”, “Deterministic vs. Stochastic”, “Reactive vs. Predictive” and “Centralized vs. Decentralized”. “Static vs. Dynamic” classification distinguishes whether the algorithm uses real-time traffic information as an input. “Deterministic vs. Stochastic” classification assesses the use of the random nature of traffic condition. Reactive algorithms solely use current traffic information, while a predictive algorithm is based on expected patterns. Decentralized algorithms optimize for the individual end-user while centralized algorithms optimize for the entire network. The paper contributes to the field by categorizing route guidance systems from different perspectives. With the above classification perspectives in mind, this thesis attempts to create new
classification dimensions and new categories of route guidance systems that divide algorithms into “2-D vs. 3-D” and “travel distance/travel time based vs. other multi-criteria based”.

### 3.3.2 Evaluation of Walkability of Routes (Scholossberg, 2006) and (Scholossberg et al., 2007)

Scholossberg (2006) presents a method of evaluating the “walkability” of routes based on GIS street data and Scholossberg et al. (2007) develop a GIS based method of auditing the walkability on the roads by collecting data from handheld PDAs. Both papers extract innovative new information called “walkability” that measures the level of difficulty of accessing certain locations, by using readily available GIS data. Extracting the “visibility” and “slopes” from DEM files for the route guidance systems in Chapter 6 are conceptually inspired by the two studies.

### 3.3.3 Enhancing In-car Navigation Systems with Personal Experience (Bederson, 2008)

The paper suggests that computer-based route guidance methods are generally weak with respect to subjective reasoning. The authors present a framework for adding subjective human experience to in-car navigation systems. They developed a set of methods to help people record their personal driving history, add rich annotations, and share their data with friends and family. Participants in this study carried mobile personal digital assistant (PDA) devices with which they could log their trips with subjective annotations and other people could access this information afterwards. However, the major disadvantage of such system is that it contains subjective and qualitative information that cannot be processed by in-car navigation programs directly. The author contributes to the field by recognizing the fact that the subjective reasoning plays a role on route choices. The route guidance module of this thesis attempts to quantify and incorporate those subjective factors into the route guidance methods.
3.3.4 Comparative Study of Decentralized Feedback Control Strategies for Route Guidance Purposes (Zuurbier, 2007)

Route guidance can improve traffic flows in a network by optimally directing drivers to different paths. The paper compares different approaches of decentralized feedback control for directing drivers. Decentralized open-loop routing finds the route based on locally available data without any feedback such as current traffic conditions. Decentralized closed-loop routing refers to the case where drivers get current road conditions as feedback and optimize for what they desire (i.e. travel time). On the other hand, centralized routing refers to the case when a centralized optimizer is used to achieve the network-wide optimal state. The author contributes to the field by categorizing levels of route guidance approaches and experimenting different versions of decentralized strategies.

3.3.5 A Behavioral Component Analysis of Route Guidance Systems Using Neural Networks (Hamad, 2003)

Even if a route guidance system suggests a route, drivers do not always choose the suggested one. In order to reduce congestion by directing traffic flows, the authors focus on the behavioral component of a practical route guidance system developed through a 4-year project at the University of Delaware. The paper assumes that different drivers perceive and behave differently in response to the information provided. Understanding the behavior of the drivers is essential for understanding the reliability of the system. A back-propagation NN is used to structure the behavioral model. The paper shares a common theme with the paper introduced in Section 3.3.1 (Bederson, 2008) that route choices are influenced by subjective and behavioral components of human minds in addition to travel distance and travel times.
3.3.6 Park and Bike: A New Multi-modal Concept for Congested Areas (Bos and Vrugt, 2008)

Bos and Vrugt (2008) provide insights in implementing “Park and Bike” strategies with which inter modal connection strategies from and to the bike mode are discussed. In order to successfully integrate the bike mode with other modes, an effective route guidance system for bikes that is sensitive to the unique features of the bike mode is required. For example, bike riders may prefer a longer flat path over a shorter inclined path since steep slopes are harder to negotiate and may be impossible to travel for some riders. With regards to the safety issues, bike riders and pedestrians are directly exposed to the external environment that they navigate through. As a result, they, would probably prefer to avoid high-crime regions.
Chapter 4
Transportation Mode Detection

4 Transportation Mode Detection

This chapter presents a methodology for identifying transportation mode via tracking GPS-equipped mobile devices in the traffic stream. Various cell phone service providers have location-based services (LBS) that track the locations of their cell phones. One major concern in using cell phones for traffic monitoring is that the phones are not necessarily in private vehicles. The mobile device can be in a car, bus or other modes of transportation that have distinct speed and acceleration profiles. In addition, querying the mobile device has financial cost implications, and the higher the number of location queries from the server the higher the associated cost. This chapter focuses on the feasibility of using the characteristics of the trail of a GPS data stream to identify the mode by which the mobile device is traveling in. Currently available LBS in Toronto can only provide GPS data once every 5 minutes. Because of the sampling limitation, a GPS data logger is used to collect the trip data and the logged data is sampled at varying frequencies as if coming from the cell phones. The analysis is conducted using NNs to determine the transportation mode.

The analysis also examines the impact of varying sampling rates (number of pings per unit time) and monitoring duration (time length of data trail) on accuracy of mode classification. In total, 60 hours of GPS data were collected while traveling on various transportation modes throughout the Greater Toronto Area (GTA). Results confirm the potential of NNs to successfully detect transportation modes from GPS data, both in peak and non-peak periods. The results indicate quantitatively that higher sampling frequency and longer monitoring duration result in higher mode detection rates. In addition, it is also found that the route-specific NNs perform better than the universal NN.
4.1 Introduction

Fixed point sensors such as loop detectors and cameras have been widely used to monitor road traffic conditions for traffic management and ITS applications. They provide relatively accurate information at strategically chosen fixed points, generally on major roads or highways. The high installation and maintenance costs of such detectors limit a city-wide deployment. Therefore, fixed detectors do not provide spatially continuous traffic information over the monitored network. There is a rapidly growing interest in moving away from fixed detectors to tracking “probe” vehicles as they travel through the network, both on major and on minor roads. Byon et al. (2006) developed a real-time traffic monitoring system named GiSTT that centrally tracks dedicated probe vehicles equipped with GPS and wireless communication devices. However, having each probe vehicle equipped with dedicated high-cost equipments and operated by a dedicated driver is not economically feasible. In order to acquire massive and economically feasible data from probe vehicles, cell phones that are widely in use and have already been dispatched throughout the city can be used. Currently, most cell phone networks provide location-based services (LBS) that can determine the locations of cell phones. The classical method of triangulating their phones against nearby cell phone towers can provide an accuracy in the similar order of separations between the cell phone towers within 100 meters. For higher accuracy, modern cell phones have GPS chips embedded in them known as Assisted GPS (AGPS) chips. AGPS eliminates the need for the classic triangulation process while providing a location accuracy of within 10-15 meters. (Byon, MASc Thesis 2005) The main use of the AGPS technology, however, has been on determining the locations of the phones, regardless of the mode of transportation the device is in. Speed and acceleration characteristics of buses, for instance, vary from those of private cars. The embedded GPS chips can also provide additional information, such as the number of satellites in view. Such information can help detecting the transportation mode the mobile device owner is using. An immediate use of such effort is to filter out the “auto” mode data and use it for traffic monitoring purposes. The data then can be regarded as live probe data since it is coming from the private automobiles.

Unlike conventional GPS data loggers from which second-by-second GPS data can be collected, the AGPS phones need to be queried from the server. Typically, the cell phone service providers
charge a certain fee for each query or “ping”. Therefore, it is important to minimize the pinging frequency while achieving sufficient “auto” mode detection accuracy. As of August 2006, in Ontario, Canada, a local LBS offers such services and it can practically query a phone every 5 minutes. In this study, both sampling rates and monitoring duration are varied and their impacts on the mode detection processes are assessed. In order to overcome the sampling limitation provided by the local LBS, a SirfstarIII-based GPS data logger is used instead of a cell phone to collect the raw data while traveling on different modes of transportation and in different areas of the GTA. The collected data are sampled at varying frequencies as if the data are coming from cell phones. In the lab, the data are sampled at varying rates to emulate real-time monitoring. It is noteworthy that this research applies to any GPS data stream, such as cell phones, in-vehicle navigation devices or any other GPS data source.

4.2 Study Objective

The objective of this study is to examine the possibility of identifying the travel mode using a relatively short trace of GPS data streams. The feasibility of using NNs for the mode identification/classification processes is assessed. Then, the impacts of sampling frequency and monitoring duration on the mode detection rates are found. In addition, the mode detection rates between peak and non-peak periods are compared. Finally, the performances of a general NN classifier are compared with route-specific NN classifiers, i.e., a classifier trained specifically to detect modes on a specific route as opposed to the entire network.

4.3 Data Collection

In the data collection phase of the study, a portable GPS data logger (Sanav Inc., GL-50-BT) is dispatched along major routes of the GTA on different modes of transportation; namely auto (arterial roads and highways), bus (arterial roads and highway), streetcar and walk modes. AGPS cell phones in Toronto can only provide a sampling frequency of up to once every 5
minutes as set by the local AGPS service provider. Therefore, in order to investigate various sampling rates, a portable GPS data logger is used to collect GPS data streams. The data are collected by the author of this thesis alone. The main purpose of this is to test the feasibility of using NN for distinguishing the different transportation modes one individual experiences. In the future, more extensive data collection from different population can be conducted in order to study the effect of driving behaviours on the performance of the mode detection processes. The collected data are then streamed into the developed classification algorithms as if they were coming from AGPS cell phones via the service provider. The main focus is on processing data from moving probes and not the underlying AGPS technology per se. The developed algorithms and analysis described in the following sections are technology-independent, i.e., any technology that can provide the same input variables can be used, be it GPS, AGPS or other future technology. A total of 60 hours of GPS data, 15 hours for each mode of auto, bus, streetcar and walking, are collected and analyzed in this study. The Sanav’s GPS data logger provides a software interface where data sampling frequency and data formats can be set before the data logger is dispatched for the data collection. Once the data are collected, the trip data are then downloaded from the data logger via a USB computer cable. It is noted that the data collection is carried out by a single person. Therefore, there can be a bias from the auto mode and walk mode since different people behave differently. Therefore, in order to implement the methodology of this thesis, a further data collection from wider range of individuals needs to be carried out. However, the methodology presented in this thesis is independent from the issue of having the true average behaviours and can be fine-tuned when such information becomes available.

4.3.1 Study Region

It would be ideal to find a route that contains all 4 modes of transportation considered in this study, but buses and streetcars do not usually operate on the same routes in the GTA. Therefore, the bus mode data are collected on different routes from streetcar routes. Of the 60 hours of collected GPS data, 15 hours are obtained from each mode: auto, bus, streetcar and walking.
The following criteria are used to select routes and time slots for the data collection process in the GTA:

1. The selected routes should contain 3 modes: auto, bus and walk or auto, streetcar and walk.

2. The selected routes should offer a variety of physical operating conditions (such as downtown area, arterial roads and highways) in order to best represent the general behavior of the different modes. This is particularly important for the universal mode detection NNs.

3. The data collection time periods should contain both rush hour and non-rush hour periods.

For the streetcar mode, Toronto Transit Commission (TTC) route number 506 was selected because it travels along a corridor that includes vehicular traffic and pedestrians while it runs from the west end to the east end of Toronto through the Central Business District (CBD). This route appears to be representative of a typical busy arterial route with mixed modes of transportation and with a variety of operating conditions for all modes under consideration.

The bus mode data are collected on different routes from those used for the streetcar mode because the two modes do not operate on the same routes in the GTA but rather complement each other throughout the TTC transit network. However, in order to keep the data as homogeneous as possible in comparison with the other 3 modes, some bus routes close to the previously selected 506 route are selected. They are TTC routes 5, 29 and 91. Figure 4.1 shows the selected routes.
Figure 4.1 Arterial routes for the data collection in the Greater Toronto Area (GTA)

For the auto-highway mode, all major highways of the GTA – highways 401, 427, 404 and QEW – are selected, as shown in Figure 4.2.
Figure 4.2 Highway data collection routes in the GTA

For the route-specific mode detection scenario, the highway 427 route is selected due to its uniqueness in that transit buses operate on this high speed freeway. It is the only case that allows comparison of the GPS data characteristics from the bus-highway mode to that of the auto-highway mode. Figure 4.3 shows the chosen route.

Figure 4.3 Highway 427 route with auto mode and bus mode only

It is also important to examine the potential capability of the mode classification during both peak periods and non-peak periods. TTC bus route 32A, which runs on Eglinton Avenue
between Yonge Street and Jane Street, is chosen for the data collection. The mode detection rates of the auto-arterial mode and the bus-arterial mode are compared for both peak and non-peak periods.

### 4.3.2 Data Post-Processing to Compute Acceleration

Typical GPS devices provide information such as longitude, latitude and speed. It is highly desirable, however, to have acceleration values for the purpose of mode detection. Different modes have different physical acceleration and deceleration characteristics. For example, auto mode tends to have higher acceleration values than bus and walk modes. Conventional GPS data loggers do not directly provide acceleration values, and therefore the user has to post-process the collected data to extract acceleration values. In the case of AGPS cell phones, a small application can be coded under the operating system of the device so that the second-by-second acceleration values are available whenever the cell phone is pinged. In addition, the pinging or sampling can either be initiated from the central LBS server or an application on the cell phone itself can be configured to transmit the GPS information at certain time intervals to the server.

### 4.4 Classification of Process Inputs and Outputs

The mode identification analysis relies on detecting mode-specific patterns from the GPS data. There are significant differences in the collected GPS data among the different modes. The auto mode has the highest average speed value and average number of satellites in view. On a bus or a streetcar, a GPS device would typically have a smaller view of the sky due to wider ceilings and relatively smaller and sharply vertical windows, limiting the chances of a direct line of sight between GPS devices and satellites. Cars on the other hand generally have wider front windshields that allow stronger and multiple satellite signals. Speed, acceleration and number of satellites in view are major input variables that are used to distinguish between different modes in this study. The developed models use as input, the values of the above variables sampled in varying frequencies (pings per unit time) over varying periods of time (monitoring duration); the
AGPS phones can hold a few most recent speed and acceleration values in their temporary buffer memory and have them available for each pinging. As a result, it is possible to access recent multiple speed and acceleration values with one pinging. The only exception is the number of satellites in view. Instead of using the variable multiple times, the average number of satellites in view is used. A preliminary observation from the collected data indicates that the number of satellites in view is strongly dependent on the mode itself and does not fluctuate as much as speed or acceleration values. It is believed that having multiple nearly unchanging input values does not provide any additional valuable information to the mode detection process. In efforts to keep the complexity of the detection algorithm to a minimum, a single value of the average number of satellites in view is used as an input. Figure 4.4 shows the distribution of number of satellites in view for each mode of transportation. On average, the number of satellites for the auto mode is the highest while it is the lowest for the streetcar mode. This indicates that GPS devices have more difficult time communicating with the satellites in the streetcar mode possibly because of the ceilings and other passengers that block the signals. Sample variance, however, is the highest with the walk mode. It is an intuitive result because the range of environments a walking person experiences is relatively greater. For example, if a person walks on the sidewalk closer to buildings, the signals will be blocked more often whereas vehicles tend to travel closer to the middle of the roads further away from the buildings.

![Distribution of number of satellites in view](image.png)

<table>
<thead>
<tr>
<th>Mode</th>
<th>Sample Mean</th>
<th>Sample Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto</td>
<td>8.8</td>
<td>1.6</td>
</tr>
<tr>
<td>Bus</td>
<td>7.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Streetcar</td>
<td>6.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Walk</td>
<td>7.5</td>
<td>3.4</td>
</tr>
</tbody>
</table>
The output of the classification process is the mode type. This thesis, first attempts to identify all modes of transportation with a universal NN. If the universal NN approach turns out to be successful, this would suggest that developing more specific NNs for detecting the auto mode from all other modes combined is possible.

### 4.5 Design of Neural Network Classifier for Mode Detection

The characteristics of GPS data streams vary depending on the transportation mode on which the GPS device is located. In this study, a pattern classification approach is adopted to identify the mode based on the characteristics of GPS data streams. NNs are used for this task because of their powerful nonlinear pattern classification capabilities (Welch, 1996). In this study, 60% of the collected data are used to train the NNs and the remaining 40% of the data are used as unseen data to measure the mode detection performances of the trained NNs.

In its most general form, a typical NN consists of multiple layers of neurons, connection weights among neurons, and associated nonlinear transfer functions within neurons. In this study, Multi-Layer Perceptron (MLP) NNs are used. More specifically, one-hidden-layer MLP NNs with M-P PEs described in Section 2.2.6 are used. The MLP is the most common supervised learning NN. Supervised learning refers to the case where there are input and output pairs of data for training the network. For the mode identification, the inputs of GPS data used include speed, acceleration and average number of satellites in view, while the outputs include the corresponding mode of transportation such as auto, streetcar, bus or walk. If the mobile device or the server can store (and transmit) both current and a few recent GPS readings, the additional past speed and acceleration values are also input to the NN to possibly improve the detection accuracy. The NN maps each set of input data to the most probable mode. The training process adjusts connection weights between neurons as an entire set of training input data, also known as

![Figure 4.4 Distribution of number of satellites in view for different modes of transportation](image-url)
an epoch, is repeatedly passed through the network. The detection system is implemented in Neuronsolutions™. (NeuroDimensions Inc. 2008) All NNs used consist of 3 layers of neurons connected in a feed-forward fashion, trained with the well-established, error-back-propagation algorithm described in Section 2.2.6. Figure 4.4 shows the configuration of the network.

Figure 4.5 Multi-layered feed-forward neural network with 4 available modes

Depending on the specifics of each scenario, the number of inputs ranges from 4 to 122 and the number of output classes is either 2 (auto or non-auto) or 4 (auto, street car, bus or walk). 2 output classes are used when the classification is for auto versus non-auto modes. 4 output
classes are used for classifying individual modes; i.e. auto, bus, street car or walk. As noted before, buses and streetcars do not operate on the same route in the GTA. However, the purpose of this 4-mode classification is to see if one general universal NN can be developed to classify the 4 modes that are not necessarily on the same route, assuming that the characteristics of the GPS data streams are not heavily dependent on the characteristics of the route on which the different modes travel. In other words, the focus in this case is to see if the 4 modes can be detected just by analyzing their dynamic performances (speed and acceleration) and GPS quality information (number of satellites in view).

It is noted that, with the increasing sensitivity of future GPS sensors and stronger signals from future GPS satellites, the Average Number of Satellites in View may become insignificant. In Figure 4.8, it seems the factor, as of now, is more significant than other factors. However, if the technologies regarding the sensitivity and the strength of the GPS units and satellites are improved, the mode detection may improve in accuracy even without the factor. In addition, the acceleration values are still significant in the same order of magnitude with speed values. As of January 2010, newly introduced GPS-embedded cellphones are also equipped with accelerometers that measure the acceleration value directly instead of deriving it with position values from GPS signals. The mode detection may improve in accuracy with the acceleration values from the accelerometers in the new cellphones.

As shown in Figure 4.5, the number of input variables varies depending on the monitoring duration and number of pings (n) per unit of time (as indicated as “Ping n” in Figure 4.5) in each scenario. For example, if the monitoring duration is 10 minutes and the number of pings or queries is 10 in that duration (one ping per minute), there are 10 speed values, 10 acceleration values and one value for average of the number of satellites in view. If the monitoring duration is one second with one ping, the NN would have one speed value, one acceleration value and one value for the number of satellites in view, totaling 3 inputs. However, it is noted that the one-shot-pinging scenario is intuitively unreliable, according to the study by Li et al. (2002).

As shown in Figure 2.16, hyperbolic tangent functions are used as the nonlinear transfer functions in the M-P PEs (McCulloch-Pitts processing elements). The number of neurons
(nodes) in the hidden layer is varied as 5, 10, 20 and 30. For different scenarios, different numbers of hidden nodes were used, depending on the number of input variables and output classes. As a rule of thumb, having too many hidden nodes forces the NN to memorize the input-output training data, and leads to poorer generalization. Having too few hidden nodes, on the other hand, gives the NN difficulties identifying and isolating the different classes. Therefore, it is desirable to keep the number of hidden nodes to a minimum without sacrificing the discriminating power.

From the 60 hours of collected GPS data, 60% are used to train the network and 40% are used for the validation process as unseen data. The final number of hidden nodes used is determined after varying the number of hidden nodes over multiple trials. The number of nodes was kept to a minimum so long as performance of the network did not degrade compared with networks with a greater number of nodes. A smaller number of hidden nodes generally results in better classification generalization as opposed to memorizing the peculiarities and noise in the training data. During the training processes, classification performance seemed to stabilize after 3000 epochs. If the error did not stabilize or continued to fluctuate after 3000 epochs, the number of hidden nodes is increased.

The suggested design of NNs also enables adaptability to the future changes. If a new mode of transportation is introduced to a route, the output layer of NN can easily be modified to include the mode as an additional output class. Once a new set of raw GPS data is collected on that route for a few days, a new NN can be trained based on the new data for that route. Alternatively, multiple universal NNs that are specific to different scenarios can be pre-built so that the scenario that best matches the new mix of modes can be used as soon as the new mode is introduced.

4.6 Experimental Scenarios and Results

The experimental scenarios are designed to validate and investigate the use of NNs for the mode detection processes. Each scenario, with its unique output classes, sampling frequency and
monitoring duration length, is associated with a unique NN. In this study, a total of 23 NNs are developed, trained and tested.

4.6.1 Universal Mode Detection Neural Network for 4 Different Modes

In this scenario, one classifier for all modes and all routes in the GTA is developed, i.e., a universal classifier. The spatial scope of the study includes the central business district, major roads in the suburbs and major highways in the GTA, as shown in Figures 4.1 and 4.2. As one arbitrary test case, a monitoring duration of 10 minutes and pinging frequency of one ping every 2 minutes are used in this scenario. Therefore, 5 speed values, 5 acceleration values and one average number of satellites in view are used as inputs. Outputs for this classification process are auto, bus, streetcar and walk. The development of this universal classifier serves as an indicator that whether using a NN-based classifier for mode detection problems in general is technically possible. The resulting mode detection rates are presented in Table 4.1. It is noteworthy that the auto mode is broken further into auto-highway and auto-arterial sub modes. Since the main objective of this universal NN is to stress test the NN approach for mode classification problems in general, it makes sense to make use of the additionally available information that is whether the data is from the arterial roads or highways. If the NN can classify finer divisions of classes, it is a positive sign that the NN-based approach may be more than adequate for classifying lesser number classes. (i.e. auto vs. non-auto) The detection rates show the percentage of unseen data the NN predicts into each mode category. For example, when the actual mode was the auto-highway mode, the network was able to recognize 60% of the data correctly as the auto-highway mode but 32% of the data was incorrectly identified as the auto-arterial mode. Similarly, the streetcar mode was detected correctly as the streetcar mode for 84% of the unseen data. Ideally speaking, if the classification process is perfect, the values on the main diagonal will be 100% and all other values will be zeros. Since the main diagonal values are the highest in each row in Table 4.1, it is a positive result that indicates the NN-based mode classifier is an acceptable approach. It is noted that Auto-Arterial mode results in higher detection rate than Auto-Highway mode. On arterial roads, the vehicles accelerate and decelerate more often than they are on highways. Therefore, the NN has higher chance of
detecting the fluctuating patterns in both speed and acceleration values and results in higher
detection rates.

Table 4.1 4 Mode detection rates (%)
(monitoring duration of 10 minutes and frequency of one ping every 2 minutes)

<table>
<thead>
<tr>
<th>Actual Mode</th>
<th>Auto-Highway</th>
<th>Auto-Arterial</th>
<th>Bus-Arterial</th>
<th>Streetcar</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto-Highway</td>
<td>60</td>
<td>32</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Auto-Arterial</td>
<td>9</td>
<td>87</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bus-Arterial</td>
<td>1</td>
<td>3</td>
<td>82</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Streetcar</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>84</td>
<td>7</td>
</tr>
<tr>
<td>Walk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>98</td>
</tr>
</tbody>
</table>

Varying the pinging frequency and the monitoring duration is expected to affect the mode
detection rates because a higher sampling frequency and a longer monitoring duration provides
more input factors that may help the NN to detect patterns from a broader input space.

4.6.2 Impact of Varying Pinging Frequency and Monitoring Duration

The results from Section 4.6.1 are promising because intuitively speaking, if the number of
output classes are reduced, it would only make the classification process easier. Ultimately, the
mode detection is primarily developed in this thesis for traffic monitoring where the classifier’s
main job is to filter out the auto mode from all other modes in order to use them as probes. In
this scenario, multiple NNs are developed with varying pinging frequencies and varying
monitoring duration. In this section, the main focus is detecting the auto-mode only, i.e.,
segregating the auto mode from all other modes lumped together. There are 2 types of detection
rates in this case; auto-auto and non-auto-non-auto detection rates. The auto-auto (a-a) detection
rates indicate what percentage of the actual auto mode data are correctly classified by the NN as the auto mode. The non-auto-non-auto (n-n) detection rates show the percentages of the actual non-auto mode data correctly identified as the non-auto mode. Higher values for both a-a and n-n show better mode detection capabilities of the NNs. Table 4.2 and Figure 4.6 show the results from the 20 different NNs developed under different conditions. The number of pings indicates how many times the mobile GPS device is queried and the monitoring duration indicates the length of the recent past time period from which queried data are used for the mode detection processes.

Table 4.2 Auto-auto and non-auto-non-auto detection rates (%) 
(under varying sampling frequency and monitoring duration)

<table>
<thead>
<tr>
<th>Monitoring Duration</th>
<th>5min</th>
<th>10min</th>
<th>15min</th>
<th>20min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a-a</td>
<td>n-n</td>
<td>a-a</td>
<td>n-n</td>
</tr>
<tr>
<td># of Pings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>92</td>
<td>65</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>94</td>
<td>66</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>67</td>
<td>96</td>
<td>68</td>
<td>97</td>
</tr>
<tr>
<td>15</td>
<td>68</td>
<td>97</td>
<td>72</td>
<td>98</td>
</tr>
<tr>
<td>20</td>
<td>69</td>
<td>98</td>
<td>74</td>
<td>98</td>
</tr>
</tbody>
</table>

*a-a refers to Auto-Auto Detection Rate and n-n refers to Non-Auto-Non-Auto Detection Rate*
Figure 4.6 indicates that the developed NNs perform better in general for n-n detection rates in all cases. It seems that the speed and acceleration characteristics of the non-auto modes are reasonably distinct and detectable. The main reason is that the non-auto mode contains multiple modes in it. Therefore, it is ready to detect different physical characteristics of different non-auto modes such as bus, street car and walk modes. Since NNs perform better when there are fluctuations in the data which would result in more distinctive patterns in the data, observing higher n-n detection rates is not surprising. Nevertheless, there seems to be some positive correlation of non-auto-non-auto detection rates with sampling frequency and monitoring duration length. For the auto-auto detection rates, on the other hand, both higher number of pings and longer monitoring duration steeply improve the detection rates. Higher frequency results in higher auto-auto detection rates because it gives fine sampled GPS data trail to the NN,
for any given duration length. Higher monitoring duration length results in higher auto-auto
detection rate because there is a higher chance of larger fluctuations in speed and acceleration
values during longer time intervals. Therefore, speed and acceleration patterns of the different
modes become more vivid and distinguishable. For example, there is a higher chance of
fluctuations in speed and acceleration patterns with 10 pings in 10 minutes than 10 pings in 10
seconds. From Figure 4.6, it seems that the monitoring duration (with its steeper slope) affects
the a-a detection rates more than the number of pings. This result means that, with everything
else being equal, focusing on increasing the monitoring duration is more effective than focusing
on increasing the sampling frequency.

4.6.3 Route-Specific Auto Mode Detection Neural Network

In this phase of the research, it is hypothesized that a more specialized classifier may improve
the performance of the mode detection. It is also suspected that some routes might have distinct
operational characteristics, such as geometry or traffic congestion levels, that would directly
impact speed and acceleration profiles. In the GTA, a unique bus route Rocket 192 is chosen that
operates between the Kipling subway station and the Pearson International Airport, as shown in
Figure 4.3. Most of the route runs on Highway 427 which is a unique highway stretch in that it
accommodates both auto traffic and bus transit. For this particular route, a route-specific NN is
developed with only 2 modes: bus-highway and auto-highway modes. In this test case, a
monitoring duration of 5 minutes and 5 pings (one ping per minute frequency) are used. The
parameter values are chosen so that the result can be compared with the results with the same
set-up in Table 4.2. With respect to other parameter values in Table 4.2, 5 minutes of monitoring
duration with 5 pings are the worst case scenario. If a route-specific NN can perform
significantly better than the non-route-specific NN even in the worst case scenario, it would be a
positive sign that shows the route-specific NN approach is an effective one. The resulting
classification performance shows significant improvement. The auto-auto detection rate is found
to be 95% and bus-highway-bus-highway detection rate is found to be 99%. Compared with the
results for the corresponding cases in Table 4.2, auto-auto detection rate is improved from 60%
to 95%. From the result of this simple experiment it is clear that the route-specific NN has
higher potential for better classification accuracy compared to the non-route-specific NN.

The major interest of this section is to test the performance of route specific NNs. The route is fixed and known for this section that there is no need for a map matching algorithm. In the field applications, the data first need to be matched to a corresponding link before it is used for mode identification purposes. This thesis focuses on the mode identification issues assuming the map matching is previously processed and available for access in a database.

Ideally, if each route can have its own route-specific NN, the mode detection rate will increase. With improving GIS technologies, link-specific NN can be implemented in the GIS database. Alternatively, due to high implementation costs and complications, multiple universal scenario-specific NNs can be built first and the NNs can be applied to other routes with similar mix of modes. The latter method assumes that as long as similar mix of modes are present on different routes with similar environmental conditions (such as high/low rise buildings, number of lanes etc.), a single universal NN is capable of detecting the modes on the different routes and can result in similar performances.

4.6.4 Auto Mode Detection Neural Network during Peak and Non-Peak Period

In this scenario, a test case with a duration of 10 minutes with 5 pings (one ping every 2 minutes) was used for both peak and non-peak periods. The parameter values are arbitrarily chosen to be fixed. The main interest in this section is not in the particular values of the parameters but rather the comparison of effects on the mode detection between the peak and non-peak periods while those parameter values are fixed. The study corridor is Eglinton Avenue between Jane Street and Yonge Street. The available modes on this route are bus-arterial, auto-arterial and walk modes. The peak period data are collected between 4 pm and 7 pm on typical weekdays. During the peak period, auto-auto detection rate is found to be 92% and non-auto-non-auto detection rate is found to be 97%. During the non-peak period (excluding both the AM peak and the PM peak of typical weekdays), the auto-auto detection rate is found to be 66% and the non-auto-non-auto detection rate is found to be 95% as seen in Table 4.2. It is interesting to note that the NNs find
it easier to detect modes during peak periods. During the peak periods, both bus-arterial mode vehicles and auto-arterial mode vehicles frequently accelerate and decelerate due to the recurrent high congestion levels on the roads. The 2 modes exhibit different profiles of acceleration and deceleration due to their unique physical characteristics. The resulting fluctuations in the data stream may help the NN to detect the patterns better. This is a useful and welcome finding because almost all traffic management strategies are more crucial during the peak congestion periods.

4.6.5 Specification of NNs in Neurosolutions

Each NN developed in this chapter is associated with various parameter values in Neurosolutions. In other words, all the main components of the software package shown in Table 2.2 are associated with specific parameter values that specify the exact set up of the NN in the software. Section 7.1 recommends a “auto vs. non-auto” mode detection scenario with a monitoring duration of 10 minutes with 10 pings. The parameter values of the scenario are included in Appendix A.

By analyzing the weights of the input layer in the developed NN, it is found that both speed and acceleration values are nearly equally important both during free flow and in congestions while the weights for the speed values tend to be slightly greater. Speed values help the NNs to capture the maximum steady speed in a free flow and acceleration values help NNs to capture the relatively higher acceleration and deceleration profiles of private automobiles from other modes.

4.6.6 Sensitivity Analysis

2 different types of sensitivity analysis methods are conducted in this section. The first type of the sensitivity analysis methods tests the effect of varying speed and acceleration values with an arbitrarily set percentage of errors on the a-a detection rate. The second type of the sensitivity analysis methods monitors the input factors to the NNs and tries to determine the most
significant input factors that influence the mode detection rates.

4.6.6.1 Sensitivity Analysis with Varying Speed and Acceleration Values

Table 4.3 and Figure 4.7 show the effect of percentage changes in speed and acceleration values on the average mode detection rates, a-a and n-n for the 10 pings for 10 minutes scenario. As the errors in speed and acceleration values are artificially increased for all collected data (assuming normal distribution), the detection rates deteriorate. A 10% change in speed according to the normal distribution need to be clarified first. If the recorded GPS speed is 100 km/hour, then the 10% of the 100 km/hour (=10 km/hour) is assumed to be equal to $6\sigma$ of the error population centered on the 100 km/hour on the speed distribution axis, assuming it would cover 99.7% of entire error population. The well-known Monte Carlo method is used to generate the new set of data that would simulate the "noisy" GPS data. It is noted that 10% change in speed values always results 10% change in acceleration values because acceleration values are linearly dependent on the speed values where they are derived from. The results are intuitive in a sense that noisy GPS data are not as useful as high quality GPS data in transportation mode detection processes.

Table 4.3 Sensitivity analysis on speed and acceleration

(20 pings for 20 minutes)

<table>
<thead>
<tr>
<th>% Change</th>
<th>Detection Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>87</td>
</tr>
<tr>
<td>10</td>
<td>76</td>
</tr>
<tr>
<td>30</td>
<td>68</td>
</tr>
<tr>
<td>50</td>
<td>62</td>
</tr>
</tbody>
</table>
4.6.6.2 Sensitivity Analysis on Input of NNs

Figure 4.7 Average mode detection rate vs. varying % change in both speed and acceleration values

(20 pings for 20 minutes scenario)

Figure 4.8 shows how much input factors; s1(current speed), s2(previous speed) ... , s10(the oldest speed), a1(current acceleration), a2(previous acceleration), ..., a10 (the oldest acceleration), influence the a-a mode detection rates as each factor is varied within a ±15% range. The graph shows the maximum fractional change of the a-a detection rates as the corresponding input variable varies within the ±15% range. Figure 4.8 for the 10 pings for 10
minutes scenario shows that speed values are generally slightly more significant than acceleration values for both auto and non-auto modes.
4.6.7 Discussion on Low Satellite Visibility Regions

In dense urban areas and tunnels, GPS signals are weak or not available. Byon (2006) in his
master’s thesis found that traffic monitoring is not feasible in the CBD due to the signal errors resulting from high-rise buildings. With increasing sensitivity of GPS chips and other improving satellite-based positioning systems, such as the European Union’s Galileo and the Russian GLONASS, the visibility of the positioning devices will increase over time. Such future technological improvements will reduce the line-of-sight problems of the GPS. In the future, with those technological improvements, by increasing the pinging rate and building a unique NN near dense urban areas, the mode detection rate may improve. Since more people benefit from services in dense urban areas, higher pinging costs can still be cost-effective or maybe will not be very expensive in the future. For the regions (i.e., tunnels) where GPS signals are not available, traffic monitoring could continue to be based on loop detectors or pseudo satellites (that transmit similar signals actual satellites produce) can be installed where the signals are not available.

4.6.8 Effect of Inclusion of Data from Incorrectly Identified Mode on Travel Time Estimation

In this section, the effect of an inclusion of incorrectly identified mode data on travel time is discussed. This thesis has illustrated the mode detection techniques with some of major transportation modes that are auto, bus, streetcar and walk modes. In addition to those modes, a bike mode has been getting more attention as an alternative to the other modes. Therefore, this section attempts to also work with the bike mode. It is noteworthy that this thesis does not attempt to treat certain modes of transportation over others or discriminate one from others either. The methodology presented in this thesis are applicable to any mixed modes of transportation. This section tests the effects of inclusion of data from the non-auto (bike) mode that is mistakenly identified as the auto mode on travel time estimations. Inclusion of the speed data from the slower mode is expected to increase the estimated travel times.

For this study, 1 hour of data from each mode (bike and auto) is collected using the GPS data logger (Sanav Inc., GL-50-BT). The data are collected on a typical weekday in September 2009 during a non-peak period. The study region is chosen in a typical west-end outskirt of the GTA.
shown in Figure 4.9 where the data collection from both bike mode and auto mode are allowed due to its bike-friendly environments (e.g. bike lanes and minimal traffic).

![Study region for the data collection of bike and auto modes](image)

**Figure 4.9** Study region for the data collection of bike and auto modes

The distance of the route in Figure 4.9 is 11.8 km. The true travel time is 9.3 minutes which is measured from a stop watch. Table 4.4 shows the result and Figure 4.10 shows the corresponding graph. “50% of Bike Data” refers to the case where 50% of the collected data from the bike mode and 50% of the collected data from the auto mode are used to find the average speed which divides the route length to estimate the travel time.
The absolute % error is computed against the true travel time collected from a stop watch.

\[
\text{Absolute % Error} = \left( \frac{\text{True Travel Time} - \text{Estimated Travel Time}}{\text{True Travel Time}} \right) \times 100
\]  
(Equation 4.2)

<table>
<thead>
<tr>
<th>% of Bike Data</th>
<th>Estimated Travel Time in Minutes</th>
<th>Absolute % Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Travel Time</td>
<td>9.3</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>9.8</td>
<td>6.1</td>
</tr>
<tr>
<td>20</td>
<td>10.9</td>
<td>17.4</td>
</tr>
<tr>
<td>30</td>
<td>12.3</td>
<td>33.3</td>
</tr>
<tr>
<td>40</td>
<td>14.6</td>
<td>58.2</td>
</tr>
<tr>
<td>50</td>
<td>15.3</td>
<td>65.1</td>
</tr>
<tr>
<td>60</td>
<td>19.5</td>
<td>110.7</td>
</tr>
<tr>
<td>70</td>
<td>21.2</td>
<td>128.9</td>
</tr>
<tr>
<td>80</td>
<td>24.9</td>
<td>169.0</td>
</tr>
<tr>
<td>90</td>
<td>32.7</td>
<td>252.0</td>
</tr>
</tbody>
</table>
As more % of bike data are incorrectly identified as coming from the auto mode, the travel time estimations deviate more from the travel time experienced by the auto mode. The results show that inclusion of slower mode being incorrectly identified as auto mode increases the estimation error.

Figure 4.10 Absolute % deviation on travel time estimations vs. % of bike data inclusion

4.7 Summary, Conclusions and Future Research

This chapter attempts to use a relatively short trace of GPS data to identify and detect...
transportation modes in real time (auto, bus, street car and walk). A pattern classification approach is used for identifying the modes of transportation utilizing the strong discriminatory capabilities of ANNs (i.e., this study analyzes the feasibility of using NNs for transportation-mode detection using speed and acceleration profiles from mobile GPS devices such as, but not limited to, AGPS phones). This chapter is designed to progress in consecutive phases. The first phase focuses on assessing the potential of NN-based mode detection in general using the GPS data. The second phase examines the impact of varying pinging frequency (data sampling resolution) and monitoring duration on the quality of the auto mode identification. The third phase tests the hypothesis that a more specialized classifier, a route-specific NN, may improve the auto mode detection performances. The last phase compares the mode detection accuracies during peak and non-peak periods. The experiments conducted in this study use 60 hours of GPS data collected from various routes and different modes of transportation across Toronto. The study leads to several interesting findings. Firstly, the universal 4-mode detection NNs proves to be a promising mode classifier. Using a GPS data stream that is 10 minutes long, sampled (pinged) once every 2 minutes, the developed NN demonstrates 60% correct classification rate for auto-mode on highways, 87% for auto-mode on arterials, 82% for buses on arterials, 84% for streetcars, and 98% for walk mode. These figures are very encouraging. It is evident that the detection performance improves as the speed and acceleration profiles become more distinct. Secondly, increasing both the sampling (pinging) frequency and monitoring duration directly improves the mode detection rates. As can be intuitively expected, higher sampling frequency and longer monitoring duration result in higher mode detection rates. Quantitatively however, it seems that it is significantly easier to detect the non-auto mode due to its more pronounced speed and acceleration patterns compared to the auto mode. The n-n rate ranges from 92% to 97% in all scenarios as shown in Table 4.2. On the other hand, the corresponding a-a ranges from 53% (2 pings per minute) to 87% (20 pings over 20 minutes). It is evident that higher pinging frequency and longer monitoring duration are more crucial for detecting the auto-mode than for the non-auto mode. Moreover, the route-specific mode detection NN is found to perform better than the universal 4-mode detection NN. Using the same pinging frequency and monitoring duration, the route-specific NN produce auto-detection accuracy of 95% compared with the corresponding 60% accuracy for the universal NN. Lastly, it
is pleasantly surprising to find that the mode classification performance significantly improves during rush hour compared with off-peak periods. Using the same experimental variables, the auto-mode detection rate for the peak period is 92% compared with 66% during the off-peak period. Most traffic surveillance and management methods suffer during congestion. It is a welcome finding in this case that the mode detection performance improves during rush hour when it is most needed.

In terms of future research, the next steps include; estimating and distributing travel times along different routes, and using the resulting travel times for dynamic route guidance. The mode detection techniques can also be used for other niche applications such as Advanced Transit Passenger Information Systems, focusing the detection and forecast of travel times on transit vehicles only.
Chapter 5
Data Fusion of Different Sensors

5 Data Fusion of Different Sensors

For a particular section of a road network, at any point in time, there could be multiple sources of quantitative and qualitative traffic information available. Quantitative sensors are usually strictly hardware-based and include loop detectors and GPS devices that produce numerical data. Qualitative sensors are usually processed data and include the traffic department’s websites and radio broadcasts that produce subjective categorical data based on hidden processes. Each sensor is characterized by a specific level of error and sampling frequency. It is a challenge to combine and utilize multiple sourced data for estimating real-time traffic conditions. Most available data fusion methods involve waiting until all sensors become available, which slows down the sampling frequencies. It may not be suitable for traffic monitoring because the most recently captured information loses its value as time passes. In the first phase of this chapter, by using single-constraint-at-a-time (SCAAT) KFs, this chapter combines multiple data sources of a section of a highway in the GTA as soon as any single sensor becomes active and estimates the current traffic conditions. However, in real-life, true traffic conditions are unknown because all sensors have associated errors with them. In the second phase of this chapter, a micro-simulation package is used in order to have access to the true traffic conditions of a simulated environment that has been calibrated for a particular road section in the GTA. Then, the performance of the developed SCAAT filters are compared with the true traffic conditions under different sampling strategies with varying number of probes and varying sampling frequencies of sensors. The use of SCAAT filters is found to be an effective method to simultaneously fuse the data and estimate current traffic conditions. It is noted that this thesis monitors for the speed values of the vehicles which can be converted to travel times. In transportation engineering field, there are well
established generalized relationships among speed, density and flow rate variables. Therefore, the speed value can also be converted into density and flow rate values that are other major traffic monitoring related variables.

5.1 Introduction

It is becoming increasingly common in urban road networks to have different types of sensors providing independent traffic measurements on the same road link. However, such sensors usually vary in the method they measure traffic resulting in varying degrees of accuracies in their output. While loop detectors collect frequent traffic information at a limited set of fixed points along a given road section, probe vehicles can provide continuous traffic measurements using GPS along the same section but at a lower frequency than loop detectors. In addition, on-line websites of traffic departments provide traffic information based on a combination of video cameras, loop detectors and human observation. Also, radio stations (e.g. AM 640 Toronto) operate their own traffic monitoring helicopters and provide live traffic reports based on observations by air-borne personnel and occasional input from listeners. Figure 5.1 illustrates the potential multiple sources of real-time traffic information along a section of a road network.
In recognition of the emerging availability of multiple sensors providing independent traffic measurements on the same road links, several research efforts are recently made to explore different research possibilities. For example, Kwon et al. (2007) (Section 3.2.1) investigate the impact of varying the number of probe runs to achieve a level of measurement accuracy similar to that of loop detectors. Qiu and Ran (2008) (Section 3.1.7) attempt to estimate traffic conditions using cellular phones with KF techniques in efforts to utilize existing facilities for traffic monitoring. They compare their estimations with the loop detectors treating the loop detectors as the ground truth data sources. However, no efforts are made to combine the multiple sources of data. Herrera and Bayen (2008) (Section 3.2.2) propose and investigate new algorithms that combine data provided by mobile sensors and location-fixed detectors, to identify the traffic flow status. The authors recognize that the availability of multiple sensors present an opportunity to obtain more accurate estimates of traffic conditions by fusing information from the different sensors. However, there is a need for a versatile methodology that can fuse data
from not only loop detectors and probe vehicles but also other available data sources, which may not necessarily have the same unit or accuracy. In fact, some information may be based either fully or partially on the human judgment, such as the information provided by websites of traffic departments (e.g. colour-coded lines) and radio stations (e.g. “moving well”, “moving slowly”, “extremely slow” or “not moving”). For an effective integration of multiple data streams while maximizing the overall benefit of all available data, there is a need for a fusion method that can integrate such data. The objective of this chapter is to develop a method of fusing data from various sources using single-constraint-at-a-time (SCAAT) KFs. The background information and formulations regarding the KF and SCAAT KFs are presented in Section 2.3.

It is noted that different sensors sample data with different time intervals. In theory, if there are multiple sensors monitoring the same section of the road continuously at the same time step, the most accurate sensor should be used over others not needing the multiple-sensor data fusion process. However, due to the higher costs associated with sampling at all times with the most accurate sensor alone, and due to different sampling frequencies of different sensors, at any moment in time, there usually is none or only 1 sensor that is sampling the data. Therefore, the focus in this chapter is about how to make use of the most recent single sensor into the traffic condition estimations even if the single sensor is not the most accurate one. The most accurate sensor is idling and its previous step data represents the past traffic condition when compared with the most recent (inaccurate) single sensor. For example, if I saw it was not raining outside as I arrived home 2 hours ago; I can assume it is not raining now because there is no updated data available. If I now hear a raining sound in my room with no windows, even before I walk out to see if it is really raining, I can start to assume it is raining even though seeing it is more accurate than hearing it.

5.1.1 Robust Traffic Information with Multiple Information Sources

Even though this chapter is inspired by the problem of having multiple sources of data, there are times when there are no data available from the main sensor (i.e. loop detectors). The data
fusion methodology presented in this chapter also can fill the data gaps with the aid of other secondary sensors. Figure 5.2 shows a traffic monitoring status of a typical weekday-afternoon-peak period on the Highway 401 in the GTA. The presented information is based on loop detectors alone. It is common to have the information gaps shown in Figure 5.2, at unexpected times and locations. If there are other sensors or information sources that are operational for the missing time slot and the location, even if those are less accurate than the currently off-line sensor, incorporating the new data would reduce the data gaps and increase the robustness of the monitoring system via the redundancy. The trajectory labeled as “Probe Vehicle Trip Profile” shows a moving probe vehicle that is moving eastbound over multiple loop detectors on Highway 401 from Highway 27 to Yonge Street. The probe vehicle in the first phase of this chapter produced the trajectory.

![Figure 5.2 An example of missing information of loop detectors](image-url)
5.1.2 Calibration of SCAAT Filter

There are multiple parameters that can influence the performance of SCAAT filters for estimating traffic conditions. The main focus of this chapter is to test the feasibility of using such filters for data fusion purposes under different sampling scenarios and not searching for the optimal set of parameter values. Therefore, constant scalar values are used for different parameters.

5.1.3 Used Parameter Values

In this chapter, the following parameter values are used for the SCAAT KFs.

A=1: There is no reason to believe the traffic condition is changed unless there is a new measurement.

B=0: There is no known external control input factor that affects speed measurements.

P values = sensor specific error variances: Each sensor has its unique specific error variance value.

Q=1e-5: There is no known process noise. However, to give the filter a room for future modifications, a very small value is assigned to the parameter.

H=1: The measured unit is speed and the desired unit for traffic monitoring is also speed. Therefore, there is no need to convert the sensor measurements.

5.2 First Phase: Data Fusion for Highway 401 in Toronto

The first phase of this chapter attempts to fuse data from different data sources such as a GPS
probe, loop detectors, radio broadcasts and the traffic department’s website. As a GPS equipped probe vehicle travels on a highway that is monitored by those other data sources, a general method of combining both quantitative and qualitative data is needed for implementing a reliable traffic monitoring system.

5.2.1 Data Collection

For the initial application and evaluation of the SCAAT KFs developed, different types of data are collected from 4 different sources: floating car survey using a GPS unit, 40 loop detectors across the Highway 401 in Toronto, radio broadcasting from AM 640 Chopper Traffic channel and the on-line freeway management system of the Ontario Ministry of Transportation (MTO). For all sensors, the data collection period spanned from March 31st 2008 to April 11th 2008 for the morning peak (7 to 10), non-peak (11 to 2pm), and afternoon peak (4pm to 7pm) on both weekdays and weekends. A 10 km section on Highway 401 between Highway 427 and Yonge St. was selected because it is one of the most travelled freeways in Toronto with various traffic monitoring sensors in operation.

5.2.2 Description of Available Data Sources

Among the various data sources, each data type has a unique error variance associated with it. A GPS device in general is capable of the lowest variance and the lowest sampling interval of 1 second. A loop detector provides the second lowest variance with sampling interval of 20 seconds. The radio broadcasting and the traffic department’s website are updated at larger time intervals with larger variances. In this study, the variance of each sensor is roughly estimated based on the combination of the device specifications and subjective judgments. The focus of this chapter is not analyzing the specifications of the information sources but it is rather presenting a methodology of how the qualitative information can be converted into numerical values and be used for the data fusion process.

One may think that using the most accurate sensor alone and discarding all other sensors without
any data fusion is idealistic. In this thesis, the most accurate sensor is the GPS device. GPS
device can produce data with varying sampling frequencies. Even though second-by-second
sampling is possible, such approach may be ineffective due to following reasons:

1. High data communication cost
2. Lower battery life of the mobile GPS device
3. Highly fluctuating micro-level data that do not necessarily represent the general traffic
   conditions

Therefore, the application of data fusion techniques still gives more options to transportation
engineers that help utilizing multiple-sourced information, lowering the combined cost and
filling the data gaps when some sensors are temporarily out of service.

5.2.2.1  GPS

GPS devices in general, including AGPS cell phones, can collect and store its location and speed
information at the maximum rate of 1 Hz (once per second). With emerging SirfStarIII GPS
chipsets, the devices are decreasing in size and becoming more sensitive. In this study, a
SirfStarIII chip-based, real-time data logger is used. The device is capable of both real-time
monitoring and off-line post-processing. It uses the same SirfStarIII chip used in most AGPS
cell phones. By sampling the GPS trip data with varying frequencies, it is possible to analyze the
effect of GPS sampling rate on the data fusion performance. Typical GPS units have 0.1 knot
(about 0.1 mph) rms. (Product Specification, GL-50, 2008) The variance is roughly converted to
0.04 km²/hour² assuming a normal distribution. 90 hours of GPS data is collected from 1 probe
vehicle and 3 continuous trips were made each day for morning peak, non-peak and afternoon
peak time slots.

5.2.2.2  Loop detector

Ki and Baik (2006) report that their developed loop-detector-based vehicular speed measurement
model can measure the speed within ±5% of error. Assuming the free flow on highways to be
100 km/hour, equivalent errors are within ±5 km/hour. With the speed population being
normally distributed with 6 standard deviations covering 99.7% of the total population between
the 10 km/hour (±5 km/h) range, the standard deviation is assumed to be 10/6 = 1.66. The
variance of the loop detectors is assumed to be \( \sigma^2 = 1.66^2 = 2.8 \text{ km}^2/\text{hour}^2 \).

5.2.2.3 Traffic Department’s Website

MTO runs a real-time traffic monitoring system known as “COMPASS”, which provides traffic
summary information on its website. The system uses various sensors that MTO operates,
including traffic cameras, loop detectors, etc. Visitors to the web site can find out about traffic
conditions in 3 categories: “Moving Well”, “Moving Slow” and “Moving Very Slow.” In
absence of the corresponding speed ranges, the 3 colour coded categories are arbitrarily
converted to be 80 to 120, 40 to 80, and 0 to 40 km/h respectively. Values in each range are
represented by their mean (mid-point) and assumed to be normally distributed around the mean.
This is rather a strong assumption made to facilitate using the KF. Expert-based systems such as
the traffic department’s website have unknown processes that results in their current estimations.
The main idea of this thesis is not to find out how they came up with such estimations but rather
present a methodology of how one can incorporate subjective information from credible
information sources that are not numerically (with their statistical distributions) available for a
KF-based data fusion process. Using the statistical fact that 3\( \sigma \) to the left and right of the mean
covers 99.7% of the total speed population assuming a normal distribution, the range of each
speed category is assumed to be equal to 6\( \sigma \) of the assumed normal distribution of the speed
values of that category with the mid-point in the category being the mean. The variance
associated with the information from this web site for each category is assumed to be \( \sigma^2 = 45 \text{ km}^2/\text{hour}^2 \). The data updating frequency is once every 3 minutes which is set by MTO.

5.2.2.4 Radio Broadcasting (AM 640 Chopper)
In Toronto, the radio station, AM 640, broadcasts real-time traffic conditions throughout the day with higher updating frequencies during the rush hour period. The station operates helicopter(s) and observes the traffic condition visually. Verbal information are highly subjective and qualitative. In order to quantify the subjective information, 5 survey participants are recruited. After listening to 3 days of the rush hour broadcasting totaling 6 hours, a subjective agreement among their perceptions are categorized into 4 categories; above 75km/hour (“looking good”, “clear”), between 50 and 75 km/hour (“starting to build”, “not bad”, “some delays”), between 25-50km/hour (“busy”, “delayed”) and below 25km/hour (“stopped”, “congested”, “not moving”). (as shown in Table 5.1) Similarly as in the section 5.2.2.3, the categorization was done arbitrarily to illustrate how one can convert subjective information and still can feed such data into numerical data fusion using KF. The main philosophy behind this is not to waste valuable information simply because they are not numerical data. This thesis suggests, the subjective data can be converted into numerical values at least arbitrarily, and still be fed into the data fusion process with larger variance values. This approach is useful in a sense that the monitoring system will have reduced number of data gaps resulting in a more reliable and robust monitoring system.

With the similar method used for the traffic department’s web site data, the variance associated with the information from the AM 640 is converted to 18km$^2$/hour$^2$. The data updating frequency is roughly once every 10 minutes during rush hours.
Table 5.1 Categorization and quantification of verbal information from radio broadcasts

<table>
<thead>
<tr>
<th>Level</th>
<th>Phrases or Sayings on Radio</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-looking good</td>
<td>75 km/h and above</td>
</tr>
<tr>
<td></td>
<td>-clear</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-starting to build</td>
<td>50–75 km/h</td>
</tr>
<tr>
<td></td>
<td>-not bad</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-some delays</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-busy</td>
<td>25–50 km/h</td>
</tr>
<tr>
<td></td>
<td>-delayed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-building</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-stopped</td>
<td>0–25 km/h</td>
</tr>
<tr>
<td></td>
<td>-congested</td>
<td></td>
</tr>
</tbody>
</table>

5.2.3 Design of Experiment

In this section, various scenarios are presented in order to analyze the performance of SCAAT KF for data fusion. The KF is both a data tracking and a forecasting tool. In the case of SCAAT KF, as new data become available from the most recent sensor measurement, the filter generates its best guess for the current state of the traffic.

In order to evaluate the performance of different SCAAT filters, the ground-truth to compare with the SCAAT estimations is required. Because one can never know the true condition of the
road, (no sensor is perfect) Qiu and Ran (2008) assumed loop detectors to be their most accurate sensor and assumed it as their ground-truth. As a probe vehicle runs on a freeway, among all the other sensors, the GPS device onboard displays the most accurate data (from the perspective of the vehicle, assuming the end-user is the driver of the probe vehicle for now) because it provides second-to-second measurement and runs with the probe vehicle, whereas other sensors are only providing general traffic-flow conditions on relatively large sections of the road. Therefore, by comparing the SCAAT KF estimates with the actual GPS readings (treating the GPS as near-true trip data), the accumulated errors with respect to the probe vehicle can be computed. At the same time, it is noteworthy that having the SCAAT-filtered estimation being too accurate is not always beneficial because the probe vehicle only experiences micro-level traffic conditions near itself. The error measurements are used to compare the behaviors of different SCAAT filters and provide insight into how varying GPS sampling rates affect the SCAAT-filtered estimations. Depending on the requirements of applications, one may prefer more stable estimations (GPS data tend to fluctuate a lot). These estimations would deviate more from the probe’s GPS data, resulting in high error measurements with respect to the GPS data while the stable estimations maybe more suitable for other ITS applications.

In this chapter, the error per unit time (EUT) is defined as follows. For any Kalman-filtered output, EUT can be computed against the GPS data which are considered to be the closest to the truth data.

\[
\text{Error per Unit Time} = EUT = \frac{\sum_{i=1}^{k} \sqrt{(\hat{\mathcal{X}} - X)^2}}{k \times \text{Total Monitoring Time}}
\]  

(Equation 5.1)

where, k is the total number of measurements

\( \hat{\mathcal{X}} \) is the SCAAT-filtered speed estimation and

\( X \) is the ground truth (the actual GPS trip data)

It is interesting to see how SCAAT KF estimates would change with varying AGPS probe sampling frequencies while the sampling rate of other sensors stays constant. Intuitively, as the
AGPS probe is sampled at a lower frequency than 1 Hz (once per second), the device will act as another useful loop detector traveling along the road. Theoretically, it would be ideal to have access to the true traffic conditions. Only then, error measures can properly be applied. Fortunately, traffic simulation packages can provide the ground truth information since they have access to second-by-second trajectories of all vehicles in their simulation environments. In the second phase of this chapter, Paramics is used to access the ground truth, simulate GPS and loop detectors, then compare the performances SCAAT KF-based data fusion of different sensor sampling strategies.

5.2.4 Results of Data Fusion

Figure 5.3 shows the output of the data fusion of multiple sensors on one particular weekday in the afternoon rush hour period, on Highway 401 between Highway 27 and Yonge street. As a probe vehicle runs on the freeway, the SCAAT filter fuses the GPS data with those of the other sensors that are monitoring the vicinity of the probe vehicle. The data points with markers are the actual readings from the different sensors while the lines represent the estimated speeds using the SCAAT filter, with the frame of reference fixed on the moving probe vehicle. As the GPS sampling interval is increased from 1 second to every 20 sec to every 3 minutes, the SCAAT filter results show less fluctuation. This is an intuitive result. As a GPS device is sampled at longer intervals, it can be viewed as a “moving” loop detector sampled at longer intervals. The SCAAT filtered estimations do not fluctuate as heavily as the ground truth data (GPS data @ 1 Hz) do, and simultaneously the filter successfully fuses other sensors’ data streams. It is noteworthy that, with slowly fluctuating speed estimations, more reliable travel times can be computed because this may filter out the micro-level experience of the probe vehicle. This is a good example of a high sampling frequency not necessarily guaranteeing better estimations even though it may actually cost more. Depending on the ITS applications and the required updating rates, the end-users should decide their desired responsiveness of the filter.
Figure 5.3  Data fusion of various sensors and comparison of SCAAT filters with varying GPS sampling frequencies.

(From 4.00 to 4.15 pm on Highway 401 Eastbound Collector on Wednesday April 2, 2008)

Based on the 2-week worth of collected data, Figure 5.4 shows the average relative EUT of SCAAT filters, with respect to the average EUT of GPS data sampled at 1 Hz. When the GPS data are sampled at the frequency of 1 Hz (once per second), the SCAAT filter is saturated with highly precise GPS data, every second, resulting in the SCAAT estimations that follow the pseudo ground truth (GPS data) closely. It is noteworthy that, even though the GPS data-based SCAAT KF output at 1 Hz is compared against the same GPS data at 1 Hz, there are errors, because the SCAAT KF output carries a “momentum” at each iteration. As the GPS data are sampled at longer intervals, the impact of other sensors on the SCAAT filtering results increases (because there are relatively less amount of GPS data in the data fusion process) and the estimations deviate more from the pseudo ground truth. It should be emphasized that there is no single optimal GPS sampling rate universally appropriate for all applications. High sampling rates will result in greater fluctuations in speed estimations that are suitable for micro-level applications, while low sampling rates will result in slow-fluctuating estimations that are more suitable for estimating travel times of the general traffic at a macro level minimizing the
influence of the micro-level experience of the probe vehicle. The major benefit of having a GPS device as an input to the multi-sourced data fusion is that, (assuming the modification of the sampling rate is relatively easier with GPS devices) by varying the GPS sampling frequency, one can fine-tune the responsiveness of the SCAAT estimations to any preferred levels. When the GPS unit is sampled every 20 seconds or less often, the major sensor that drives the estimation process is no longer the GPS unit because, in this case, loop detectors are being sampled more often than the GPS device. Similarly, by varying the sampling rates of other sensors, the resulting SCAAT estimations will behave differently because each sensor has a unique variance value. The major advantage of adopting this SCAAT KF-based data fusion for traffic monitoring is that any change in the sampling rate or addition/removal of any new or old sensor can be handled with no additional major modifications to the filtering framework. The SCAAT filter simply uses the single most recent measurement from any existing sensor, and it is not bound to any particular sensor or pre-determined sampling frequency. In practice, the flexible nature of SCAAT filtering can enable robust and easy-to-implement traffic monitoring systems.

Figure 5.4 Relative EUT of SCAAT filters with varying sampling intervals with respect to SCAAT with GPS sampled at 1 Hz
5.2.5 Traffic Monitoring GIS Database

5.2.5.1 Prototype of Traffic Monitoring GIS Database

After successfully obtaining the fused data from various data sources, a procedure for maintaining a GIS database and estimating travel times is still required. The fused or SCAAT-filtered data, need to be disaggregated into their corresponding road sections, also known as “arcs” in a GIS database. The “hot-zone” map-matching algorithm by Byon et al. (2006), is used because data are collected only from Highway 401, a pre-determined route. By using associated road links (arcs) with their lengths (distance attribute) in the GIS database, the estimated travel times can be computed from simply dividing the length by SCAAT-estimated speed values. A prototype of a real-time traffic monitoring GIS database is developed using the ArcObjects in the ArcGIS platform as shown in Figure 5.5.

Figure 5.5 A prototype of real-time traffic monitoring GIS database
5.2.5.2 Evaluation of Travel Time Estimation

In order to evaluate the performance of the travel time estimations from Section 5.2.5.1, a particular link on Highway 401 Eastbound Collector lane (from Kipling Avenue to Dufferin Street) is monitored during the afternoon rush hour for 10 days. The actual length of the link from the GIS database is divided by the SCAAT KF estimated speed. The 90th percentile deviation in percentage from the pseudo ground truth (GPS data from the probe vehicle) is 5%, 9%, 11% and 14% for the GPS sampling rate of 1 sec, 20 sec, 1 min and 3 min, respectively. (Table 5.2) The pseudo ground truth is measured directly from the time tags of GPS data as the vehicle enters and exits the link. It is noteworthy that Table 5.2 shows the error of 5% for 1 second sampling interval scenario. It is because of the momentum effect (mentioned earlier in Section 5.2.4) of KF process that still results in some errors even if the sampling rate is 1 Hz which is the same sampling rate as the collected GPS data.

Table 5.2 Travel time estimation deviation from the truth with varying sampling interval

<table>
<thead>
<tr>
<th>Sampling Interval</th>
<th>Deviation in % from the Pseudo Ground Truth (GPS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sec</td>
<td>5%</td>
</tr>
<tr>
<td>20 sec</td>
<td>9%</td>
</tr>
<tr>
<td>1 min</td>
<td>11%</td>
</tr>
<tr>
<td>3 min</td>
<td>14%</td>
</tr>
</tbody>
</table>

The results indicate the SCAAT-estimated travel times are reasonably close to what a road user
actually experiences. (GPS probe vehicles can temporarily be assumed to be any random road users’ vehicles that use the SCAAT-based route guidance applications per se.)

5.3 Second Phase: Data Fusion in Micro-simulation

A traffic simulation software such as Paramics is capable of measuring attributes of all vehicles at every time step on a particular road section. By averaging the speed values of all vehicles on the section of the road, it is possible to compute the “true” traffic conditions in the simulation, and results of the developed SCAAT filters can be compared to the true conditions for their performance evaluation under different sampling strategies. Unlike the first phase of this chapter where GPS data is treated as the pseudo ground truth, in this section, it is finally possible to access the actual ground truth speed values at least within the simulation environment. Since the qualitative information sources (radio broadcasts and traffic department’s website) in the first phase of this study cannot be simulated, and do not affect the SCAAT KF output significantly due to relatively high variance and low sampling rate, only GPS and loop detectors are used in this second phase of the study. It is noted that the main contribution of the first phase of the study is to illustrate how even qualitative information can be fed into numerical data fusion.

In this thesis, the ground truth is defined as an average speed of all vehicles on the road link. Figure 5.6 shows multiple speed values from multiple vehicles at each time step on one section of the road. The average speed of all the vehicles are considered to be the ground truth. The line in Figure 5.6 is the average speed of the link at each simulation time step.
5.3.1 Network Modeling in Paramics

EMME/2 is used at the planning level to perform the user equilibrium (UE) traffic assignment based on the Transportation Tomorrow Survey (DMG, 2006). Then, the network known as the “Queen’s Quay (QQ) Network” in Figure 5.7 is cut from the larger GTA network and an origin-destination (OD) matrix is calibrated at the planning level to reproduce the observed traffic counts on the boundaries of the network on EMME/2. Then this OD is fed to Paramics to perform the dynamic traffic assignment. The network consists of 546 nodes or intersections of which 58 are signalized, 1047 links and 53 zones. Aerial photos from Google Maps (Google Inc.) are used to determine the number of lanes, exact locations of divergence and convergence points, exclusive left and right lanes, curb points, location of on and off-ramps to the Gardiner Express Way in the GTA.

The traffic counts that are used to calibrate the Paramics model of the internal QQ network are provided to IntelliCAN by the City of Toronto. One typical weekday’s example traffic counts near the intersection of Front Street and Church Street are included in Appendix B.

The QQ network includes the Gardiner Expressway from Bathurst Street in the West to the Don...
Valley Parkway in the East, and a “buffer area” extending from the Toronto Waterfront in the South to Front/Wellington Street in the North.
There are two major groups of parameters used in Paramics. The network-wide parameters shown in Table 5.3 specify characteristics of the entire network behaviours. The vehicle characteristics parameters shown in Table 5.4 specify the physical behavior of individual vehicles.
Table 5.3 Network-wide parameters used in Paramics

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Used Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback Interval</td>
<td>Drivers are informed at set intervals so they can reroute to less congested routes if available.</td>
<td>2 minutes</td>
</tr>
<tr>
<td>Mean Headway</td>
<td>Each vehicle has a target headway with a mean value depending on factors such as weather, highway type, vehicle type, driver's aggressiveness and awareness.</td>
<td>0.74 second</td>
</tr>
<tr>
<td>Mean Reaction Time</td>
<td>Drivers need time to react to perceived situations. For example, when a driver sees a car in front of him/her slowing down, the driver spends some time until the breaking actually takes place</td>
<td>0.61 second</td>
</tr>
<tr>
<td>Time Steps per Second</td>
<td>Specifies the number of discrete simulation intervals that are simulated per second</td>
<td>2 steps per second</td>
</tr>
<tr>
<td>Speed Memory</td>
<td>Each vehicle has the facility to remember its own speed for a number of time steps. This mechanism is used implement driver reaction time by basing the change in speed of the following vehicle on the speed of the leading vehicle. This variable adds the flexibility to model larger reaction times or smaller time steps</td>
<td>3</td>
</tr>
<tr>
<td>Curve Speed Factor</td>
<td>Specifies how much the vehicles slow down due to the curvature on roads</td>
<td>1</td>
</tr>
<tr>
<td>Headway Factor</td>
<td>Target headway can be modulated using this factor. For example, in a tunnel, all vehicles tend to extend their headways by certain percentage</td>
<td>1</td>
</tr>
<tr>
<td>Link Speed</td>
<td>The desired speed when the traffic volume is low</td>
<td>65 mph</td>
</tr>
</tbody>
</table>
Table 5.4  Vehicle characteristics parameters

<table>
<thead>
<tr>
<th>Vehicle Type</th>
<th>Top Speed (km/hour)</th>
<th>Maximum Acceleration (m/s²)</th>
<th>Max Deceleration (m/s²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>158.4</td>
<td>2.5</td>
<td>4.5</td>
</tr>
<tr>
<td>LGV</td>
<td>126</td>
<td>1.8</td>
<td>3.9</td>
</tr>
<tr>
<td>OGV1</td>
<td>104.4</td>
<td>1.1</td>
<td>3.2</td>
</tr>
<tr>
<td>OGV2</td>
<td>118.8</td>
<td>1.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Coach</td>
<td>126</td>
<td>1.2</td>
<td>3.7</td>
</tr>
</tbody>
</table>

5.3.3  Data Collection from Micro-simulation

A busy part of Downtown Toronto, namely the financial district of Toronto shown in Figure 5.8, is used as the study area. It is the most densely built-up area in the GTA and about 100,000 commuters enter and leave the district every working day. The road network is modelled for the period from 7:30 AM to 9:00 AM on weekdays. During the simulation, the first 30 minutes of the simulation is regarded as a “warm-up” period, and no data are collected during that period. On the road link of Wellington St. W., 4 loop detectors are placed on each of the 4 lanes between Bay St. and Yonge St. The simulated loop detectors provide speed information whenever new vehicles pass over them. The road link is a west bound one-way street. From 8:00AM to 9:00AM, 1328 vehicles have appeared on the link and passed 1 of the 4 loop detectors.
Figure 5.8 Data collection area for the data fusion using Paramics
5.3.4 Evaluation of Estimation Error

Using the true current traffic conditions in the simulation, a set of conventional estimation error measures can be computed for different sampling strategies. Table 5.5 lists those measures.
### Table 5.5 Conventional performance measures

(Qiu and Ran, 2008)

<table>
<thead>
<tr>
<th>Performance Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Error (ME)</strong></td>
</tr>
<tr>
<td>[ \frac{1}{N_T} \sum_{k=1}^{N_T} (\hat{v}_k - v_k) ]</td>
</tr>
<tr>
<td><strong>Mean Absolute Error (MAE)</strong></td>
</tr>
<tr>
<td>[ \frac{1}{N_T} \sum_{k=1}^{N_T}</td>
</tr>
<tr>
<td><strong>Mean Relative Error (MRE)</strong></td>
</tr>
<tr>
<td>[ \frac{1}{N_T} \sum_{k=1}^{N_T} \left( \frac{\hat{v}_k - v_k}{v_k} \right) ]</td>
</tr>
<tr>
<td><strong>Mean Absolute Relative Error (MARE)</strong></td>
</tr>
<tr>
<td>[ \frac{1}{N_T} \sum_{k=1}^{N_T} \left</td>
</tr>
<tr>
<td><strong>Mean Square Error (MSE)</strong></td>
</tr>
<tr>
<td>[ \frac{1}{N_T} \sum_{k=1}^{N_T} (\hat{v}_k - v_k)^2 ]</td>
</tr>
<tr>
<td><strong>Root of Mean Square Error (RMSE)</strong></td>
</tr>
<tr>
<td>[ \sqrt{\frac{1}{N_T} \sum_{k=1}^{N_T} (\hat{v}_k - v_k)^2} ]</td>
</tr>
</tbody>
</table>

Where; \( N_T \) = number of time steps or measurements,

\( \hat{v}_k \) = speed estimation from the SCAAT filters,

\( v_k \) = true current average speed.
5.3.5 Second Phase Results of Data Fusion

5.3.5.1 Effect of Varying Number of Probe Vehicles

In this section, the error measures are used to compare different sampling scenarios. The sampling interval used on all sensors is 1 second. It is noted that the 1 second interval is arbitrarily set as a fixed quantity to see the effect of other factors, such as the number of probes and the number of loop detectors being monitored.

**Scenarios**

A. 1 Loop Det. : Using 1 loop detector on lane 1
B. 4 Loop Det. Avg. : Using the average of 4 loop detectors on the 4 lanes
C. 1GPS SCAAT: SCAAT filter applied to 1 GPS probe vehicle at a time
D. 2GPS SCAAT: SCAAT filter applied to 2 GPS probe vehicle at a time
E. 4GPS SCAAT: SCAAT filter applied to 4 GPS probe vehicle at a time
F. 6GPS SCAAT: SCAAT filter applied to 6 GPS probe vehicle at a time

Scenario A monitors a single loop detector on lane 1. Even though the sampling interval is 1 second, the data from the loop detector would not be updated until a new vehicle passes over it. Therefore, the relatively high sampling rate would not improve the performance because it sometimes takes minutes of waiting time to see a new vehicle passes over the loop detector and until then the old measurement will not be updated. Scenario B monitors the average of 4 loop detectors on all 4 lanes. Any vehicle that passes over any 1 of the 4 lanes will update the current average speed. However, it is important to note that both in Scenario A and B, the loop detectors are monitoring a fixed point on the link. The data readings at the fixed point do not necessarily represent the current traffic conditions on the link. Scenario C tracks a random probe vehicle on the link. Scenario D, E and F track 2, 4, 6 random probe vehicles respectively, and the average speed of the multiple probes are monitored by the SCAAT filter.
Table 5.6 shows the results. It is found that Scenario F results in the lowest error measures. It is interesting to note that Scenario C with just a single probe performs better than Scenario A and B with loop detector(s). Even though it is not an economically feasible option, conceptually, if a single GPS unit is tracked every 1 second, it is as if there are so many loop detectors right on the single probe's path that update information every 1 second. In other words, there is no idling time for the GPS sensor, while loop detectors only update their predictions when new vehicles pass them. Therefore, it is not surprising that Scenario C performs better than Scenarios A and B. However, having access to the GPS unit every 1 second may not be realistic (even though technically not impossible) due to high communication costs. This section verifies that even 1 GPS unit can perform better than loop detector(s) as long as the sampling frequency is reasonably high.

### Table 5.6 Effect of varying number of probe vehicles

<table>
<thead>
<tr>
<th></th>
<th>A.1 Loop Det.</th>
<th>B.4 Loop Det. Avg.</th>
<th>C.1GPS SCAAT</th>
<th>D.2GPS SCAAT</th>
<th>E.4GPS SCAAT</th>
<th>F.6GPS SCAAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>18.04</td>
<td>17.54</td>
<td>-3.55</td>
<td>0.55</td>
<td>0.32</td>
<td>0.08</td>
</tr>
<tr>
<td>MAE</td>
<td>20.71</td>
<td>18.59</td>
<td>10.55</td>
<td>10.52</td>
<td>6.99</td>
<td>3.60</td>
</tr>
<tr>
<td>MRE</td>
<td>11.59</td>
<td>13.29</td>
<td>2.83</td>
<td>3.17</td>
<td>0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td>MARE</td>
<td>11.66</td>
<td>13.31</td>
<td>3.23</td>
<td>3.44</td>
<td>0.32</td>
<td>0.15</td>
</tr>
<tr>
<td>MSE</td>
<td>626.75</td>
<td>506.83</td>
<td>172.07</td>
<td>153.09</td>
<td>94.14</td>
<td>39.17</td>
</tr>
<tr>
<td>RMSE</td>
<td>25.03</td>
<td>22.51</td>
<td>13.12</td>
<td>12.37</td>
<td>9.70</td>
<td>6.26</td>
</tr>
</tbody>
</table>
5.3.5.2 Effect of Varying Sampling Strategies

This section focuses on the effect of different sampling strategies. The following describes the 3 scenarios tested.

**Scenarios**

A. 1GPS SCAAT every 1sec: A single probe vehicle is tracked at each time step for every 1 second

B. Loop + GPS every 1 sec: The SCAAT filter switches input source between a single loop detector and 1 probe vehicle every 1 second

C. Loop + GPS every 30 sec: The SCAAT filter switches input source between a single loop detector and 1 probe vehicle every 30 seconds (1 loop data and 1 GPS data every 1 minute)

Scenario A can be thought of as the most ideal case in this section where a single probe vehicle is tracked every 1 second. Scenario B switches between the loop detector and the probe vehicle every 1 second. It is important to note that the loop detector however does not update its estimation until a new vehicle passes over it. Even if the loop detector is sampled every 2 seconds in Scenario B, the data would still be intact if no new vehicles have passed during that interval. Table 5.7 shows that Scenario B resulted in poorer accuracies than Scenario A. Scenario C increases the sampling interval by 30 times from Scenario B. However it is interesting that Scenario C performs not much worse than Scenario B. MRE and MARE are actually even lower for Scenario C. The similar performances of B and C show that the most important factor is the percentage mix of the loop detector and the GPS data. The SCAAT filter should only use the most accurate sensor at any simultaneous time step. However, this statement is not to be confused with the main philosophy behind the multiple-sourced data fusion. If multiple information sources are available in the same time step, then the most accurate sensor
should be chosen. But, in real life, this is not the case most of the time. Different sensors update their information at different frequencies and usually no two sensors are active in the same time step. This section is intentionally designed with the extreme scenarios to compare the effect of different sampling strategies that in real-life do not happen often. The conclusion from this section is that the data should be combined only if there are data gaps from the most accurate sensor. Fusing the lower quality loop detectors even though there are other better sensors simultaneously available would only increase the error measures. For example, if there are data gaps from GPS data and the loop detector data is available in the gaps, only if the duration of GPS data gaps is reasonably large, the loop detector should be fused in. Whether to include the loop detector data or not is a subjective call that results in trade-off between estimation accuracy and service reliability that is induced from multiple sources of information. In this section, the sampling rates were intentionally pushed to the limit (1 Hz) so that the comparisons between the Scenario A, B and C can easily be compared. In reality, GPS data from the probes will not be available every 1 second and the loop detector’s speed value will only update when a vehicle passed over it. Even if the loop detector is sampled every 1 second, unless a new vehicle passed over it, it will only provide the speed of the previous vehicle and this will deteriorate the estimations. Therefore, the use of SCAAT filters are still useful in real life where there will be data gaps temporally even with both GPS probes and loop detectors in operations.
Table 5.7 Effect of varying sampling strategy

<table>
<thead>
<tr>
<th></th>
<th>A. 1GPS SCAAT every 1 sec</th>
<th>B. Loop+GPS every 1 sec</th>
<th>C. Loop + GPS every 30 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>-3.55</td>
<td>6.38</td>
<td>7.24</td>
</tr>
<tr>
<td>MAE</td>
<td>10.55</td>
<td>15.09</td>
<td>19.11</td>
</tr>
<tr>
<td>MRE</td>
<td>2.83</td>
<td>6.45</td>
<td>1.21</td>
</tr>
<tr>
<td>MARE</td>
<td>3.23</td>
<td>6.89</td>
<td>1.63</td>
</tr>
<tr>
<td>MSE</td>
<td>172.07</td>
<td>351.38</td>
<td>554.31</td>
</tr>
<tr>
<td>RMSE</td>
<td>13.12</td>
<td>18.75</td>
<td>23.54</td>
</tr>
</tbody>
</table>

5.3.5.3 Varying Sampling Interval vs. Varying Number of Probes

This section compares 2 roughly equivalent scenarios.

**Scenarios**

A. 1GPS SCAAT every 10 sec: The SCAAT filter is applied to a single GPS probe that is sampled every 10 seconds.

B. 3GPS SCAAT every 30 sec: The SCAAT filter is applied to the average of 3 GPS probes that is sampled every 30 seconds.

Scenarios A and B are similar in that they both end up sampling 3 GPS data every 30 seconds. Scenario A samples 1 probe vehicle 3 times more often than Scenario B, and Scenario B samples
3 times less frequently but samples 3 GPS probes. The result shown in Table 5.8 is interesting in that the 2 scenarios performed about the same. Neither A nor B is a clear winner. From the result, it can be concluded that the number of probe vehicles monitored are not as important as the total number of pings per unit time regardless of the number of different probes. However, this is based on a major assumption that those monitored vehicles behave similarly. In the case where there are large distributions of bahaviour of probes, then, it is recommended to monitor more number of probes less often than monitoring one probe more often. This would minimize the effect of the behavioural component of one driver on the speed estimation accuracies.

**Table 5.8 Varying sampling interval vs. number of probes**

<table>
<thead>
<tr>
<th></th>
<th>A. 1GPS SCAAT every 10 sec</th>
<th>B. 3GPS SCAAT every 30 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ME</strong></td>
<td>-4.72</td>
<td>1.38</td>
</tr>
<tr>
<td><strong>MAE</strong></td>
<td>12.87</td>
<td>14.13</td>
</tr>
<tr>
<td><strong>MRE</strong></td>
<td>-0.04</td>
<td>4.25</td>
</tr>
<tr>
<td><strong>MARE</strong></td>
<td>0.81</td>
<td>4.56</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>227.42</td>
<td>273.66</td>
</tr>
<tr>
<td><strong>RMSE</strong></td>
<td>15.08</td>
<td>16.54</td>
</tr>
</tbody>
</table>

5.3.5.4 Effect of Varying Sampling Interval

In this section, 3 scenarios are used to test the effect of varying the sampling interval.
**Scenarios**

A. 1GPS SCAAT every 10 sec: A SCAAT filter is applied to 1 random probe vehicle that is sampled every 10 seconds.

B. 1GPS SCAAT every 30 sec: A SCAAT filter is applied to 1 random probe vehicle that is sampled every 30 seconds.

C. 1GPS SCAAT every 60 sec: A SCAAT filter is applied to 1 random probe vehicle that is sampled every 60 seconds.

Table 5.9 shows the results as expected. Scenario A performs the best simply because it is sampling the probe vehicle at the highest frequency among the 3 scenarios while other factors are unchanged.

**Table 5.9 Effect of varying sampling interval**

<table>
<thead>
<tr>
<th></th>
<th>A. 1GPS SCAAT every 10 sec</th>
<th>B. 1GPS SCAAT every 30 sec</th>
<th>C. 1GPS SCAAT every 60 sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>-4.72</td>
<td>-5.09</td>
<td>-0.16</td>
</tr>
<tr>
<td>MAE</td>
<td>12.87</td>
<td>14.05</td>
<td>17.58</td>
</tr>
<tr>
<td>MRE</td>
<td>-0.04</td>
<td>-0.24</td>
<td>0.62</td>
</tr>
<tr>
<td>MARE</td>
<td>0.81</td>
<td>0.62</td>
<td>1.09</td>
</tr>
<tr>
<td>MSE</td>
<td>227.42</td>
<td>332.04</td>
<td>503.73</td>
</tr>
<tr>
<td>RMSE</td>
<td>15.08</td>
<td>18.22</td>
<td>22.44</td>
</tr>
</tbody>
</table>
5.4 Conclusion

This chapter presents a data fusion method that can deal with both quantitative hardware-based sensors (GPS probes and loop detectors) and qualitative expert-based sensors (radio broadcasts and traffic department’s website). A special type of KFs known as a SCAAT filter is used to fuse multiple data sources and estimate current traffic conditions. In the first phase of this chapter, the developed SCAAT filters are found to be capable of handling multiple information sources (sensors) that are GPS probes, loop detectors, radio broadcasts and traffic department’s website. In the second phase, in order to have access to the true traffic conditions, a micro-simulation package is used and the developed SCAAT filters are applied with different sampling strategies with GPS probes and loop detectors. Various error measures are computed and compared for the different sampling scenarios.

In conclusion, it is found that using SCAAT KFs for traffic monitoring is an effective way of dealing with various types of traffic sensors and can provide more reliable traffic conditions utilizing every available existing traffic infrastructure.
Chapter 6
Multi-criteria Route Guidance

6 Multi-criteria Route Guidance

Chapter 4 deals with detecting the mode of transportation from GPS data and filtering out the speed information from the GPS data for the “auto” mode. Once the speed information is collected, they are fused with other traffic sensors for more reliable travel time estimations in Chapter 5. The most natural use of such information is finding the fastest route which is an advanced version of finding the shortest (in distance) route. This thesis first finds the shortest and fastest routes and then finds routes based on other criteria by using advanced GIS. With GPS navigators, drivers are now more confident in exploring ‘out of the ordinary’ routes. Most route guidance applications minimize travel distance and time which are important factors, but are not the only navigational criteria of interest to users especially in urban environments. With the aid of advanced features of GIS, new geospatial factors such as the three-dimensional (3-D) nature of the roads and crime rates can be included in the route guidance for broader applications. For instance, 3-D GIS can generate information on visible sceneries along a given route (for travellers) or the slopes of the consecutive road segments (for bike riders). In addition, pedestrians and bike riders who are more directly exposed to external environment may opt to avoid regions with high crime rates. The scope of this chapter is limited to incorporating sceneries, road slopes and homicide rates into route guidance. In the future, this concept of incorporating new geospatial information can be extended for computing for example, low-elevation areas which are susceptible to flooding and hilly regions with heavy traffic experiencing pollutions. As an initial attempt of endless number of applications using new types of geospatial data, this thesis presents methods of incorporating 3-D features of the roads and
geospatial crime rates information for route guidance purposes. It is found that the three
dimensional nature of the roads and crime rate related information can result in different route
choices.

6.1 Introduction

Due to the emergence of cheaper, smaller and more sensitive GPS devices, the demand for
consumer-level GPS navigators has been increasing. The integrated route guidance algorithms
that accompany such devices, primarily, use travel distances and times as their main input. Some
of the more advanced devices incorporate real-time traffic conditions and points of interest in
computing the optimal path. In addition, some commercial GPS navigators are now available
specifically for pedestrians in a form of GPS embedded cell phones, and for bike riders in a form
of GPS watch or bike-mount GPS navigators with bike road maps saved in them. However,
most existing route guidance algorithms treat the road networks as two-dimensional (2-D)
entities. This approach misses out potential benefits that can be harvested by incorporating
three-dimensional (3-D) characteristics of the roads into the algorithms. There have been some
initial attempts to include 3-D features of the road into the car navigation. SONY XYZ™ (Tech
Japan Inc., 2009) navigation product lines sold in Japan can reproduce 3-D images of buildings
near roads aiding the driver with visual confirmation between the actual environments and the
image shown on the navigation device. Some researchers already have started working on
including live video into the car navigation. However, none of them yet, fully incorporate
benefits that 3-D GIS can offer.

GIS can aid the driver in choosing a scenic route between a particular origin-destination pair. A
scenic route is roughly defined as a route with visually pleasing sights. However, if hills or other
man-made structures block the view, even a route in the middle of scenic regions can be
considered as a non-scenic route. Therefore, GIS computes a visibility factor which is useful in
determining scenic ratings of different routes. Visibility heavily depends on the height of the
roads, which is information available via digital elevation models (DEM). With aids of 3-D
tools in GIS, the DEM can be processed to produce valuable information on roads such as
visibility and slopes.
It is noted that computing scenic route can especially benefit tourist drivers who are unfamiliar with the city and have predetermined entrance and exit points (of major highways) of the city that they are bypassing on their trip route. Instead of simply bypassing the city for their final destinations using major highways, drivers can find an alternative more scenic path by sacrificing some extra travel time.

 Visibility information can assist the driver in finding the best scenic route. In order to have accurate visibility information, one needs to know both the DEM and height information of man-made buildings. Then the layers of DEM and the building information sum up to provide effective elevation information for computing the visibility. At the time of writing, the height information of the buildings is not available and only the DEM is available. However, in the future, as such data are collected; the visibility information will be more accurate. Google Earth™ has already started including 3-D models of famous buildings around the world in their “3-D Buildings” layer with height information built into the models. Figure 6.1 shows 3-D building models of Tokyo, Japan and Toronto, Canada. In Germany, CityGML (CityGML, 2009) develops a common information model for 3-D urban objects. In addition, an international industry consortium known as the Open Geospatial Consortium (OGC) (Open Geospatial Consortium Inc., 2009) also deals with specifying 3-D building standards within the sector of 3D Information Management (3DIM) Working Group. Canadian cities will follow the 3-D modeling trend shown in Figure 6.1. Having access to 3-D models of the Toronto skyline is no longer a distant possibility. When such practices lead to a complete 3-D GIS map of Toronto, the visibility information will mature fully as well. This chapter develops a methodology of using such 3-D GIS maps assuming that the same procedure will work with future mature versions of the maps when they become available.
The slope information can provide bike riders and pedestrians with the least inclined path and can also be fed into pollution-generating models to estimate pollution levels on the roads for transportation planning purposes. Such 3-D applications can also assist city planners, evacuation planners and other public officials in many aspects of their operations. Figure 6.2 shows comparative diagrams of 2-D and 3-D road maps of Toronto.

In addition, current GIS maps can be upgraded by adding geospatial information such as crime rates near the roads. As a foreigner visiting a new city, the traveler may opt to avoid dangerous regions especially when walking or riding a bike. Portable GPS navigators with such geospatial information can also benefit travelers using non-auto modes of transportation.

The main objective of this chapter is to exploit the 3-D road and crime rate information of the urban environment to aid new ITS route guidance applications and to suggest a guideline for
future ITS applications using new geospatial data. Several case studies applied in the GTA are presented in this chapter.

![2D GIS vs 3D GIS](image)

Figure 6.2 Comparison of 2-D and 3-D road maps in Toronto

6.2 Multi-criteria Data Collection

6.2.1 Scenic View and Attraction Zone Survey

From May 19\textsuperscript{th} to 23\textsuperscript{rd}, 2008, a survey is conducted for the evaluation of scenic views in different zones in Toronto. An area is assumed to be scenic if it is attractive to travelers. Scenic values are subjective and vary among the population. In order to capture as objective as possible trend in scenic values across the GTA, the survey was conducted on a sample of professionals in the tourism industry in the GTA. 29 professional travel agents from 12 different agencies participated in the survey. 20 participants were interviewed in person and 9 participants
responded via an on-line survey. The GTA is sub-divided into 13 zones by major roads. Figure 6.3 shows the 13 zones. The downtown area is divided into smaller zones because relatively more attractions are located in the downtown area. (E.g. C.N. Tower, Rogers Centre etc.) The sizes and locations of the zones are determined arbitrarily in similar sizes in order to illustrate the procedure of how one may incorporate scenic ratings into a GIS layer. Developing an objective method of determining specifics of each zone is outside of the scope of this chapter. Figure 6.4 shows the survey form used in the scenic route survey. The zones in Figure 6.3 can be divided into more number of zones with smaller sizes. However, the survey participants can have difficult time rating many smaller zones differently. In addition, when people rate a scenery, it is common to associate it with relatively larger zones. (“South end of Toronto looks great” or “North-west corner of Toronto does not have an excellent scenery”)

Figure 6.3 Zonal division of the GTA
Survey for Studies of Scenic Route Determination

1. How long have you been in Toronto?
   - Less than 1 year
   - 1 year to less than 5 years
   - 5 years to less than 15 years
   - 15 years to 30 years
   - More than 30 years
   - Other __________

2. How often do you travel around Toronto per month?
   - Everyday
   - 2~3 times a week
   - 2~3 times a month
   - Once a week
   - Once a month
   - Other __________

3. If you rate the City of Toronto as a sightseeing area between 0 and 5, what is the rate that you would give?
   - 0
   - 1
   - 2
   - 3
   - 4
   - 5

4. Do you commute from a different city to work? If yes, please indicate the city or the region in the space provided.
   - Yes ____________________________
   - No

5. If you rate the following areas in the City of Toronto as sightseeing areas, select the corresponding rate that you would give to each area.

   | 0 | 1 | 2 | 3 | 4 | 5 | I don’t know |
---|---|---|---|---|---|---|-------------|
1. |   |   |   |   |   |   |             |
2. |   |   |   |   |   |   |             |
3. |   |   |   |   |   |   |             |
4. |   |   |   |   |   |   |             |
5. |   |   |   |   |   |   |             |
6. |   |   |   |   |   |   |             |
7. |   |   |   |   |   |   |             |
8. |   |   |   |   |   |   |             |
9. |   |   |   |   |   |   |             |
10. |   |   |   |   |   |   |             |
11. |   |   |   |   |   |   |             |
12. |   |   |   |   |   |   |             |
13. |   |   |   |   |   |   |             |

Figure 6.4 Scenic route survey form used
The scenic values of each zone are averaged based on the Question 5 of the survey form where a higher number indicate better scenery. The survey results are listed in the following. Region 12 which contains the C.N. Tower and the Rogers Centre, the most famous attractions in the GTA, is given the highest scenic value of 4.7 out of 5.

Zone 1: 2.33
Zone 2: 1.85
Zone 3: 3.12
Zone 4: 1.73
Zone 5: 2.7
Zone 6: 3.04
Zone 7: 2.31
Zone 8: 2.37
Zone 9: 2.63
Zone 10: 2.97
Zone 11: 3.97
Zone 12: 4.7
Zone 13: 4.45

6.2.2 Digital Elevation Model

A digital elevation model (DEM) contains information about the elevation of a ground surface. It is represented in a form of a raster file either in a grid of squares or triangles. In this study, the
Canadian Digital Elevation Model (CDEM) of Toronto from the Natural Resources Canada (University of Toronto Map Library, 2008) is used in this study. The data is in a raster format in squares with a resolution of 10 meters by 10 meters. This means that the theoretical minimum size of the zones in Figure 6.3 is 10 meters by 10 meters. Figure 6.5 shows the general trend in elevation values across the GTA. The brighter colour indicates the higher elevation.

Figure 6.5 Canadian Digital Elevation Model (CDEM) of Toronto.

(University of Toronto Map Library, 2008)
6.2.3 Bike Paths

The Toronto Cycling Map from the City of Toronto (City of Toronto, 2008) is considered for carefully determining the test case routes that are shown in Section 6.4.1. The test case routes of C and D in Figure 6.12 contain some of the available bicycle routes.

6.3 Construction of Various GIS Layers

6.3.1 Scenic View Layer (SVL)

The construction of the scenic view layer (SVL) requires two different GIS layers: the visibility layer (VL) and the zonal scenic worth raster layer (ZSWRL). From a point on a surface, only certain parts of the land are visible because of the differences in elevation and the absence of obstacles in the line of sight. The visible areas are not always in vicinity of the road the user is on. At the same time, the area near the current location can occasionally be invisible due to abrupt changes in elevation.

A spatial analysis tool called ‘viewshed’ can identify the cells in an input raster that can be seen from one or more observation points. The ‘viewshed’ tool has 9 characteristics that can be controlled depending on the needs of applications. Figure 6.6 shows the visual illustration of the 9 variables and the descriptions of the 9 variables are listed in the following. (ESRI Inc., 2008)

- Spot (SPOT): The surface elevations for the observation point(s).
- OffsetA (OF1): The vertical distance to be added to the z-value of the observation point(s).
- OffsetB (OF2): The vertical distance to be added to the z-value of each cell as it is considered for
  the visibility and responsible for blocking the line of sight.
- Azimuth1 (AZ1): The start of the horizontal angle to limit the scan.
- Azimuth2 (AZ2): The end of the horizontal angle to limit the scan.
- Vert1 (V1): The top of the vertical angle to limit the scan.
- Vert2 (V2): The bottom of the vertical angle to limit the scan.
- Radius1 (R1): The inner radius that limits the search distance when identifying areas visible from each observation point.
- Radius2 (R2): The outer radius that limits the search distance when identifying areas visible from each observation point.

Figure 6.6 The 9 controllable characteristics of the viewshed tool

(ESRI Inc., 2008)

In this study, the SPOT is directly set to the elevation value from CDEM. OF1 and OF2 are set to zero assuming there are no extra elevations on the land. If the CDEM is improved in the future and contains the information about the heights of man-made structures, OF1 and OF2
values are, in a sense, embedded into DEM that the two variables are not needed. AZ1 and AZ2 are set to 0 and 360 indicating that the ‘viewshed’ function will look out the entire range of azimuth. V1 and V2 are set to 0 and 90 to cover the entire surface below the current location. R1 and R2 are set to 0 and an arbitrarily large value respectively so that the algorithm will look out on the entire DEM raster map. However, depending on person, R2, the maximum sight distance can vary. In future research in the field of road safety, the distribution of R2 among drivers can play a major role. However, this thesis assumes drivers can see very far.

The tool is applied to two major highways in Toronto for illustration purposes. Figure 6.7 shows visible areas from the Highway 401 and Gardiner Expressway. The visible areas are coloured in white. 3-D building models in Figure 6.1 is only available in Google Earth (for a limited number of famous buildings in the downtown of the GTA) for graphical displaying purposes and exporting the height data is not currently available. In this thesis, visible zones are computed solely based on the DEM only without the building height information due to the following reasons.

1. Building height information is only available for a small number of buildings in downtown.
2. Google Earth currently uses the 3-D building models for displaying purposes only and does not have such data available in DEM format.
3. The scenery computing procedure will remain unchanged even if the complete city-wide building height information will be available sometime in the future and the focus of this chapter is to illustrate this procedure.
Figure 6.7 The visibility from Highway 401 based on the 10 observation points (top) and the visibility from the Gardiner Expressway based on 10 observation points (bottom).

The VL (visibility layer) is constructed for each link of the road network. The VL is a Boolean raster file that has a value of 1 when the cell is visible from the observation point and 0 otherwise. From the conducted scenic view survey and relative attractiveness values for each zone, ZSWRL is constructed. By applying a mathematical raster operation (multiplication) on the intersecting parts of the two layers (the VL and ZSWR layer), visible and scenic regions are identified. By replacing travel times with the reciprocals of scenic worth values, one can maximize the scenic values associated with the chosen route instead of minimizing the travel times. However, depending on the observation point on the link, the scenic values can vary even
on the same link. Therefore, three observation points are selected at the beginning, the end and in the middle of each link. The scenic value of the link is, then, calculated by averaging the scenic values of the three observation points. Figure 6.8 illustrates how the scenic value of a particular road section can be computed.

\[ \text{Scenic Value of a Point (SVP)} = \sum_{\text{Entire Map}} \text{Boolean Visibility Layer} \cap \text{ZSWRL} \]

(Equation 6.1)

\[ \text{Scenic Value of a Road Link} = \frac{\text{SVP at the starting point} + \text{SVP at the mid point} + \text{SVP at the ending point}}{3} \]

(Equation 6.2)
6.3.2 Road Slope Layer (RSL)

Unlike the scenic value layer, the road slope layer does not need subjective opinions of people and, therefore, does not need the extra information beyond the factual elevation information. The slope layer can be generated using a 3-D spatial analysis tool called “slope”. The slope tool identifies the rate of maximum change in z-value (elevation) from each cell. (ESRI Inc., 2008) Figure 6.9 shows the slope variations near the intersection of Eglinton Avenue and Don Valley Parkway in the GTA. The routes that overlap with the yellow lines in Figure 6.9 have more hills.

![Slope variations in the GTA](image)

**Figure 6.9** Slope variations in the GTA

Similar to the computation of the scenic view layer, the slope layer is computed based on the three observation points. At each observation point, the absolute value of the maximum slope, in a circular buffer of 20 meter radius, is chosen. In this application, the radius is two times the cell size of DEM. The maximum slope values of the three observation points along the link at the
beginning, at the end and in the middle are then averaged. The averaged slope value is used as a representative slope of the link in this chapter.

6.3.3 Crime Rate Layer (CRL)

Information about crime rates of regions can further enhance the route guidance algorithms. Figure 6.10 shows different kinds of crimes that Toronto Police Service recognizes. (Toronto Police Service, 2009) Unfortunately, not all crimes are geospatially coded and available to the public except homicides. For the past a few years, the number of the homicides with respect to the total number of all crimes consistently ranges from 0.15% to 0.19%. This means that the number of homicides is consistently proportional to the total number of all crimes and can still be a good indicator of crime trends in general. In this chapter, as an illustrative procedure, it is assumed that the higher number of homicides occurring in certain regions represent increased danger in those regions. The number of homicides associated with each zone is counted and the region with the highest count receives the highest rating of 5 and the region with the lowest count get the rating of 1. The remaining zones are then categorized into the 5 point rating system and given a value from 1 to 5. Figure 6.10 shows the actual geospatially-coded homicides with a map (2005-2008), and historic counts of all crimes in the GTA from 2005 to October 2009.
Figure 6.10 Homicides in Toronto from 2005 to 2008

(Toronto Police Service, 2009)

The computation of the crime rate layer adopts methods similar to those used in computing the scenic view layer and the road slope layer. The crime rate ratings at the three observation points are averaged and the average value is used as a proxy of the safety risk associated with the link. Figure 6.11 illustrates the risk assessment process. The risk rating is determined by categorizing each zone into the 1-to-5 rating based on the number of homicides that occurred in each zone.
6.3.4 Determination of Minimum Cost Path

The Dijkstra’s shortest path algorithm (Cormen et al. 2001) is used to find the minimum cost path where the distance in the shortest path algorithm is replaced with costs from SVL, RSL and CRL. The algorithm finds the minimum cost paths from a single source vertex to the final destination vertex in a weighted, directed graph.

6.4 Illustrative Case Studies in Toronto

6.4.1 Across Toronto Scenario

The developed GIS layers are applied to some of the possible routes from the south-west end to
the north-east end of the GTA. The suggested origin-destination pair is the most common route for the American visitors to Canada who are travelling north while passing through Toronto. Depending on how zones are selected and how zonal scenic values are determined, the resulting route guidance can vary considerably. However, this proof-of-concept case study focuses on introducing the new concept of 3-D GIS based route guidance and only 6 routes are arbitrarily selected from the origin to the destination pair. Each route is sub-divided into 10 equally spaced links and each link contains the three observation points for the computation of the SVL, RSL and CRL. Figure 6.12 shows the 6 test routes labeled from A to F. The route B is the shortest route in distance. With the maximum legal speed applied, the route B is also the traditional fastest route. It is noted that not all routes have bike paths throughout the route. Therefore, when computing the most flat route for bikes, it is assumed that the bike riders will travel on alternative parallel roads near their chosen route whenever the bike path is not available.

Figure 6.12 Test case scenario routes for the evaluation of 3-D GIS route guidance
The most scenic route among them is found to be the route F based on the zonal scenic layer alone. The zonal scenic layer is built based on the survey that the participants gave higher values to the zones near the lakeshore area. The result is intuitive because the route passes right through the major attractions of the GTA along the lakeshore where many survey participants gave high ratings. With respect to the fastest route B, the route F is travelling on the opposite ends of the city. It is interesting that a different criterion can result in a very different route. However, in terms of the distance, they are about the same. This means, for tourist drivers, they can still reach the same exit point from the same entry point of the city while taking more scenic route by sacrificing a little more travel time in this case. The least hilly route is found to be the route A which directs vehicle around the hilly areas of the GTA. This results in a different and longer route from the fastest route B. In this case, even though the route A is computed to be the most flat route, people on bikes would not travel too much more distance. In order to properly address this issue, one needs to further investigate how they weigh different criteria. Then the combination of weights of multi-criteria can be properly modeled. However, the research scope of this thesis does not include the modeling of multi-criteria with varying weights in order to clearly show the different effects of each criteria. It is noteworthy that route B which runs mainly through Highway 401 is computed to be hilly on the RSL because there are many steep valleys in the route. But, in real-life, highway 401 is connected with bridges that levels out the route to make it more flat. One limitation of CDEM is that it only contains the natural elevation information and ignores the existing man-made structures. If the DEM files are improved in the future with information about man-made structures, more accurate results can be expected. The safest route is also found to be the route A because it runs around the corners of Toronto where crime rates are relatively lower. In order to fully benefit from the crime rate layer, one would need to construct a micro-level-crime-rate layer with finer zone divisions so that the traveler can still travel through the middle of the downtown instead of completely avoiding the area. In the case of multicriteria route guidance, weights can be assigned to each criterion; for instance,

\[
\text{Combined Weight} = a \times \text{Criterion A} + b \times \text{Criterion B} \ldots + e \times \text{Criterion E}
\]

where, \(a + b + \ldots + e = 1\)
For illustration, a, b, ..., e are chosen to be 0.2, and the five criteria are travel time, travel
distance, scenery along the route, road slope and crime rate. The route C is found to be optimal.
It is interesting to note that the route C is optimizing for none of each individual criterion and
those weights can be varied according to the traveller’s preferences.

6.4.2 Across Downtown Scenario for Safest Route Computation

A small region in Downtown Toronto is chosen and finer zonal divisions are made as in Figure
6.13. Unlike the zonal division made earlier at a larger scale, in this micro division, a method of
how to handle crimes that occur on zonal boundary lines needs to be addressed. In this study, the
cri mes that occur on the boundaries are included in both zones that are in contact. The number
of homicides from 2005 to 2008 in each zone is labeled in Figure 6.13. The 3 routes are chosen
under the following conditions:

1. They start and end at the same points.
2. They cover all zones together while being apart as far as they can from each other.
3. Their route lengths are similar and no back-tracing is allowed.
4. Their lengths and travel times are roughly equal. (Distance and travel time conditions are
equal and only GIS based factors can influence the route choice.)

Each route is divided into 10 links split apart by equal distances. Each link has three observation
points, as described in the earlier section. The route C is found to be the safest route because it
stays within the safer zones for a longer period of time than other routes. The danger level
associated with the Route A is found to be slightly higher than the Route C. Route B runs
through the zones with high crime rates and it is deemed to be the most dangerous route. Zone
sizes, zone divisions and resolutions are all expected to impact the results. However, a proper
sensitivity analysis of these factors is beyond the scope of this paper.
6.4.3 Further Potential Application of 3-D GIS Methods

The application of 3-D GIS techniques can benefit various fields of transportation engineering. For example, the visibility analysis can provide drivers with sight distance information at different locations and help evaluate the safety of the road. The visibility information can also be used in the analysis of drivers’ dynamic route choice models. For example, if a driver can see further, taking advantage of the higher elevation, the driver may be able to view and better anticipate approaching congestions and may use that information to alter their route choice at the next decision point. The slope layer can provide input for microscopic traffic pollution models as a hilly terrain causes higher fuel consumption rates and generates more pollutants. Finally, the elevation information can be used to find usable roads in a network subjected to floods. The
route guidance algorithm, in this case, would direct the traffic around the flooded region by using only the roads that are higher than the anticipated water level.

6.5 Conclusion

This chapter develops new methods for route guidance based on multiple criteria. The core methodology utilizes information from 3-D GIS databases and geo-coded crime rates GIS layers. Road segments are assigned with scores based on multiple criteria such as scenic values, road slopes and crime rates. With the aid of such 3-D GIS platform and information, rich-content route guidance is enabled. The application of the new concept is illustrated by several case studies in the GTA, where travelers are guided across the city according to their choice of criteria, based on scenery, road slope and safety. The presented applications are meant to be illustrative in nature. In-depth analysis of all pertinent factors in each case is beyond the scope at this conceptual stage.
Chapter 7
Conclusions and Recommendations

7 Conclusions and Recommendations

7.1 Summary

This thesis presents a system for traffic monitoring and route guidance named, GPS and GIS for Traffic Monitoring and Route Guidance (GISTMARG). GISTMARG is a real-time traffic monitoring system that can utilize mobile GPS devices as traffic probes and fuse the data with existing conventional traffic sensors. With 3-D GIS and geospatial data, GISTMARG also incorporates multiple criteria into route guidance methods for finding the most scenic, the most level and the safest routes.

The first component of GISTMARG is the transportation mode detection module. The study attempts to use relatively short traces of recent GPS data to identify and detect transportation modes in real time. This study initially analyzes the feasibility of using NNs as the mode classifier. The first phase of the study attempts to identify different modes of transportation with the proposed method and it was found to be an effective tool. The second phase examines the impact of varying the pinging frequency and monitoring duration length on the performance of mode identification processes. It is found that higher pinging frequency and longer monitoring duration length result in higher successful mode detection rates. As an example, if 68% auto mode detection rate is required for an application, it is recommended that the GPS probes are sampled 10 times for 10 minutes. Instead of recommending one particular parameter configuration, this thesis rather presents results from various settings and leaves the decision to the end-user because cost implications and desired accuracy level vary with different ITS
applications. The third phase finds that a more specialized classifier that is route-specific leads to further improvements on the mode identification performances. The last phase of the study compares the mode detection accuracy during peak and non-peak periods. It is found that the mode detection performs better during the peak periods because there are higher chances of speed and acceleration values fluctuating during the peak periods due to the frequent stopping that occurs during peak periods.

The second component of GISTMARG presents a method for fusing traffic information from different sources of data. It is found that by using a special type of KF known as SCAAT (Single-Constraint-At-A-Time) filter, it is possible to fuse the data from multiple sources and predict the near future traffic conditions. The SCAAT KF is found to be able to effectively update the current prediction based on the most recent sensor data. The study suggests a method of how qualitative traffic information such as radio broadcasts and on-line traffic information can be quantitatively incorporated into the data fusion processes. It is shown that by fusing the data from different data sources, robust, reliable and accurate real-time traffic information can be produced. When the data fusion methods were applied to the simulated data from Paramics, various sensor sampling scenarios were compared against the ground truth that only the simulated data can provide. The comparisons reveal that a single probe vehicle sampled at a high frequency can replace loop detector(s). When multiple probes are tracked at the same time, the error measures improves even further.

The third component of GISTMARG develops innovative methods of incorporating scenic view, road slope and crimes rates into the route guidance systems where advanced GIS features are fully utilized. By using the digital elevation model and the crime rate GIS layer of the GTA, the most scenic, the most level (the least hilly) and the safest routes are found. Such information is especially valuable to tourists, cyclists and pedestrians.

In this thesis, a new set of strategies is successfully implemented towards the new traffic monitoring and route guidance system named GISTMARG.
7.2 Recommendations

GISTMARG is applicable to regions considering monitoring traffic conditions using mobile AGPS devices in conjunction with conventionally existing traffic sensors such as loop detectors. In order to correctly detect the auto mode with 68% detection rate, it is recommended to query the mobile AGPS device 10 times for 10 minutes of duration. If a higher detection rate is required, 15 pings in 15 minutes can provide 80% detection rate. In order to implement the mode detection module in a new city, the following steps are suggested:

1. Collect 1~2 weeks of raw GPS data from various types of vehicles.
2. Apply GISTMARG mode detection and check if desired mode detection rates are achieved.
3. Use the developed NN for the entire city at the initial stage of the deployment.
4. For major sections of the roads, develop the route specific NN for more accurate detection rates.

The mobile AGPS device can have a small software application installed on it so that it carries previous 10 speed and acceleration values in its local memory buffer. If querying the phones through the wireless location services provided by the service provider is not economically feasible (too expensive), it is recommended that the user subscribe to an unlimited road-side internet service to avoid service fees incurred for each query by the service provider as in the case of using WLS. In addition, in order to attract cell phone owners to voluntarily participate in the traffic monitoring system as probes, traffic departments can subsidize the unlimited internet cost for them and use them as less expensive probes. The cell phone owners in return can enjoy free web browsing.

From the perspective of the data fusion module, in theory, GISTMARG does not ignore any existing sensors or information sources. GISTMARG based on the SCAAT KF data fusion technique can handle multiple sourced sensor data which can utilize all available sensors regardless of their accuracies. GISTMARG trusts sensors with higher accuracies but does not ignore sensors with lower accuracies either. By accessing all available sensors including the
lower accuracy information sources, traffic monitoring systems can especially improve their robustness by reducing the information gaps. When one develops an application that requires estimating current traffic condition, the data fusion module of GISTMARG can serve as an illustrative example of how both quantitative and qualitative information can be used for traffic monitoring purposes.

From the perspective of the route guidance module, it is recommended that future route guidance applications start to consider geospatial data layers in their algorithms to fully utilize available data. GISTMARG illustrates how such data layers can be incorporated into the route guidance and similar approaches can be included in other future applications with different types of data layers.

### 7.3 Future Research

In the topic of traffic monitoring, instead of only using private automobiles as probes and disregarding other modes of transportation, the traffic conditions experienced by buses, bikes, and pedestrians could be used to infer about the traffic conditions of the auto mode. There may be relationships between the traffic conditions of auto mode and other modes. If this issue can further be researched, none of the raw GPS data need to be thrown away but all of them can be used to estimate the current traffic conditions by certain mapping process from one mode to the other.

In the topic of route guidance, using the powerful 3-D GIS functions, the effect of visibilities on the route choices at route choice decision points can be researched further. Obviously, a driver clearly seeing the heavily congested road (due to the geographical advantage of the current location) in front of him/her will avoid that path. The route choice is at the end made by the driver, and not by the in-car navigation devices. This means that the route choice is heavily influenced by psychological factors. By understanding how people think better, we can attempt to develop a centralized closed-loop route guidance system that can better interact with tricky minds of people and result in near network-wide optimal strategies. The topic of analyzing the
behavioral factors of the general public is out of scope of this thesis which on its own is a major branch of transportation researches.

With the aid of 3-D GIS, it is possible to integrate live weather conditions as a GIS layer and the information can be applied to each road link. Rain or snow can impede the traffic conditions especially on hilly sections of the road. Live weather information combined with road slope information can help computing more realistic optimal routes.

For transportation planners, GISTMARG can provide valuable information for activity-based models. By analyzing the location and duration of mobile GPS devices, it is possible to predict the pattern of the activities. It is also possible to develop a time-of-day models by monitoring the network-connected mobile GPS devices. In addition, for the traditional 4-stage modeling, the traffic assignment step can benefit from the route guidance module of GISTMARG. Instead of considering only travel times for the comparison of different routes, the weights of the road links can be updated with GIS based factors and can influence the route decisions.
References


DMG. “Transportation Tomorrow Survey: Travel Survey Summaries for the Greater Toronto Area”, Joint Program in Transportation, University of Toronto, 2006.


Faghri, A., Hamad, K., Kerr, D., Magdefrau, C., Stokes, J.R., Yeung, W. 2000. Application of Global Positioning System (GPS) to Travel Time and Delay Measurements, University of Delaware, Newark, Delaware Center for Transportation, Newark NJ.


Pilli-Sihvola, Y. Floating Car Road Weather Information Monitoring System. Transportation Research Record: Journal of the Transportation Research Board of the National Academies, No. 1741, pp. 3-5, 2001.


http://www.paramics-online.com/


Quiroga, C.A., and Bullock, D. A Data Model For Managing GPS Travel Time Data, Fifth International Conference on Applications of Advanced Technologies in Transportation Engineering, American Society of Civil Engineers, Newport Beach, California, pp. 403-410. 1998


Quiroga, C.A., and Perez. GETT (GPS-Based Evaluation of Travel Time) 1.0. User Manual, Texas Department of Transportation, Austin District, Austin, TX, p.46, 2002

Sanav Inc., GL-50-BT Bluetooth GPS Receiver & Data Logger with SiRF III, Mobile GPS
Sanav Inc. http://www.sanav.com


Smith, B.L., Zhang, H., Fontaine, M., and Green, M. Cellphone Probes as an ATMS Tool. Research Project Report for the National ITS Implementation Research Center and US DOT University Transportation Center, Center for Transportation Studies at the University of Virginia, 2003.


Tech Japan Inc. Sony’s New Navigation System XYZ,


The Toronto Star, Homicides in the GTA, 2005-08. Available at:


Toronto Police Service, TPS Crime Statistics: Crime Indicators, 2005-09,


Welch, G. SCAAT: Incremental Tracking with Incomplete Information. TR96-051, October 1996 University of North Carolina at Chapel Hill, NC, 1996.


Appendices

Appendix A. Weights of the trained auto mode classifier for 10 pings per 10 minutes scenario. (one-hidden-layer M-P perceptrons in Neurosolutions Version 5)

Appendix B. Traffic count example of the intersection at Front Street and Church Street used for calibration of Paramics simulation
Appendix A.

Weights of the trained auto mode classifier for 10 pings per 10 minutes scenario.

(one-hidden-layer M-P perceptrons in Neurosolutions Version 5)

*Appendix A serves as a set of reference values that can be used to reproduce the experiment with Neurosolutions Version 5.

// Weights saved from breadboard.

#NSWeightFileVersion 235

#inputFile File (Weights of Speed and Acceleration and Average Number of Satellites in View)

21

1.636363636363636e-001
1.928227102303161e-002
1.928227102303161e-002
1.936316695352840e-002
1.936316695352840e-002
1.936316695352840e-002
1.936316695352840e-002
1.928227102303161e-002
1.928227102303161e-002
1.936316695352840e-002
2.313624678663239e-002
#desiredFile File

2

1.800000000000000e+000 -9.000000000000000e-001
1.800000000000000e+000 -9.000000000000000e-001
0

#inputAxon Axon

21 1

1

0

#hidden1Axon TanhAxon

10 1
<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
</table>

#outputAxon TanhAxon

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 1 2 4.151280251472518e-001 3.868221076082645e-002</td>
<td></td>
</tr>
</tbody>
</table>

#criterion L2Criterion

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 1 1 0 2.353434858241523e-001 10 3.306833094271676e-001 1.40583513901181e-001 -1.083559678945280e-001 -2.353434858241523e-001 -4.522995696890164e-001 1.501052888576921e-001 -4.846797082430494e-001 3.218024231696524e-001 4.375591296121097e-001</td>
<td></td>
</tr>
</tbody>
</table>

#hidden1Synapse FullSynapse

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
</table>
1.800000000000000e+000 -9.000000000000000e-001
1.800000000000000e+000 -9.000000000000000e-001

#CVOutputDesiredDataGraph DataGraph

2
1.800000000000000e+000 -9.000000000000000e-001
1.800000000000000e+000 -9.000000000000000e-001

#desiredViewer MatrixViewer

2
1.800000000000000e+000 -9.000000000000000e-001
1.800000000000000e+000 -9.000000000000000e-001

#CVDesiredViewer MatrixViewer

2
1.800000000000000e+000 -9.000000000000000e-001
1.800000000000000e+000 -9.000000000000000e-001

#hidden1SynapseBackpropGradient Momentum

210
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000 0.000000000000000e+000
#hidden1AxonBackpropGradient Momentum

10 0.0000000000000000e+00 0.0000000000000000e+00 0.0000000000000000e+00 0.0000000000000000e+00

#outputSynapseBackpropGradient Momentum

20 0.0000000000000000e+00 0.0000000000000000e+00 0.0000000000000000e+00 0.0000000000000000e+00
0.0000000000000000e+000 0.0000000000000000e+000 0.0000000000000000e+000 0.0000000000000000e+000 
0.0000000000000000e+000 0.0000000000000000e+000 0.0000000000000000e+000 0.0000000000000000e+000 

#outputAxonBackpropGradient Momentum

2 0.0000000000000000e+000 0.0000000000000000e+000
Appendix B

Traffic count example of the intersection at Front Street and Church Street used for calibration of Paramics simulation
# 24-Hour Count Summary Report

<table>
<thead>
<tr>
<th>FRONT ST</th>
<th>STAT CODE</th>
<th>ARTERY CODE</th>
<th>COUNT DATE</th>
<th>AM PEAK</th>
<th>AM PEAK HOUR</th>
<th>PM PEAK</th>
<th>PM PEAK HOUR</th>
<th>OFF HOUR PEAK</th>
<th>OFF HOUR PEAK HOUR</th>
<th>24 HOUR</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EASTBOUND</td>
<td>CATEGORY: 24 HOUR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRONT ST E/W OF CHURCH ST</td>
<td>26525</td>
<td>26525</td>
<td>5/3/05 Tue</td>
<td>626</td>
<td>08:45 - 09:45</td>
<td>1,144</td>
<td>16:45 - 17:45</td>
<td>190</td>
<td>14:15 - 15:15</td>
<td>11,321</td>
<td></td>
</tr>
<tr>
<td>FRONT ST E/W OF CHURCH ST</td>
<td>26525</td>
<td>26525</td>
<td>5/4/05 Wed</td>
<td>632</td>
<td>08:45 - 09:45</td>
<td>1,109</td>
<td>17:15 - 18:15</td>
<td>129</td>
<td>11:45 - 12:45</td>
<td>11,577</td>
<td></td>
</tr>
<tr>
<td>FRONT ST E/W OF CHURCH ST</td>
<td>26525</td>
<td>26525</td>
<td>5/5/05 Thu</td>
<td>616</td>
<td>08:30 - 09:30</td>
<td>1,227</td>
<td>17:00 - 18:00</td>
<td>143</td>
<td>11:45 - 12:45</td>
<td>12,249</td>
<td></td>
</tr>
<tr>
<td>EASTBOUND TOTAL</td>
<td></td>
<td></td>
<td></td>
<td>1,873</td>
<td>3,480</td>
<td>2,162</td>
<td></td>
<td></td>
<td></td>
<td>35,047</td>
<td></td>
</tr>
<tr>
<td>EASTBOUND AVERAGE</td>
<td></td>
<td></td>
<td></td>
<td>624</td>
<td>1,160</td>
<td>721</td>
<td></td>
<td></td>
<td></td>
<td>11,682</td>
<td></td>
</tr>
<tr>
<td>FRONT ST</td>
<td></td>
<td></td>
<td></td>
<td>1,673</td>
<td>3,480</td>
<td>2,162</td>
<td></td>
<td></td>
<td></td>
<td>35,047</td>
<td></td>
</tr>
</tbody>
</table>