Development of Real-Time Traffic Adaptive Crash Reduction Measures for the Westbound I-94/35W Commons Section

Final Report

Prepared by:
John Hourdos
Wuping Xin
Panos Michalopoulos

Department of Civil Engineering
University of Minnesota

CTS 08-28
Technical Report Documentation Page

CTS 08-28

2.  

3. Recipients Accession No.  

4. Title and Subtitle  
Development of Real-Time Traffic Adaptive Crash Reduction Measures for the Westbound I-94/35W Commons Section

5. Report Date  
December 2008

6.  

7. Author(s)  
John Hourdos, Wuping Xin, Panos Michalopoulos


9. Performing Organization Name and Address  
Department of Civil Engineering  
University of Minnesota  
500 Pillsbury Drive S.E.  
Minneapolis, MN 55455

10. Project/Task/Work Unit No.  
CTS project # 2005056

11. Contract (C) or Grant (G) No.  

12. Sponsoring Organization Name and Address  
Intelligent Transportation Systems Institute  
University of Minnesota  
200 Transportation and Safety Building  
511 Washington Ave. SE  
Minneapolis, Minnesota 55455

13. Type of Report and Period Covered  
Final Report


15. Supplementary Notes  
http://www.cts.umn.edu/Publications/ResearchReports/

16. Abstract (Limit: 200 words)  
Much research has been conducted in the development, implementation, and evaluation of innovative ITS technologies aiming to improve traffic operations and driving safety. An earlier project succeeded in supporting the hypothesis that certain traffic conditions are favorable to crashes and in developing real-time algorithms for the estimation of crash probability from detector measurements. Following this accomplishment a natural question is “how can this help prevent crashes?” This project has the ambitious plan of not only answering this question but also providing a multifaceted approach that can offer different types of solutions to an agency aimed at reducing crashes in this and other similar locations. This project has two major objectives; first it aimed at utilizing a cutting edge 3D virtual reality system to design and visualize different driver warning systems specifically for the I-94 westbound high crash location in Minneapolis, MN. Second, in view of the desire of local engineers for a more traditional approach, this project explored the use of existing micro-simulation models in the evaluation of safety improvements for the aforementioned high crash area. This report describes the results of these investigations but more importantly describes the lessons learned in the process of the research. These lessons are important because they highlight gaps of technology and knowledge that hampered this and other research projects with similar objectives.

17. Document Analysis/Descriptors  
Traffic safety, rear-end crash, microscopic simulation, crash causal analysis, statistical inference, car-following models, traffic shockwaves, flow breakdown

18. Availability Statement  

19. Security Class (this report)  
Unclassified

20. Security Class (this page)  
Unclassified

21. No. of Pages  
119

22. Price  

Development of Real-Time Traffic Adaptive Crash Reduction Measures for the Westbound I-94/35W Commons Section

Final Report

Prepared by

John Hourdos
Wuping Xin
Panos Michalopoulos

Department of Civil Engineering
University of Minnesota

December 2008

Published by

Intelligent Transportation Systems Institute
Center for Transportation Studies
University of Minnesota
200 Transportation and Safety Building
511 Washington Ave SE
Minneapolis, MN 55455

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. This report does not necessarily reflect the official views or policy of the Intelligent Transportation Systems Institute or the University of Minnesota.

The authors, the Intelligent Transportation Systems Institute, the University of Minnesota and the U.S. Government do not endorse products or manufacturers. Trade or manufacturers’ names appear herein solely because they are considered essential to this report.
Acknowledgements

This study followed on the footsteps of the “Accident Prevention Based on Automatic Detection of Accident Prone Traffic Conditions: Phase I” project which used the advanced detection and surveillance stations maintained by the Minnesota Traffic Observatory to analyze the causes for crashes in the most frequent crash section in Minnesota. Support for this study was provided by the ITS Institute of the University of Minnesota. The authors wish to thank Professor Gary Davis of the University of Minnesota for his assistance in this project and the Minnesota Department of Transportation for their support in data collection, field survey work and, for implementing some of the projects recommendations.

The authors wish to acknowledge those who made this research possible. The study was funded by the Intelligent Transportation Systems (ITS) Institute, a program of the University of Minnesota’s Center for Transportation Studies (CTS). Financial support was provided by the United States Department of Transportation Research and Innovative Technologies Administration (RITA).
# Table of Contents

1 Introduction .......................................................................................................................... 1
   Primary Objective .................................................................................................................. 2
   Secondary Objective ............................................................................................................ 3

2 Existing Crash Prevention Systems ...................................................................................... 6
   Manual Driver Alert Systems ............................................................................................... 6
   Automated Driver Alert Systems .......................................................................................... 7

3 Crash Causes and Mechanism ............................................................................................ 10
   Introduction ......................................................................................................................... 10
   Correlation and Causation ................................................................................................. 11
   Research Model for the Identification of Causal Structure ................................................ 12
   Methodology ....................................................................................................................... 14
      Reliable Associations ........................................................................................................ 14
      Empirical Causal Connections ....................................................................................... 15
   Crash Causal Mechanism ................................................................................................... 18
   Causality Investigation ....................................................................................................... 18
      Study Area and Measurements ....................................................................................... 18
      Suspected Causal Factors ............................................................................................... 20
      Preliminary Causal Structure ....................................................................................... 24
      Modeling the Flow Breakdown ..................................................................................... 25
      Location Specific Preliminary Crash Causal Structure .................................................. 32
      Boundary Conditions .................................................................................................... 36
      “I-94 Flyover” Ramp Conditions ................................................................................... 37
      Post-Portland Right Lane Conditions ............................................................................ 37
      Post-Portland Middle Lane Conditions .......................................................................... 38
      Combined Effects ........................................................................................................... 39
      Graphical Causal Model Refinement ............................................................................. 41
   Conclusions ......................................................................................................................... 41

4 Immersive 4-D Visual Simulation of Driver Warning Systems ............................................ 45
   Introduction ......................................................................................................................... 45
   DEN 3D Environment And Object Geometry Modeling Process ........................................ 46
      3D Vehicle Models .......................................................................................................... 53
   DEN Microsimulation Vehicle Trajectory Emulation .......................................................... 54
      3D Road and Microsimulation Road Geometry Matching .............................................. 54
      Real-Time 3D Visualization of Vehicle Trajectories ...................................................... 55
      Limitations Of Visual Simulation of Vehicle Motion Trajectories .................................. 59
   Current Limitations of Simulation Observation ................................................................ 59
      In Vehicle Observation ................................................................................................... 59
      Simulated Visual Driver Field Of View ........................................................................... 60
      Driver Warning Subsystems ......................................................................................... 62
   Conclusions ......................................................................................................................... 62

5 Car-Following Models: Historical Review and Implications on Freeway Traffic Safety Modeling ....................................................................................................................... 63
   Issues and Current Solution ............................................................................................... 63
List of Tables

Table 3.1 Estimation summary for logit model predicting flow breakdown.................................28
Table 3.2 Logit modeling of flow breakdown likelihood..............................................................29
Table 3.3 Univariate analysis of all available variables. ...............................................................33
Table 3.4 Logit model #1: measurements at the boundaries. .........................................................36
Table 3.5 Logit model #2: addition of measurements at the I-94 flyover entrance ramp..............37
Table 3.6 Logit model #3: addition of measurements on the right lane downstream of Portland Ave. .....................................................................................................................38
Table 3.7 Logit model #4: addition of measurements on the middle lane downstream of Portland Ave. .....................................................................................................................39
Table 3.8 Logit model #4: addition of measurements on the middle lane downstream of Portland Ave. .....................................................................................................................40
Table 4.1 Point elevation errors along image pair boundaries.......................................................48
Table 4.2 Vehicle model Level Of Detail (LOD) node structure. ...................................................53
Table 6.1 Error statistics and optimal number of local polynomials when post-processing vehicle trajectories from the crash occurred on Aug 8, 2006. ...........................................86
Table 7.1 Safety surrogate measure for traffic conflict. .................................................................96
Table 7.2 Potential surrogate measures in a micro-simulator in our project. ..............................100

List of Figures

Figure 1.1 I-94 westbound freeway simulation model (green) and high crash location (red)........5
Figure 2.1 Lowry Hill tunnel lane signs. Minneapolis, MN........................................................7
Figure 2.2 Tokyo - Osaka freeway rear-end crash prevention system.......................................8
Figure 2.3 Sangubashi curve rear-end crash prevention system...............................................9
Figure 3.1 Possible causal diagrams for lung cancer.................................................................12
Figure 3.2 Possible relationships between crash and speed variance......................................15
Figure 3.3 Data collection locations. Source: Mn/DOT web site.............................................19
Figure 3.4 V/O graph, right lane of section downstream of Portland Ave. August 11th, 2005. ....21
Figure 3.5 (a) Flow in vph moving from the auxiliary lane to the middle lane.
(b) Right lane speed and right – middle lane speed difference............................................23
Figure 3.6 Comparison between speeds on the right and middle lanes 1 minute prior to crash..........................................................24
Figure 3.7 Rough causal structure based on prior knowledge..................................................25
Figure 3.8 Flow chart of the preliminary flow breakdown causal structure..........................27
Figure 3.9 Final flow chart of flow breakdown causal structure.............................................31
Figure 3.10 Final causal model of flow breakdown.................................................................31
Figure 3.11 Location specific flow chart of preliminary causal structure based on prior knowledge...........................................................34
Figure 3.12 Location specific preliminary causal model based on statistical covariation and prior knowledge.................................................................................................35
Figure 3.13 Critical values of speeds: middle, right, and difference...........................................42
Figure 3.14 Flow chart of crash causal structure as supported by the analysis results..............43
Figure 3.15 Crash causal model as supported by the data..........................................................44
Figure 4.1 Stereo pair coverage tiles covering the Interstate 35W and Interstate 94 commons freeway section. The high crash rate area is within MnnW_3...........................................47
Figure 4.2 Epipolar geometry, the base line, CC’, and CX, C’X, all lie in the same plane (shaded). For the flight path data, overlapping aerial image pairs are reprojected as in (b) onto common plane parallel to a line between optical centers (in direction of flight path)..........................................................................................................................47
Figure 4.3 Extracted road and overpass edges of the Interstate 35W/94 commons from east of Cedar Avenue up to the lowery tunnel...............................................................48
Figure 4.4 The view shown is just beyond the over pass, which is immediately westward and downstream of most crash and near miss occurrences. Lane marking strips are drawn and projected on-top of the ortho-corrected texture (processed within the SimWright pluggin). ..........................................................................................................49
Figures 4.5 (a – g) Final 3D road model, progressing west bound as indicated on map. The critical accident area is just beyond D and F.......................................................49-53
Figure 4.6 software configuration for the application developed for the DEN is indicated above the dotted line. The previously developed DEN base class libraries are below the dotted line.....................................................................................................................55
Figure 4.7 Z-elevation LUT ‘voxel’ (a x,y,z cell point in the table) rendering of roadway. The different colors represent the ‘layer’ attribute, h, in the elevation LUT which in this case is due to road overpasses.....................................................................................58
Figure 4.8 Scaled DEN placed ‘inside’ a Chevrolet Blazer SUV. Note field of view (indicated by shaded frustums) will be affected by observer’s head movement. ..............61
Figure 5.1 Measured car-following behavior. Data were collected from a Japan test track using GPS. This type of non-damped oscillations is typical of car-following process, sometimes referred to as “closing following spirals.” ...............................................................72
Figure 6.1 Proposed 3-stage data collection and processing methodology....................................77
Figure 6.2 Schematic work flow of the spline based bi-level programming process.........................82
Figure 6.3 (a) Data collection site: a high crash-rate freeway section of Interstate-94 WB..........84
Figure 6.3 (b) Ten highest crash sections in 2002 ........................................................................84
Figure 6.3 (c) Surveillance cameras that monitor the site. .............................................................84
Figure 6.4 Trajectories of 8 vehicles in a crash happened on 08-08-2006 (w/o post-processing) .......................................................................................................................85
Figure 6.5 Post-processing the trajectory of Veh09 of the crash occurred on Aug 8, 2006 ..........87
Figure 6.6 Data used in sensitivity test: ground truth vs. ground truth added with noise...........91
Figure 6.7 Comparing the bi-level optimization algorithm to locally weighted regression: mean absolute errors between ground truth position data and filtered position data.........92
Figure 6.8 (a) Comparing the bi-level optimization algorithm to locally weighted regression: mean absolute errors for position inconsistencies. ..................................................93
Figure 6.8 (b) Comparing the bi-level optimization algorithm to locally weighted regression: mean absolute errors for speed inconsistencies.................................94
Executive Summary

Much research has been conducted in the development, implementation, and evaluation of innovative ITS technologies aiming to improve traffic operations and driving safety. The “Accident Prevention Based on Automatic Detection of Accident Prone Traffic Conditions: Phase I” project, concluded earlier this year, succeeded in supporting the hypothesis that certain traffic conditions are favorable to crashes and in developing real-time algorithms for the estimation of crash probability from detector measurements. Following this accomplishment a natural question is “how this can help prevent crashes?” This project had the ambitious plan of not only answering this question but also to provide a multifaceted approach that can offer different types of solutions to an agency aiming in reducing crashes in this and other similar locations. This project had two major objectives; first it aimed in utilizing a cutting edge 3D virtual reality system to design and visualize different driver warning systems specifically for the I-94 westbound high crash location in Minneapolis, MN. Second, in view of the desire of local engineers for a more traditional approach, this project explored the use of existing micro-simulation models in the evaluation of safety improvements for the aforementioned high crash area. Although, in the course of this project, unforeseen difficulties did not allow for the complete accomplishment of the aforementioned objectives, the experience produced a number of very useful and important findings. In the course towards the first objective a new 3D visualization approach was designed that allow for the faithful reproduction of real roadway environments in a computer model. This method utilizes data from the USGS, road construction plans and simple photos to create very accurate, life size environments. In the process of accomplishing the second objective two issues became apparent. The first deals with the inability of all currently available micro-simulation traffic models to produce realistic vehicle collisions and the second was that although it was possible to create algorithms that estimate crash probability, solutions must target the causes for the crashes while models need to replicate the unsafe behavior of drivers involved in these crashes. These realizations steered the project progress away from the original objectives into utilizing its resources in placing the foundations for subsequent research in new crash capable driver behavior models (successful result of the subsequent “Enhanced Micro-Simulation Models for Accurate Safety Assessment of Traffic Management ITS Solutions” project), a PhD Thesis describing a new methodology for the causal analysis of high crash freeway sections, and immediate solutions for reducing crash risk in this high crash area. The latter have been successfully implemented by Mn/DOT in the field.
1 Introduction

According to Mn/DOT statistics, the westbound section of I-94 at the 94/35W commons in the south of downtown Minneapolis, MN is the highest crash location in the Twin Cities. In simpler terms, there is one crash every two days approximately. Research on detection of accident prone conditions and accident prevention has been carried out at the University of Minnesota since 2001. Although preliminary results showed promise it wasn’t until the design, implementation and field deployment of advanced detection and surveillance stations [Hourdos et al., 2004] at the I-94/35W commons section that real crashes could be recorded and crash prone conditions detection algorithms developed. The detection and surveillance system integrates machine vision sensors for data collection, compressed digital video for surveillance and wireless communications for information retrieval and remote control. This resulted in a unique laboratory for performing a variety of traffic flow and safety studies. Four of these advanced detection systems were deployed at the 1.7 mile long site for wireless, continuous coverage allowing detailed real time data collection and surveillance.

The four detection and surveillance stations overseeing the deployment site are a unique source of data, observations, and information. The detail and amount of wide area measurements in conjunction with the video recordings represent a unique database which allows advanced studies of traffic characteristics, flow dynamics as well as safety and traffic management concepts. The sensors collect and transmit in real time individual vehicle speed, length, time headways as well as aggregated 10 seconds measurements of flow rate, volumes, time and space mean speed, space occupancy, density, LOS and vehicle classification. The aforementioned data are collected on a 24/7 basis in six areas of the freeway site. Each of these areas is 300 to 500 feet long and measurements are extracted approximately every 100 feet for a total of 51 detection points (18 series times 3 lanes, plus exit and entrance ramps). The laboratory has been operational during several heavy Minnesota winters and stormy summer seasons providing detailed measurements during extreme weather conditions. For data management purposes, video recording takes place between 7:00AM and 8:00PM during weekdays and 12:00PM until 8:00PM Saturdays and Sundays. Currently, more than 150 crashes have been recorded and stored over a period of 12 months along with 215 unreported near crashes. With the exception of a few days where transmission was interrupted due to malfunctions of the Ethernet radios, traffic measurements were collected continuously during this period resulting in more than 5000 hours of detailed traffic measurements both microscopic and macroscopic.

From the qualitative analysis of the measurement time series plots it was concluded that frequently there is a significant increase in speed variance and traffic pressure before the time at which the crashes and near crashes occurred. This indicates that traffic conditions prior to crashes are significantly different from the norm. In addition, analysis was carried out for speed variance, traffic pressure and kinetic energy, using binary logit models. Normal traffic conditions (crash free), in 20-minute periods, were selected in periods with the same environmental and traffic characteristics as the accident ones. In addition to the traffic measurements other contributing factors considered by the model included lighting conditions, position of sun, weather, time, and day of accident occurrence. The analysis suggests that speed variance, traffic pressure and kinetic energy are statistically significant for detecting crash prone conditions.
The next logical step was to find a way to utilize these relationships in the detection of crash-prone conditions. The method that showed the greatest potential was based on statistical analysis and specifically in the development of models for the estimation of crash likelihood. Based on the real-time measurements and metrics defined, many metric variants were developed. Through the process of logistic regression the variants that exhibited the largest correlation with pre-crash conditions were incorporated into crash likelihood estimation models. These models were then integrated into crash-prone condition detection algorithms combining them with more realistic detection aggregation periods as well as a heuristic test that enhanced performance. The crash-prone traffic condition detection algorithms were tested for 10 days of varying weather and traffic conditions not earlier employed in the model & algorithm development. From a number of threshold groups identified through a trial and error process, detection performance curves were developed (Hourdos et al., 2005 and Hourdos et al., 2006).

The aforementioned research succeeded in proving the existence of crash-prone conditions and presented tools for their real-time detection. Algorithms with similar structures have been developed in the past for the detection of freeway incidents. In the case of incidents the next step was simple, following the incident detection, the traffic operators verified and initiated the relevant procedures for clearing it up and restore road capacity. In the case of crash-prone conditions, even with an alarm signaling that crash probability is high, response is not that simple. Since no incident has happened yet traffic engineers need to employ proactive methods influencing driver behavior and in extend traffic flow conditions to reduce the crash probability. There are several possible ways this can be achieved, some of them are:

- Driver warning systems. Infrastructure or in-vehicle.
- Traffic control strategies like variable speed limits, lane changing prohibitions, or
- Demand control strategies like ramp metering and forced diversion.

**Primary Objective**

The project presented in this report proposed to investigate systems belonging mainly in the first two categories. Considering the plethora of different alternative designs and recognizing that it would be impossible to test all of them in a driving simulator with human subjects this project proposed the following staged approach:

1. Tasks 1 and 2 of the project aimed in defining a number of possible driver warning and traffic control system designs. This first task of the project was a combination of literature and market search along with in-house engineering expertise. Since effective countermeasures are the ones that target the causes of the crash problem, the project team quickly realized that it would not be possible to propose any system designs before performing a causal analysis for the crashes on I-94 Westbound. While the literature and market review can be found in the next chapter (chapter 2) the causal analysis proved to be a lot more complicated requiring a new methodology to be developed. Such a methodology and its application in this freeway site were proposed by Hourdos in a successful PhD dissertation. The dissertation inspired and supported by this project has been accepted by the research community as innovative. An excerpt from the dissertation dealing with the causal analysis of freeway crashes can be found in Chapter 3 of this report.
2. Tasks 3, 4, and 5 were dedicated in the design and development of a realistic representation of the I-94 high crash area inside a 3D virtual reality visualization. Such a system, similar to the visualization system of a driving simulator, would allow the researchers to experiment with different driver warning system design alternatives. Considering that the geometry and traffic weaving patterns are integral parts of the crash causes the regular “flat” 3D visualization many times employed in driving simulator experiments would not have been enough. Instead this experiment required a faithful representation of the real site. Such an objective proven to be a lot more complicated and time consuming than originally anticipated and while it was accomplished it drained project resources to the level where actual implementation of driver warning systems and experimentation was not possible (tasks 5 to 9 were not accomplished). Regardless, the exercise of developing such a unique visualization has its own rewards and lessons learned. The process and the lessons learned from this experience can be found in chapter 4 of this report.

Secondary Objective

The project, recognizing the needs of transportation engineers for simpler solutions, had a secondary objective. During the last decade, traffic simulation has grown to be an integral part of transportation engineers both in designing new facilities as well as optimizing the operations of existing ones. Although traffic simulators have proved to be a great tool in terms of traffic operations they have severe limitations in the area of traffic safety. The project’s secondary objective was to set the foundation for further development on micro-simulation models by examining the safety related potential of existing simulators. This objective was approached from the following two separate angles:

1. Tasks 10 and 12 set the foundation for research in new car-following models that can produce crashes by emulating the behavior of the “less than perfect” driver. First an extensive investigation of the state-of-the-art in microscopic traffic modeling was performed. The result of the investigation is described in chapter 5 of this report. To assist subsequent research developing new car-following models this project invested in understanding the details of the crashes in the I-94 site. This investigation was performed from the macroscopic view, described in the thesis by Hourdos [Hourdos, 2005] and summarized in chapter 3, to the microscopic individual vehicle interactions by extracting and analyzing individual vehicle trajectories of vehicles involved in crashes. Specifically for the latter a new methodology was developed [Xin et al., 2008] to improve the data produced from the NG-VIDEO system [Zhang et al. 2006]. This methodology and a summary of the extracted data can be found in chapter 6 of this report. The work on tasks 10 and 12 has generated two peer reviewed papers by Xin, Hourdos, and Michalopoulos, 2008 and Xin, Hourdos, Michalopoulos, and Davis, 2008 as well as one successful peer reviewed project “Enhanced Micro-Simulation Models for Accurate Safety Assessment of Traffic Management ITS Solutions” [Xin et al., 2008]

2. Task 11 of the project aimed on a more direct approach of utilizing simulation for freeway safety assessment. This approach was to utilize an existing microscopic simulator in the extraction of surrogate measures of safety. These measures theoretically have the same trend as crash probability therefore can potentially be used to rank alternative designs in terms of safety regardless if they are unable to predict expected
number of crashes. A limited review of the literature regarding Surrogate Safety Measures can be found in chapter 7. Similar work to this was accomplished in parallel by the Office of Safety RD&T Turner Fairbank Highway Research Center [Gettman et al., 2008]. The scope of the task was to replicate the I-94 high crash area in the AIMSUN micro-simulator, collect enough information in order to construct reliable Origin/Destination matrixes for 8 hours of simulation, and calibrate. Having a reliable model at hand we were to enhance the application in order to extract surrogate measures of safety. Specifically, Acceleration Noise [Herman 1959, Montroll 1961], Speed Variance, and Unsafe Headway [Torday et al., 2003]. Finally, having a platform where one can compare different freeway geometric designs and rank them in terms of safety; two alternative designs for complete area reconstruction were implemented and compared. For the last step the research team worked with Mn/DOT to acquire the designs of two proposed designs created under contract by CH2M Hill for the Downtown Minneapolis Freeway study. These designs were total area reconstructions specifically addressing the safety and congestion concerns in both the I-94 and I-35W freeways. Unfortunately, due to lack of resources and other problems encountered this task was not completed although it reached its final stages. Specifically, The following steps were accomplished:

a. Modeling of current I-94 Westbound from TH-280 (3 miles upstream of high crash area) to Broadway Avenue (2 miles downstream of high crash area). A map of the simulated area can be seen in Figure 1.1.

b. Manual collection of O/D patterns between key entrances and exits. License plate capture in video was used for the task.

c. Acquisition of area O/D matrixes from the CH2M Hill team and adjustment based on the manual counts and loop detector data.

d. Very detailed calibration of the micro-simulation model with validation of individual lane volume and speed patterns.

e. Modeling of two alternative geometric designs for the area.

f. Development of an API extension to the simulator for the extraction of surrogate safety measures based on individual vehicle trajectories. The final application was not tested although it passed preliminary validation based on historical crash frequencies per freeway segment.
Figure 1.1 I-94 westbound freeway simulation model (green) and high crash location (red). Source: Google Maps.
2 Existing Crash Prevention Systems

Development and implementation of crash prevention systems can be divided into two categories. The first category deals with In-vehicle crash avoidance systems like adaptive cruise control, forward and backward collision avoidance systems, run off-road driver assist systems, as well as the larger Intelligent Vehicle Highway System program aimed at removing the driver from the equation. The majority of these systems depend on instruments in the vehicle which measure speed, headway, acceleration/deceleration, and distances from the surrounding vehicles and use these to influence the vehicles trajectory and avoid a collision or running off the road. In general, none of these systems depend on information concerning the general traffic flow conditions or even what happens 2-3 vehicles ahead. Although a number of these systems have reached the market, very few vehicles in the road are currently equipped with such systems. In a study by Rajamani and Levinson (2005) it was illustrated through microscopic simulation that unless the market penetration of, in their case, adaptive cruise control systems exceeds 90% there will be no visible improvement in the traffic stream. Considering the scope of this research no further discussion will be devoted to in-vehicle systems.

The second category, the one where this works belongs, covers infrastructure crash prevention systems. Unlike the first category, very few advances have taken place in the development and implementation of infrastructure-based crash prevention systems. The few that have been developed and deployed are either manual in operation (require human intervention) or very simplistic in the underlying automation. For completeness, three examples of infrastructure based systems are presented in this section.

Manual Driver Alert Systems

The most common example of manually operated driver alert system is the decades old Highway Advisory Radio (HAR) system. In the United States, this system is comprised of short range FM transmitters manually activated by the traffic operators. Electronic signs or Variable Message Signs (VMS) on the side of the road alert the drivers about a problem downstream and prompt them to tune-in to the prescribed frequency for more details. In general HAR is used for incident management alerting vehicles upstream of the location of a crash, the approaching congestion, and of possible alternate routes. In Europe HAR-like systems, called RDS-TMC, are more sophisticated in the sense that each vehicle’s radio is capable of automatically tuning to the emergency frequency, therefore not requiring the use of special roadside signs. As an incident management tool HAR has proven very successful but is still a reactive measure in need of human control (I-95 Corridor Coalition Report, 2001).

Another manually operated system commonly encountered in tunnels involves Lane Signs. As can be seen in Figure 2.1 lane signs are dynamic displays capable of changing from a green arrow indicating no problem ahead, to a yellow arrow signaling congestion in the lane downstream of the present location, or a red X signaling that the lane is blocked either in the tunnel or immediately after it. Lane signs are usually activated manually by the traffic operators, although a few automated systems exist in European tunnels. Such implementation depends on sensors measuring the speed of the stream in each lane and/or the existence of stopped vehicles.
In Europe, as part of the TABASCO and INFOTEN projects, a freeway driver alert system has been deployed. This system, called COMPANION, depends on Light Sticks on the side of the road. These light sticks resemble the traditional reflective sticks usually found on the roadside in locations where heavy snowfall is frequent, but are equipped with dynamic LEDs capable of changing color and flashing pattern. In their current state the light sticks are manually activated and change from yellow to red signals varying congestion and lane blocks downstream.

**Automated Driver Alert Systems**

Automated driver warning and crash prevention systems based on dynamic weather and pavement information have been deployed in the field during the last ten years. These systems depend on sensor information on rain, fog, and pavement ice formation to trigger warning mechanisms like VMS, variable speed limit signs, entrance ramp gates, and de-icing systems. None of the deployed systems take into account the prevailing traffic conditions.

The most notable effort in the development and deployment of crash prone traffic condition detection and alert systems is currently active in Japan. Under the Advanced cruise-assist Highway System (AHS) program, two systems relevant to this research are currently in the field deployment and evaluation stage. The first is an automated roll-over alert system. This system is comprised of several sensors providing information on individual vehicle speed, height, and weight and estimates the probability this vehicle might roll-over in the approaching sharp curve. If this probability is greater than a defined threshold variable, message signs and HAR-like radio alert the specific driver of the danger and require him to slow down. These systems are based on simple physical models for their determination of the roll-over probability and are mainly aimed at preventing large truck crashes.

The second system developed under the AHS program is more relevant to the scope of this research. It is an automated rear-end crash prevention driver alert system aimed at roadway sections with geometric designs that limit visibility. Two such systems have been deployed in high crash areas in Japan. The first system is on the freeway connecting Tokyo and Osaka (Figure 2.2) while the second system is located in the Sangubashi section of the Shinjuku Line on Tokyo Metropolitan Expressway No. 4 (Figure 2.3). Both deployments are very similar. The problem is that the sharp curve of the roadway limits visibility. Additionally, in both locations...
frequent congestion, either recurrent or non-recurrent, causes vehicles to queue-up immediately after the curve. The forming queue is not visible to the approaching vehicles, resulting in frequent and sometimes fatal rear-end crashes. The solution in both cases involves traffic sensors detecting the formation of a queue, or in the case of a previous crash, lane blockages and the automatic activation of several driver warning devices. The method of the warning depends on the distance to the curve and can be seen in Figures 2.2 and 2.3.

The logic in these systems is simple. If there is a queue in the predefined position or any other lane obstruction, the warning mechanism is activated. These systems are not sensitive to the level of congestion or the existence of compression waves in the traffic stream.

![Figure 2.2 Tokyo - Osaka freeway rear-end crash prevention system.](image)
Figure 2.3 Sangubashi curve rear-end crash prevention system. Source: http://www.ahsra.or.jp/eng/index_e.htm.
3 Crash Causes and Mechanism

(Excerpt from Hourdos 2005 PhD dissertation)

The present contains nothing more than the past, and what is found in the effect was already in the cause. 
Henri Bergson (1859 - 1941)

Introduction

When considering the plethora of roadway geometric designs combined with all the possible driving behaviors, the goal of determining all possible crash causes and mechanisms, even if one focuses only on high crash areas, sounds unattainable. However, it is possible to develop investigative methodologies that organize the data analysis and possibly incorporate methods infrequently utilized in transportation studies. This chapter develops such a methodology. Specifically, by combining research methodologies used in medicine with elements of inferred causation from Artificial Intelligence (AI) and traffic flow theory concepts, we were able to investigate the crash causal factors in the I-94 high crash section. Although this methodology has been developed around a specific example, we believe that the underlying analysis can be replicated in other high crash freeway sections.

The causality investigation methodology presented in this chapter can be divided into three major sections. Initially, the differences between correlation and causation are presented with help from the literature. In this section it is emphasized that correlation does not guarantee causation and that only through experimentation can one truly identify causal factors. Unfortunately, in this and many other problems, experimentation is not feasible and only observational studies are possible. Based on this constraint, methodologies originally developed for medical research are adapted for the purposes of this work. This section concludes with an outline of the procedure to be followed as well as the additional tools, apart from correlation, available in the causality investigation.

The second section presents a preliminary investigation of possible causal factors. The first causal factor identified, based on results presented in prior chapters, is flow breakdown. The flow breakdown phenomenon is investigated following a procedure similar to the bigger crash causality investigation. A flow breakdown mechanism is proposed and refined according to the available real-time measurements and traffic flow principles. The final section expands the investigation towards the identification of crash causal factors and mechanism. The process followed is based on a combination of statistical covariation, tests of conditional independence, and prior knowledge of temporal/spatial order and traffic flow principles. A graphical model of the crash mechanism is proposed and refined based on the above cues. The final causal model presented supports the existence of three causal factors: flow breakdown, drivers’ selection of short headways, and driver distraction. As supported by the analysis the flow breakdown is caused by a combination of the traffic coming down the I-94 flyover ramp and the traffic on the right lane upstream of the merge. The interesting finding is that the downstream conditions at the
Lowry Hill Tunnel are not associated with the flow breakdown. Although it was not possible to explore in detail the reasons behind the selection by the drivers of dangerously short headways on the right lane it was possible to investigate further the possible causes for driver distraction. As explained in detail in later sections, the merge area upstream of Portland Avenue presents some challenging lane changing conditions attracting the attention of the drivers from looking ahead for possible attitude changes by the leading vehicles. Specifically, the speed differential between the right and middle lanes is identified as a possible causal factor for crashes in the study area.

**Correlation and Causation**

Williams et al. (1982) supported the concept that road crashes result from combinations of a variety of interacting factors rather than from single identifiable causes. Chira-Chavala et al. (1986) stated that crashes are complex phenomena and the problems at a location are likely to be site specific. However, Baldwin (1966) supported the idea that all crash causes are equal if, by the elimination of any one of the crash causal factors, the crash might have been prevented. The last statement is in agreement with the general idea of causation given by Baker (1990) who defined a “causal factor” as any circumstance “contributing to a result without which the result could not have occurred.” The former two opinions, although not in perfect agreement, do not contradict the general definition since it does not exclude the possibility of multiple causal factors. The study of causation is central to the understanding of human reasoning. Diagnosis, be it in traffic safety or epidemiology, depends on finding a satisfactory explanation for a given set of observations and the meaning of explanation is intimately related to the notion of causation. Diagnostic methods are generally divided into two categories: observational and experimental.

As Glymour (1987) notes, modern science greatly depends on experimental methods. Even so, many questions cannot be answered by experiments, and many answers did not come from experimental studies. The limitations of experimentation are both practical and ethical. For practical reasons it is not possible to do experiments with the economies of nations or with the arrangements of galaxies. For ethical reasons, it is not possible to do experiments in humans on the cause of disease or on the causes of traffic crashes. Faced with both the urgent need for knowledge and with stringent limitations on the scope of experimentation, researchers resort to observation and statistics. In traffic safety it is rarely possible to conduct proper randomized experiments; in the majority of cases, observational study of measurements and statistical modeling is the only viable method.

As stated in the introduction chapter, in recent years, several observational studies exploring the association of traffic conditions (as they manifest themselves through measurements) with crash events, have gained popularity. Indeed, in a work by Lee et al. (2002) the authors go beyond establishing associations, in identifying crash precursors. Although, no specific connection is made between crash precursors and crash causal factors the similarity between concepts is clear. So far, such observational methodologies when used to establish reliable associations between crash events and traffic conditions, are unable to distinguish between legitimate causal connections and spurious covariation. This analysis limitation, although recently encountered in traffic safety, is not new in the field of medicine.
According to Thagard (1998) an explanation of why people get a particular disease usually begins by noticing associations between the disease and possible causal factors. For example, a progression from correlation to causation has taken place with various kinds of cancer. In the eighteenth century, rough correlations were noticed between cancers and various practices: using snuff and nose cancer, pipe smoking and lip cancer, chimney sweeping and scrotum cancer, and being a nun and breast cancer [Proctor, 1995]. A more recent example is the landmark case of the causal connection between smoking and lung cancer. Lung cancer rates increased significantly in Great Britain and the United States during the first half of the twentieth century, correlating with an increase in smoking. By 1950 observational studies in several countries had established very strong associations between smoking and lung cancer. Regardless of these strong associations, the causal connection between smoking and lung cancer was not accepted until 1964 largely due to R. A. Fisher, a fierce experimentalist, who was prominent in declaring, in 1957, that the case against tobacco smoking was not proven since observational studies can be misleading. Amongst other arguments, the most compelling explanation he gave was that there may be a genetic factor which predisposes people both to smoke and to lung cancer. Indeed, there is evidence of a genetic component to smoking. Fisher showed that among pairs of identical twins where one member smokes, the other twin is more likely also to smoke than is the case for fraternal twins (who are no closer genetically than other brothers and sisters) thus implying that smokers and nonsmokers have different genotypes. Fisher’s argument illustrated the inability of observational studies to distinguish between the two graphs in Figure 3.1. It took several years and a study examining pairs of identical twins where only one of the two was a smoker, to statistically demonstrate that when controlling for the gene factor, smoking and lung cancer are not conditionally independent. Although the association is now well accepted, the mechanism by which smoking causes lung cancer is still a mystery possibly involving little understood cell mutations.

![Figure 3.1 Possible causal diagrams for lung cancer.](image)

**Research Model for the Identification of Causal Structure**

Prompted by the aforementioned inability of statistical covariation to reliably identify causation, but unable to resort to experimentation, several researchers as well as philosophers searched for study models able to guide and organize observational studies in uncovering causal relationships. One such research model was proposed by Thagard (1999) for guiding medical research. This model has the following three stages:

- **Stage 1**: Identify reliable associations.
- **Stage 2**: Empirically establish causal connections.
- **Stage 3**: Identify mechanism explaining causal connections.
The investigation methodology followed in this study is based on the above model. The available studies and methodologies identifying associations between traffic conditions and crash events are still floundering around stage 1, in large part due to the temporal/spatial scales for which data are available. In this study, the analysis was at much finer spatial (a single high crash area) and temporal (detailed traffic measurements) scales, in the hope of moving from stage 1 to stage 2 and, at least partially, to stage 3. Thagard’s model for medical research sets the general guidelines for the exploration of causation. Although stage 1 can be approached through well defined statistical methods, stages 2 and 3 require the acquisition and organization of knowledge about the world in which the causes and the effects reside.

When investigating causality a basic distinction can be made between structure and strength [Lagnado et al., 2004]. The former concerns the qualitative causal relations/links that hold between variables e.g., driver inattention causes crashes, lower speed limits prevent crashes etc. The latter is concerned with the quantitative aspect of these relations e.g., to what degree does inattention cause crashes, or low speed prevent them. Conceptually, the question of structure is more basic than that of strength i.e., one needs to know or assume the existence of a link before one can estimate its strength. Regardless, as already mentioned, in crash prevention studies researchers traditionally work the opposite way – they measure the strength of covariation between traffic and other measurements and crash events and based on this, implicitly establishing causal connections (structure). As a consequence it is impossible to distinguish between associations that link correlated factors (e.g., speed variance and crash) from underlying causal relations (flow breakdown and crash). More generally, these models are incapable of distinguishing between direct and indirect causal relations, or covariations that are generated by hidden causal factors [Waldmann, 1996].

The goal in this study was to establish both a causal structure as well as to measure the strength of the causal factors. According to Lagnado (2004) there are cues one can use in the identification of the causal structure, including:

- Statistical relations
- Temporal and spatial order
- Intervention
- Prior knowledge

Temporal and, less often, spatial order are normally assumed essential for defining causation, and are undoubtedly the most important cues that people use to distinguish causes from effects. Prior knowledge can be a powerful cue if used carefully. But it is often difficult to distinguish between a belief and a fact. Regardless, knowledge of the physical laws and theories governing the world of plausible causes and effects can play a pivotal role in identifying causal connections. For example, in this study, traffic flow theory outlining the known properties of traffic and its dynamics can be used to propose or explain possible connections. Useful as these cues may be, the statistical and philosophical literature has adamantly warned analysts that, unless one knows in advance all causally relevant factors, or unless one can carefully manipulate some variables, no certain causal inferences are possible (Cartwright, 1989, Fisher, 1953). However, recently Pearl (1991) has proposed a minimal-model theory of causation which provides a plausible account for how causal models could be inferred from observations. Pearl, and in essence this study, view the task of causal modeling as an identification game which scientists play against nature. The fundamental assumption is that nature possesses stable causal
mechanisms which, on a microscopic level are deterministic functional relationships between variables, some of which are unobservable. According to Pearl, “A causal model of a set of variables \( U \) is a directed acyclic graph (DAG), in which each node corresponds to a distinct element of \( U \).” The links between nodes denote direct causal influences among the variables. The causal model serves as a blueprint for forming a “causal theory” i.e., a precise specification of how each variable is influenced by its parents in the DAG. Formally, “A causal theory is a pair \( T = <D, P_D> \) consisting of a causal model \( D \) and a set of parameters \( P_D \) compatible with \( D \). \( P_D \) assigns a function \( x_i = f_i[pa(x_i), e_i] \) and a probability measure \( g_i \), to each \( x_i \) belonging to \( U \), where \( pa(x_i) \) are the parents of \( x_i \) in \( D \) and each \( e_i \) is a random disturbance distributed according to \( g_i \), independently of the other \( e \)’s and of any preceding variable \( x_j: 0 < j < i \).” The definition of causal theory is important to the methodology followed in this study.

**Methodology**

The methodology followed in this study is based on the general research model proposed by Thagard. First, reliable associations were established between crash events and traffic as well as crash events and environmental variables. Starting with this preliminary idea of what are the possible relationships between conditions on the road and crashes, the qualitative and quantitative knowledge acquired through all the analysis described in the previous chapters is used to establish preliminary causal connections. Through the use of statistical tests of conditional independence and following a step-by-step process, these causal connections are refined, accepted, or refuted. The goal is to propose a minimal causal model supported by the data, adequately describing the causal chain leading to crashes. Such a model can help engineers propose, design, and implement crash prevention/reduction interventions at the specific high crash area.

**Reliable Associations**

The qualitative and quantitative analysis aimed at empirically identifying the various traffic conditions encountered in the selected site as well as to quantify some of their characteristics. From these observations, it was possible to identify some apparent associations between traffic conditions and crash events. Unfortunately, such knowledge, albeit helpful, is not enough. Methodologies that can reliably detect statistical co-variation are needed.

There are numerous measurements and metrics describing different aspects of traffic flow in a particular location. In the previous chapter some of these measurements were so reliably associated with pre-crash traffic conditions that they could be utilized in early detection of such conditions. Since most of these metrics were constructed by aggregating or combining more fundamental traffic flow variables, their connection with the underlying causal process has been obscured. For example, knowing only that the speed variance measured at a point is reliably correlated with pre-crash conditions does not allow distinguishing between case (a) & case (b) in Figure 3.2. If it is possible, through further investigation, to identify the source of the increase in speed variance, for example a compression wave propagating backwards which demands weaving movements, or a particular driving behavior, then the original statistical covariation would have value in explaining causal structure and possibly later in measuring strength. Therefore, we are looking for measurements in different spots on the roadway that can describe
the prevailing traffic conditions. In particular, in a high crash area the investigation should start at the boundaries.

As was described earlier in this document, traffic is usually considered as a continuous medium into which patterns or disturbances propagate upstream or downstream depending on the conditions. Specifically, during uncongested conditions, perturbations in speed and density propagate in the same direction as the traffic (upstream to downstream). In contrast, during congested conditions, such perturbations propagate contrary to the direction of the traffic stream. These well-established relationships support a useful set of temporal and spatial relationships that one can look for in the collected measurements. Starting at the boundaries, an investigation of conditional independence can in a step-by-step process, narrow the investigation area and reduce the number of explanatory variables. Regardless, the associations between environmental conditions and crash likelihood established in the previous chapter are applicable to this investigation.

**Empirical Causal Connections**

The previous step provides a set of reliable associations between traffic/environmental measurements and crash events. The goal was for the measurements in the set to be conditionally independent when one controls for the crash or no-crash condition. Initially this goal is not always fully achieved, indicating possible interactions between variables or the existence of hidden causal factors. Regardless, this first set of associations between traffic and environmental variables combined with empirical knowledge of suspected causal factors allows for the development of a preliminary causal model. This directed but not necessarily acyclic, graph has for nodes the aforementioned variables, while a directed arrow between nodes indicates a causal relationship. Further statistical analysis of additional measurements as well as prior knowledge based on spatial/temporal order as well as traffic flow principles will guide the refinement of this preliminary structure and model. Ultimately, for each arrow in the graph there must either be prior reasons explaining the causal connection, or conditional independence tests using the study data. In the rest of this section the methods and cues utilized in this step of the methodology are presented.

**Analytical tools**

There are several software applications designed to assist in causal inference. Examples of such applications are TETRAD, developed at Carnegie Mellon, and RX developed at Stanford. Some
of these applications, like TETRAD, are for general use while others are developed for specific fields, such as RX, specifically designed for medical research. Considering the timeline of this work and the learn-as-we-go progress, more basic tools were utilized. The basic need at this stage of the research was to uncover correlations between traffic measurements and to conduct tests for conditional independence. The logistic regression utilized in the previous chapter combines the aforementioned needs and was used to create models of crash likelihood. Logit modeling and in general Generalized Linear Model (GLM) procedures are commonly used in transportation research and specifically in traffic safety. Logit models are driven by the covariation between the dichotomous dependent variable (crash or no-crash) and the various explanatory variables (traffic and environmental measurements). Additionally, part of the modeling process is the determination of each variable’s significance in determining the outcome. If a proper step-by-step modeling process is followed, cases of conditional independence can be identified among the explanatory variables. Finally, based on the estimate of the coefficient of each covariate, we can identify their relationship with the probability of the outcome. These relationships must agree with the independent findings of the qualitative and quantitative analysis as well as with traffic flow theory principles.

As mentioned earlier, temporal and spatial order as well as known traffic flow dynamics are useful in the investigation of crash causality in this high crash area. In this section, facts and observations are presented and analyzed, most of which have already been discussed during the qualitative and quantitative analysis but now are offered under a different point of view.

Prior Knowledge: Spatial independence

One of the driving considerations in this causality investigation is the observed correlations between measurements on different locations in the study area. Interpreting these correlations correctly is important. A road network is a unified system where humans through their choices and driving behaviors generate the traffic we finally measure. Because of these facts, all traffic measurements are correlated to each other. For example, barring small fluctuations and time shifts, all roadway segments exhibit the same large scale fluctuations in volume and to some extent speed and density. Society is on a schedule, many people leave their offices and head home all more or less at the same time. Especially, in an area so close to the start or end of work-related trips, all roadways will experience a synchronized rise and fall of demand levels. Due to these facts, any simple correlation analysis will tend to find all measurements highly correlated, thereby confusing the investigation.

In this work the aforementioned problem is reduced by increasing the data resolution, focusing more on short period correlations. But in order for correlation between two measurements, simultaneously taken in separate locations, to serve as evidence of causal connection, the two roadway subsystems must be physically linked. Consider two different examples; the first is concerned with measurements taken from two different detector stations on the same roadway, half a mile apart from each other, while the other is concerned with two other sensors one on the mainline of the freeway, upstream of an adjacent freeway-to-freeway entrance ramp, and the other at the upstream end of that ramp. In both cases, analysis reveals that measurements between the members of these two pairs are highly correlated. Assume that, in the first example, both locations of the roadway are under the same traffic state (un-congested or congested), while in the second example there is no congestion spilling backwards from the freeway proper to the ramp. In the first example the correlation can be accepted as indicating a causal connection since
what happens on the upstream location will affect the downstream and vice versa, depending on the state of traffic. In the second example, a correlation cannot be due to causal connection because the two roadway subsystems are physically isolated (under the assumption of uncongested traffic). In this case the prior knowledge of spatial independence can dismiss the connection as a result of spurious covariation. Additionally, if it is the case that both the conditions on the ramp as well as on the freeway proper are shown to be causally connected to the effect (crash in this case), spatial independence supports treating the two variables are exogenous.

**Prior Knowledge: Temporal order**

Extending the issues raised in the previous section, by assuming that roadway geometry and the prevailing traffic conditions are known, prior knowledge of temporal order can be the deciding factor in determining the direction of a causal connection, which is otherwise ambiguous with just correlation information. Additionally, temporal order will also guide the selection of variables as well as the interpretation of the results from conditional independence tests. Finally, later in the analysis, the temporal order of observed patterns in the traffic stream prior to crashes will assist in sorting through the suspected crash causal factors.

**Prior Knowledge: Traffic flow theory**

In the previous two sections the assumptions made stem mainly from traffic flow theory principles. The knowledge acquired through the observation of traffic is only useful if it is coupled with well founded traffic flow characteristics e.g., the direction of disturbance propagation in different traffic states, or the independence of subsystems in the presence of consecutive changes in traffic state (see the case in the beginning of chapter six). Later in the analysis, knowledge of traffic flow characteristics and behavior will be instrumental in deciding between otherwise ambiguous choices.

**Search for the minimal model**

As expected our initial model contains many possible causal connections. These proposed connections are then to be tested against all the aforementioned cues as well as additional measurements inside the influence area. This stage of “pruning,” as it is often called, aims for a minimal model [Pearl, 2000]. The need for a minimal model stems from the well accepted scientific principle of parsimony, often referred to as Occam's razor [Heylighen, 2000]. The principle admonishes us to choose from a set of otherwise equivalent models of a given phenomenon the simplest one. In any given model, Occam's razor helps in to ‘shave off” those concepts, variables, or constructs that are not really needed to explain the phenomenon. By doing that, developing the model will become much easier, and there is less chance of introducing inconsistencies, ambiguities, and redundancies. Though the principle may seem rather trivial, it is essential for model building because of what is known as the “under-determination of theories by data”. For a given set of observations or data, there are always an infinite number of possible models explaining those same data. This is because a model normally represents an infinite number of possible cases, of which the observed cases are only a finite subset. The non-observed cases are inferred by postulating general rules covering both actual and potential observations.
Crash Causal Mechanism

Ideally, by following the steps described in the previous section, a supported causal mechanism will be presented explaining how, why, and under what conditions a crash takes place in this particular location. Reaching this ideal state is very difficult and requires a guarantee that all relevant factors are known and their interactions are all defined. Such a mechanism can be presented graphically in a DAG or through the structural equations described in Pearl’s definition of causal theory. Through this mechanism one can simulate interventions by finding answers to counterfactual statements of the form “if through an intervention, causal factor A is changed to value \( x_2 \) instead \( x_1 \), while all other variables remain the same, will the crash be prevented?” In this work, we tried to come as close as possible to this ideal outcome. Whether or not the goal is fully reached, this investigation methodology organizes the search for causes, provides support by actual measurements for most of the empirical causal connections, and should identify the causal relationships in enough detail to successfully guide intervention designs. The rest of this chapter presents the implementation of this investigation methodology on the I-94 high crash area.

Causality Investigation

The causality investigation and its results presented in the rest of this chapter are the joined product of two knowledge seeking approaches. One approach capitalizes on the qualitative and quantitative analysis presented in previous chapters while the other is based on several additional logistic regression models. Regarding the latter, in contrast to the previous chapter where the goal was to find the best model of crash likelihood for integration with a detection algorithm, in this chapter the information gathered during the fitting process is utilized in the evolution of knowledge from reliable associations to causal relationships. The investigation started with the search of a preliminary causal structure based on the empirical information collected. This information is quantitative (traffic measurements) as well as qualitative (crash video observations). The measurements used in this stage of the research are presented next.

Study Area and Measurements

Most of the measurements used in the following sections were already discussed in previous chapters. Nevertheless, for better comprehension of the process the measurements utilized in identifying the crash causal factors and mechanism are presented in this section. Figure 3.3 identifies all the loop detector locations in the study section. In each of these locations volume and time occupancy measurements per lane are available every 30 seconds. Additionally, individual vehicle speeds and headways are available at the post-Portland station (800 feet (243m) downstream of Portland Avenue) as well as at the most common crash location (100 feet (30.48m) downstream of Portland Avenue, marked as “crash site” in the figure.
Based on the results of the qualitative and quantitative analysis, the study section boundaries were set as follows:

- **Upstream boundary**
  - I-94 upstream station
  - I-35W combination ramp
- **Downstream boundary**
  - Tunnel entrance
  - Hennepin Exit

Assuming that the findings of the previous analysis were not considered and one selected boundaries further upstream and downstream, a simple test of correlation would probably have shown that measurements in further away stations exhibit decreasing correlation to conditions at the crash site. Also, when measurements at the aforementioned boundaries are introduced, the further away ones are conditionally independent of the likelihood of a crash. As will be illustrated later in this chapter, the study area shrunk even further as more details regarding causal relationships were uncovered.

For modeling purposes, cases (crash events) and controls (no-crash events) were identified. During the study period 173 crashes were recorded (51 collisions and 122 near-misses). Considering that we are interested in crash prone conditions and based on the findings regarding the speed and frequency of the compression waves encountered in the area, four consecutive 30 second intervals prior to the time of the crash were considered crash prone, bringing the number of cases considered to 692. Measurements collected between 20 and 2 minutes prior to a crash are discarded to reduce the danger of misrepresenting the prevailing traffic conditions (could be crash prone but it is not guaranteed). All remaining time periods between 7 a.m. and 8 p.m. during the days of the study are considered crash-free (approximately 22,308). Controls were selected randomly from the total number of non-crash periods. Various alternatives for the selection ratio were explored ranging from 1:2 to 1:32 (essentially all available non-crash periods). The results in regards to causal connection cues were identical in all alternative ratios.
For brevity, results from the 1:32 ratio are presented in this chapter. For the purposes of logit modeling the dependent variable for cases is equal to 1 and for controls is equal to 0.

**Suspected Causal Factors**

From the recorded video of crash cases as well as that of non-crash periods, several observations were made concerning suspected crash causes. A thorough account of these observations can be found in chapter five. In the following sections these observations will be revisited to facilitate the search of specific causal factors.

**Flow breakdown**

The easiest and most persistent observation made was that almost all crashes, 98% to be exact, were preceded by one or more compression waves traveling upstream. The empirical analysis in chapter six attempted to use observations on the measurements, coupled with traffic flow characteristics, to narrow down the possibilities for the location of the flow breakdown that generated the compression waves. Strong evidence was found that the bottleneck was located between the I-94 flyover entrance ramp and the 2nd four lane station (Figure 3.3). Regardless, it was not possible to make connections between actual crash cases, the compression wave that created them, and the traffic conditions that created the wave in the first place. If the assumption that compression waves are one of the crash causal factors is correct, a crash prevention intervention would need to target the compression wave causal factors. Therefore, one path the causality investigation must follow is the one that locates evidence supporting the entire chain of events.

**Dangerous driving**

Another observation made from the analysis of the collected data is that during certain periods, suspiciously coinciding with pre-crash ones, vehicles in the right lane are dangerously driven i.e., have very short headways. Figure 3.4 presents a typical volume/occupancy scatter plot from 30 second data on the post-Portland right lane detector.
Figure 3.4 V/O graph, right lane of section downstream of Portland Ave. August 11th, 2005.

The figure represents typical traffic encountered in this location. As can be readily seen, flow on this lane reaches values as high as 2700 veh/h/lane. A very simple calculation reveals that, during those periods the average time headway between vehicles is as low as 1.3sec. This is the time-distance between the front bumpers of two successive vehicles, so the actual following time-distance can be as low as 0.75sec depending on vehicle length. In a study comparing driver reaction times in a simulator and a test track [McGehee et. al., 2000], total brake reaction time (defined as the period between the point at which the driver began to release the accelerator pedal up to the maximum brake application point) was 2.2 seconds for the Iowa Driving Simulator (IDS) and 2.3 seconds on the test track. Time to initial steering (defined as the point at which the driver first began to use steering to avoid the crash) was 1.64 seconds on the IDS and 1.67 seconds on the test track. Time to throttle release was also compared between the two studies. Mean time from incursion start to throttle release was 0.96 seconds for the IDS study and 1.28 seconds for the test track study. It is clear that, for the short periods of 10 to 20 minutes where the 2400-2700 veh/h/lane flow rates prevail in the study site, vehicles are driving very close to each other. The reason these periods do not last longer is because this flow is highly unstable and small disturbances lead to breakdowns (see metastable flow state definition in chapter six). The challenge in this investigation is to find measurement-supported evidence that this behavior is associated with crash events, and so possibly explains the causal connection with crashes.

**Driver inattention**

According to the Minnesota Motor Vehicle Crash Facts report of 2002, in multi-vehicle crashes “driver inattention” is the primary factor followed by “following too closely (tailgating)” and “speeding.” When single and multi-vehicle crashes are combined “driver inattention” is the primary factor for less severe crashes (the majority). In chapter five, observations showed that the maneuvers drivers use to negotiate upstream of Portland Avenue can draw their attention away from forward conditions long enough to cause a crash. In fact, Figure 3.5 illustrates that the
large number of vehicles moving from the first auxiliary lane, to the right mainline lane, ending in the middle mainline lane can reach an equivalent flow of more than 1500 veh/h. This flow is comparable to the flow of a high demand one-lane entrance ramp. All these vehicles need or wish to perform these two successive lane changes (auxiliary to right and right to middle) in a distance less than 2500 feet (762m) while, as the second graph in Figure 3.5 shows, the speed difference between the right and middle lanes can be as high as 20mph (32.18 km/h).

In determining the theoretically minimum ideal gap for merging, gap acceptance theory, suggests that this gap is the sum of three time intervals: (1) a safe time headway between the merging vehicle ahead in the destination lane $T_r$, (2) the time lost accelerating during the merging maneuver $T_L$, and (3) a safe time headway between the second vehicle in the destination lane and the merging vehicle $T_f$ [Drew, 1968]. The bigger the speed difference between lanes, the larger the time lost $T_L$ must be, increasing the size of the minimum ideal gap. It follows that all gap acceptance models agree that the probability of selecting a gap varies with gap size. Therefore, in the case of the lane changes performed in the study area, the larger the speed difference the bigger the minimum gap required and the longer a driver will have to look for it. Logically, the time spent looking in the side or in the rear view mirror is time taken away from looking ahead and time added to the reaction time for handling attitude changes of the leading vehicle in the lane. As final evidence that speed difference between the right and middle lanes is suspected for good reason to be a crash causal factor, Figure 3.6 presents the ratio of speeds in the right and middle lane during pre-crash periods. It is clear that the trend is very consistent. In the course of the causality investigation, the relationship of other upstream traffic conditions, both with the resulting speed difference as well as crashes will be explored. The causal structure connecting these variables is important in determining solutions.
Figure 3.5 (a) Flow in vph moving from the auxiliary lane to the middle lane. (b) Right lane speed and right–middle lane speed difference.
Figure 3.6 Comparison between speeds on the right and middle lanes 1 minute prior to crash.

**Preliminary Causal Structure**

Based on the aforementioned suspected causal factors, a rough causal structure is presented in Figure 3.7. In this graph the three suspected causal factors discussed in the previous section are included as well as possible unknown at the moment factors. Considering that very little information has so far been acquired concerning causal connections, all possibilities are included. The plan is to then keep revising this structure based on the directions the data indicate until a minimal model is achieved that provides enough information for engineers designing crash prevention interventions.
Figure 3.7 Rough causal structure based on prior knowledge.

Figure 3.7 is a conceptual description of the preliminary causal model where the causal factors are in this case represented by true/false variables. The goal of the investigation is to associate the crash effect with variables representing measurable traffic conditions related with these causal factors. In the following section a subsystem investigation of the flow breakdown factor is presented.

Modeling the Flow Breakdown

Before proceeding in the analysis of crash causal factors, it is more efficient to explore the causal factors of the flow breakdown. The steps or methods followed in the investigation of the flow breakdown causal factors are the same as in the main crash causal factor analysis. This subsystem implementation of the method can serve as an example of the larger analysis. Similar to the crash cases, a number of flow breakdown cases were collected with the help of the video records\(^1\), identifying the first 60 instances of compression waves. In addition, 250 samples of uncongested conditions were identified. To guarantee good results, prior to all cases and controls there were a minimum of 20 minutes of uncongested traffic. Based on traffic flow theory and engineering judgment, a preliminary causal structure for the flow breakdown can be identified and presented in the form of the flow chart in Figure 3.8. Some of the elements contributing to the development of the preliminary causal structure are as follows:

- At the instant of the breakdown, the upstream and downstream systems are totally separated. By definition, a compression wave is a discontinuity in the traffic stream.

---
\(^1\) Actually, this investigation was performed almost a year after the original crash information collection. The fact that there were enough crash cases to support the analysis allowed the re-aiming of the surveillance and detection equipment to cover the ramp merge point as well as approximately 1000 feet upstream, inside the ramp.
Additionally, for time intervals shorter than upstream to downstream travel time, the conditions just prior to the breakdown are also separate since the conditions in the upstream boundary will cause changes in the downstream only after a time period approximately equal to the travel time between the two locations.

- As explained in the example of spatial independence, the I-94 flyover ramp subsystem is separate and independent from both the upstream and downstream ones, only when there is no congestion, which explains the reasoning behind selecting only the very first instance of a compression wave.

- From the two previous statements as well as principles of traffic flow theory, it is plausible to assume that the upstream boundary conditions are responsible for the metastable flow state encountered at the breakpoint prior to the actual breakdown. An unknown detail is if the upstream conditions are responsible for the metastable state only through the effect they have on the right lane or if there is another parallel mechanism, currently unknown. The answer will be found later when statistical covariations are explored.

- According to the definition of metastable flow, a system is stable only because a disturbance of sufficient magnitude has not happened yet. There can be two sources for such a disturbance, an internal or an external source. The internal source is concerned with the actions of an individual driver, indicated in the graph labeled driver error. Although not uncommon, such cases are beyond the scope of this analysis. Additionally, such an abnormal behavior would have been noticed on the video records. The external (to the traffic stream) source assumes that the disturbance triggering the flow breakdown is generated either by the conditions propagating down the entrance ramp or due to lane changing maneuvers of traffic heading for the I-394 exit (right lane of the tunnel) and/or the Hennepin exit. The ambiguity, if such exists, will be resolved through tests of conditional independence.

The first step in the analysis is to determine the statistical covariation of each variable with the flow breakdown event. This is accomplished by fitting several Univariate logit models and checking the statistical significance of the associated parameters. The results of this analysis are presented in Table 3.1.
Figure 3.8 Flow chart of the preliminary flow breakdown causal structure.
Table 3.1 Estimation summary for logit model predicting flow breakdown.

<table>
<thead>
<tr>
<th>Univariable (single variable model)</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hennepin_Occupancy</td>
<td>0.0536623</td>
<td>0.153524</td>
<td>0.350</td>
<td>0.7267</td>
</tr>
<tr>
<td>Hennepin_Volume</td>
<td>0.00998233</td>
<td>0.0294986</td>
<td>0.338</td>
<td>0.7351</td>
</tr>
<tr>
<td>TunnelEntrance_AvgOfSpeed</td>
<td>-0.0271698</td>
<td>0.0233993</td>
<td>-1.161</td>
<td>0.2456</td>
</tr>
<tr>
<td>TunnelEntrance_MaxOfOccupancy</td>
<td>0.0115072</td>
<td>0.0129947</td>
<td>0.886</td>
<td>0.3759</td>
</tr>
<tr>
<td>TunnelEntrance_SumOfVolume</td>
<td>0.00985792</td>
<td>0.0166882</td>
<td>0.591</td>
<td>0.5547</td>
</tr>
<tr>
<td>TunnelEntrance_AvgOfOccupancy</td>
<td>0.0289842</td>
<td>0.0228051</td>
<td>1.271</td>
<td>0.2037</td>
</tr>
<tr>
<td>UpstreamEntrance_AvgOfOccupancy</td>
<td>0.0449391</td>
<td>0.0367392</td>
<td>1.223</td>
<td>0.2213</td>
</tr>
<tr>
<td>UpstreamRamp_AvgOfOccupancy</td>
<td>0.0348238</td>
<td>0.0250234</td>
<td>1.392</td>
<td>0.1640</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.305370</td>
<td>0.0374422</td>
<td>8.156</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Volume</td>
<td>0.267183</td>
<td>0.0448629</td>
<td>5.956</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Speed</td>
<td>-0.274114</td>
<td>0.0358101</td>
<td>-7.655</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Occupancy</td>
<td>0.158112</td>
<td>0.0287348</td>
<td>5.502</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Volume</td>
<td>0.0530350</td>
<td>0.0376174</td>
<td>1.410</td>
<td>0.1586</td>
</tr>
<tr>
<td>Portland_right_lane_Speed</td>
<td>-0.165136</td>
<td>0.0241484</td>
<td>-6.838</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_middle_lane_Speed</td>
<td>-0.0611394</td>
<td>0.0192900</td>
<td>-3.169</td>
<td>0.0015</td>
</tr>
<tr>
<td>Portland_middle_lane_Occupancy</td>
<td>0.0965356</td>
<td>0.0355002</td>
<td>2.719</td>
<td>0.0065</td>
</tr>
</tbody>
</table>

Grey = Low significance  
Bold = Significant variables

From the Univariate analysis results we observe that the Hennepin exit traffic conditions are not associated with the formation of the flow breakdown, while the tunnel entrance volume and maximum occupancy (right lane leading to I-394 exit) are also not relevant. Individually all other variables are statistically associated with the flow breakdown phenomenon. The next step involves a step-by-step process of fitting a logit model in which we observed the reaction of the included variables to the introduction of a new one. The analysis begins from the area boundaries and then proceeds by introducing variables closer to the bottleneck location.
According to the results of the statistical covariation test, presented in Table 3.2, there was no cause for ambiguity after all. The downstream boundary conditions (tunnel entrance) are not responsible for the disturbance triggering the breakdown while the traffic coming down the ramp is.

Based on this elimination order the following can be concluded:

- From the results of step 1 we can conclude that the flow breakdown phenomenon is independent of the tunnel traffic variables given the ramp occupancy and the upstream boundary traffic conditions. The original correlation between the breakdown event and the tunnel traffic was the result of spurious covariation due to the correlation between the traffic conditions on all points of the mainline. The stronger association between the

---

Table 3.2 Logit modeling of flow breakdown likelihood.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TunnelEntrance_AvgOfSpeed</td>
<td>-0.0316986</td>
<td>0.0502497</td>
<td>-0.631</td>
<td>0.5282</td>
</tr>
<tr>
<td>TunnelEntrance_AvgOfOccupancy</td>
<td>0.0114915</td>
<td>0.0502073</td>
<td>0.229</td>
<td>0.8190</td>
</tr>
<tr>
<td>UpstreamEntrance_AvgOfOccupancy</td>
<td>0.0646288</td>
<td>0.0490352</td>
<td>1.318</td>
<td>0.1875</td>
</tr>
<tr>
<td>UpstreamRamp_AvgOfOccupancy</td>
<td>0.0374710</td>
<td>0.0490352</td>
<td>0.778</td>
<td>0.4383</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.3245000</td>
<td>0.0400077</td>
<td>8.111</td>
<td>0.0000</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UpstreamEntrance_AvgOfOccupancy</td>
<td>0.0494616</td>
<td>0.0485930</td>
<td>1.018</td>
<td>0.3087</td>
</tr>
<tr>
<td>UpstreamRamp_AvgOfOccupancy</td>
<td>0.0457227</td>
<td>0.0308983</td>
<td>1.480</td>
<td>0.1389</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.261244</td>
<td>0.0415402</td>
<td>6.289</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Speed</td>
<td>-0.179816</td>
<td>0.0442885</td>
<td>-4.060</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UpstreamEntrance_AvgOfOccupancy</td>
<td>0.00699306</td>
<td>0.0555961</td>
<td>0.126</td>
<td>0.8999</td>
</tr>
<tr>
<td>UpstreamRamp_AvgOfOccupancy</td>
<td>0.0253491</td>
<td>0.0344881</td>
<td>0.735</td>
<td>0.4623</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.319735</td>
<td>0.0483632</td>
<td>6.611</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Speed</td>
<td>-0.128378</td>
<td>0.0494927</td>
<td>-2.594</td>
<td>0.0095</td>
</tr>
<tr>
<td>Portland_right_lan e_Occupancy</td>
<td>-0.135113</td>
<td>0.192880</td>
<td>-0.701</td>
<td>0.4836</td>
</tr>
<tr>
<td>Portland_right_lane_Volume</td>
<td>0.205668</td>
<td>0.221858</td>
<td>0.927</td>
<td>0.3539</td>
</tr>
<tr>
<td>Portland_right_lane_Speed</td>
<td>-0.253318</td>
<td>0.0974371</td>
<td>-2.600</td>
<td>0.0093</td>
</tr>
<tr>
<td><strong>Step 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.27554</td>
<td>5.59672</td>
<td>0.407</td>
<td>0.6843</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.322078</td>
<td>0.0480638</td>
<td>6.701</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Speed</td>
<td>-0.124952</td>
<td>0.0507339</td>
<td>-2.463</td>
<td>0.0138</td>
</tr>
<tr>
<td>Portland_middle_lane_Speed</td>
<td>-0.193332</td>
<td>0.0359463</td>
<td>-5.378</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_middle_lane_Occupancy</td>
<td>0.0460056</td>
<td>0.102856</td>
<td>0.447</td>
<td>0.6547</td>
</tr>
<tr>
<td>Portland_middle_lane_Occupancy</td>
<td>0.249328</td>
<td>0.369595</td>
<td>0.675</td>
<td>0.4999</td>
</tr>
<tr>
<td>Portland_MvsR_SpeedDifference</td>
<td>-0.185424</td>
<td>0.366472</td>
<td>-0.506</td>
<td>0.6129</td>
</tr>
</tbody>
</table>

Grey = Low significance
Bold = Significant variables
upstream conditions and the breakdown event disassociates the tunnel entrance conditions.

- In step 2 we observe that the upstream boundary conditions are independent of the likelihood of a flow breakdown on the condition of the inclusion of the right lane speed. In the same note, based on results of step 3, the right lane occupancy and volume are also independent on the condition of the right lane speed. Again these results indicate that although the upstream conditions help generate the traffic conditions in the right lane when we explicitly control for the right lane speed they become independent, suggesting a sequential causal relationship.

- Finally from step 4 results we conclude that the middle lane conditions are independent of the likelihood of a flow breakdown event on the condition of the inclusion of the right lane speed. This indicates that the flow breakdown event is not associated with any weaving activity between the right and middle lane.

Interestingly, based on the estimated signs the following observations can be made:

- The density of the stream coming down the ramp is positively related to the breakdown.
- The lower the speed of the platoons coming down the ramp, the higher the breakdown likelihood. This is an interesting result because, one would think that the more vehicles coming down the ramp fast-moving and closely-spaced, the bigger the disturbance. Instead, high density with low speeds (low volume) is actually worse. Following the analysis, a closer inspection of the video records revealed that as the speed of the vehicles coming down the ramp reduces, the sooner they try to merge on the right lane, causing bigger disturbances. In contrast, when they come down faster they stay in the auxiliary lane longer.
- Additionally, contrary to usual notions about metastable flow, the lower the speed on the right lane, the bigger the likelihood of a breakdown. An interesting future research topic would be to explore the relationship between the speeds of the two merging streams and the drivers’ gap selection and/or distance to merge.

Figure 3.9 presents the final flow chart of the causal structure as supported by the data while Figure 3.10 presents the final causal model of flow breakdown. In the model, for completeness, the driver error variable is included as a True/False one that overrides the ramp conditions.
Figure 3.9 Final flow chart of flow breakdown causal structure.

Figure 3.10 Final causal model of flow breakdown.
Location Specific Preliminary Crash Causal Structure

Exploring the causal factors of flow breakdown was greatly assisted by well defined traffic flow principles; the exploration of crash causal factors will be more complicated since there may be several sub-mechanisms unknown and unmeasured. Previous work by Davis and Swenson (2003) developed a causal model of rear-end crashes observed in the study area. That was a very detailed model based on equations of motion and individual driver actions. It identified the causal chain between driver actions and the crash event. The model in this work follows a top-down approach in an attempt to explore the general traffic dynamics setting the stage for the crashes.

Based on prior knowledge of spatial independence and temporal order, the preliminary graphical causal model can evolve to match more closely the specific high crash area. To better understand the individual concepts, a flow chart of the preliminary causal structure is also presented in Figure 3.11. The measurements will be utilized to justify statements about traffic conditions. The chart in Figure 3.11 is still a conceptual design based on engineering judgment and the prior analysis of video and measurements. Some of the relationships developed can be explained as follows starting from the upstream boundary.

- From the video and measurement records, it is known that prior to crashes the upstream end of I-94 is uncongested. In this case, spatial independence removes any possible causal connections between the upstream entrance and upstream mainline nodes. These two systems are independent.
- The merge of the two aforementioned traffic streams in the first four-lane section generates new traffic conditions. At this point we can hypothesize that one of the effects is the merging difficulty, eventually causing driver distraction. The second effect of this mix is the generation of the fast moving, dense platoons of vehicles responsible for the observed very short time headways.
- Another possible causal factor for the dangerous conditions in the right lane could stem from route selection of vehicles destined for the right lane of the tunnel (I-394 exit) and the Hennepin exit.
- Although the conditions at the I-94 flyover combination ramp have not yet been included in the logit model, we can safely assume that there are no causal connections between the ramp and the upstream conditions. Additionally, since the merge point of the ramp is downstream of the crash impact location, a causal connection can only exist during congested conditions. This explains the arrow to the flow breakdown factor. Only during uncongested conditions might there be some causal connection between the flyover ramp and the tunnel/Hennepin boundaries. Since we are only concerned with crashes and temporally they are preceded by the flow breakdown such connection is not included.
- The direct causal connections between the nodes representing the boundary conditions (dashed lines) were included to emphasize the fact that there might be other, unknown at the moment, causal relationships in addition to the three suspected causal factors.

The next step is to build a logit model with the dependent variable being either 0 for non-crash or 1 for crash, and traffic measurements at the boundaries as the independent variables. Following standard modeling practice the process begins with a careful Univariate analysis of each variable. The results are presented in Table 3.3. This table describes the statistical covariation between each available variable and the crash events. Based on these statistical covariations and the
preliminary structure in Figure 3.11 a preliminary causal model was developed and presented in Figure 3.12. In the causal model only the variables correlated with the crash events were considered. At this stage all causal connections are preliminary since it is not yet clear if they are direct, indirect, or associations due to spurious covariation.

Table 3.3 Univariate analysis of all available variables.

<table>
<thead>
<tr>
<th>Univariable (single variable model)</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hennepin_Occupancy</td>
<td>0.393698</td>
<td>0.0574899</td>
<td>6.848</td>
<td>0.0000</td>
</tr>
<tr>
<td>Hennepin_Volume</td>
<td>0.0750078</td>
<td>0.0110418</td>
<td>6.793</td>
<td>0.0000</td>
</tr>
<tr>
<td>TunnelEntrance_AvgOfSpeed</td>
<td>-0.0293110</td>
<td>0.00867803</td>
<td>-3.378</td>
<td>0.0007</td>
</tr>
<tr>
<td>TunnelEntrance_MaxOfOccupancy</td>
<td>-0.00479274</td>
<td>0.00743847</td>
<td>-0.644</td>
<td>0.5194</td>
</tr>
<tr>
<td>TunnelEntrance_MinOfSpeed</td>
<td>-0.00578124</td>
<td>0.00705245</td>
<td>-0.820</td>
<td>0.4124</td>
</tr>
<tr>
<td>TunnelEntrance_SumOfVolume</td>
<td>0.0560166</td>
<td>0.00676674</td>
<td>8.278</td>
<td>0.0000</td>
</tr>
<tr>
<td>UpstreamEntrance_AvgOfOccupancy</td>
<td>-0.00221027</td>
<td>0.00389029</td>
<td>-0.568</td>
<td>0.5699</td>
</tr>
<tr>
<td>UpstreamEntrance_SumOfVolume</td>
<td>0.0326384</td>
<td>0.00619636</td>
<td>5.267</td>
<td>0.0000</td>
</tr>
<tr>
<td>UpstreamRamp_AvgOfOccupancy</td>
<td>-0.00175303</td>
<td>0.00348552</td>
<td>-0.503</td>
<td>0.6150</td>
</tr>
<tr>
<td>UpstreamRamp_SumOfVolume</td>
<td>0.0483612</td>
<td>0.00890565</td>
<td>5.430</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.0397192</td>
<td>0.00250366</td>
<td>15.864</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Volume</td>
<td>0.0980891</td>
<td>0.0139810</td>
<td>7.016</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Occupancy</td>
<td>0.0391135</td>
<td>0.00247511</td>
<td>15.803</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Volume</td>
<td>-0.0658068</td>
<td>0.0127622</td>
<td>-5.156</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Speed</td>
<td>-0.0281325</td>
<td>0.00269524</td>
<td>-10.438</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_AveHeadway</td>
<td>-0.000511096</td>
<td>0.0000717524</td>
<td>-7.123</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_VehLength</td>
<td>0.00871386</td>
<td>0.00546234</td>
<td>1.595</td>
<td>0.1107</td>
</tr>
<tr>
<td>Portland_middle_lane_Speed</td>
<td>-0.00933157</td>
<td>0.00307382</td>
<td>-3.036</td>
<td>0.0024</td>
</tr>
<tr>
<td>Portland_middle_lane_AveHeadway</td>
<td>-0.000535187</td>
<td>0.0000673640</td>
<td>-7.945</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_MvsR_SpeedDifference</td>
<td>0.247450</td>
<td>0.00936415</td>
<td>26.425</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Grey = Low significance
Bold = Significant variables
Figure 3.11 Location specific flow chart of preliminary causal structure based on prior knowledge.
Figure 3.12 Location specific preliminary causal model based on statistical covariation and prior knowledge.
Since the boundary locations have at least three lanes, additional measurements describing traffic conditions over all lanes are included. The table presents the first evidence of statistical covariation between traffic measurements and crashes. So far all boundary conditions are associated with the crash. Preliminary observations from Table 3.3 note that:

- In generating conditions favorable to crashes, an important factor is the upstream boundary volume, while the occupancy (density) in this case does not seem to play a vital role. This result is not unreasonable considering the suspected mechanism. It is the number of vehicles merging in the right and middle lanes at the beginning of the first four-lane section that is critical. If the volume and density both were high at that point, a bottleneck would be formed reducing right lane congestion at the end of the four-lane section. High density, low volume signifies congestion which was also contrary to the video observations of the upstream boundary.

- On the downstream boundary only the variables describing average traffic levels are important, prompting the suspicion that it might be product of spurious covariation. Further investigation is required.

- Based on the results of the flow breakdown analysis it is not surprising that the I-94 flyover ramp conditions are associated with the likelihood of a crash.

**Boundary Conditions**

Selecting only the significant variables of the boundary conditions a multivariable model was fitted. Table 3.4 presents the results from the boundary multivariable logit model.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-6.30740</td>
<td>0.580670</td>
<td>-10.862</td>
<td>0.0000</td>
</tr>
<tr>
<td>Hennepin_Occupancy</td>
<td>1.22449</td>
<td>1.00199</td>
<td>1.222</td>
<td>0.2217</td>
</tr>
<tr>
<td>Hennepin_SumOfVolume</td>
<td>-0.148757</td>
<td>0.192501</td>
<td>-0.773</td>
<td>0.4397</td>
</tr>
<tr>
<td>TunnelEntrance_AvgOfSpeed</td>
<td>-0.0422661</td>
<td>0.0104311</td>
<td>-4.052</td>
<td>0.0001</td>
</tr>
<tr>
<td>TunnelEntrance_SumOfVolume</td>
<td>0.0512798</td>
<td>0.00691488</td>
<td>7.416</td>
<td>0.0000</td>
</tr>
<tr>
<td>UpstreamEntrance_SumOfVolume</td>
<td>0.0233057</td>
<td>0.00642443</td>
<td>3.628</td>
<td>0.0003</td>
</tr>
<tr>
<td>UpstreamRamp_SumOfVolume</td>
<td>0.0408293</td>
<td>0.00912679</td>
<td>4.474</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

The model was developed following a step-wise process. The significant variables could have been selected during the step-wise process based on change in deviance or as in the table above, based on the resulting p-value compared with level of significance $\alpha=0.05$. From the results it is worth noticing that the variables describing the conditions at the Hennepin exit are highly correlated to each other and the introduction of one greatly changes the estimate and significance of the other. For the subsequent steps we opt to include only the one describing the exit volume. Based on these results no changes are warranted to the preliminary causal model of Figure 3.12. The procedure from this point on is to introduce variables from locations closer to the crash point and see if any of the earlier ones are dropped. If the significance of a variable is greatly reduced

---

1 Actually experiments were conducted also with the Occupancy variable. The choice of the one over the other had no effect on the analysis results.
when a new one is introduced it suggests that the old variable is now conditionally independent to the effect given the inclusion of the new factor. The process is described in steps.

“I-94 flyover” Ramp Conditions

In the following steps variables describing the conditions on the I-94 flyover ramp were introduced and the effect they have on the rest was explored.

Table 3.5 Logit model #2: addition of measurements at the I-94 flyover entrance ramp.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-9.68463</td>
<td>0.623549</td>
<td>-15.531</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.0510219</td>
<td>0.00325090</td>
<td>15.695</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Volume</td>
<td>0.103184</td>
<td>0.0156786</td>
<td>6.581</td>
<td>0.0000</td>
</tr>
<tr>
<td>Hennepin_SumOfVolume</td>
<td>0.100954</td>
<td>0.0117543</td>
<td>8.589</td>
<td>0.0000</td>
</tr>
<tr>
<td>TunnelEntrance_AvgOfSpeed</td>
<td>-0.00817032</td>
<td>0.0109522</td>
<td>-0.746</td>
<td>0.4557</td>
</tr>
<tr>
<td>TunnelEntrance_SumOfVolume</td>
<td>0.0191241</td>
<td>0.00735164</td>
<td>2.601</td>
<td>0.0093</td>
</tr>
<tr>
<td>UpstreamEntrance_SumOfVolume</td>
<td>0.0276969</td>
<td>0.00682587</td>
<td>4.058</td>
<td>0.0001</td>
</tr>
<tr>
<td>UpstreamRamp_SumOfVolume</td>
<td>0.0528203</td>
<td>0.00960616</td>
<td>5.499</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

The first noticeable change is the drop in significance of the tunnel speed variable. Although minimum confounding was observed on the rest of the variable estimates (their estimate values and significance is virtually unchanged), the introduction of the ramp variables reduced the importance of the tunnel speed. Although not a clear indication, this response reinforces the suspicion that the association of the tunnel conditions with crashes is due to spurious covariation. Logically, if the tunnel traffic was independent of the crash likelihood on the condition of the ramp variables, both tunnel variables should have been dropped.

Post-Portland Right Lane Conditions

Special attention must be given in the introduction of the variables describing the conditions on the right lane immediately downstream of Portland Avenue. This location is 1000 feet (304m) upstream of the merge point of the entrance ramp and is the location where ~95% of the crashes take place. Additional variables describing in greater detail conditions in this location are introduced. Since these variables are bound to be highly correlated with each other special attention is given to confounding and interaction effects.
Table 3.6 Logit model #3: addition of measurements on the right lane downstream of Portland Ave.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-9.84924</td>
<td>0.693434</td>
<td>-14.204</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Occupancy</td>
<td>0.0795214</td>
<td>0.00555019</td>
<td>14.328</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Volume</td>
<td>0.0427855</td>
<td>0.0202158</td>
<td>2.116</td>
<td>0.0343</td>
</tr>
<tr>
<td>Portland_right_lane_Speed</td>
<td>0.0325038</td>
<td>0.00693469</td>
<td>4.687</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Headway</td>
<td>-0.00130639</td>
<td>0.0000983782</td>
<td>-13.279</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.0527998</td>
<td>0.00447147</td>
<td>11.808</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Volume</td>
<td>0.104503</td>
<td>0.0170140</td>
<td>6.142</td>
<td>0.0000</td>
</tr>
<tr>
<td>Hennepin_SumOfVolume</td>
<td>0.0836156</td>
<td>0.0123153</td>
<td>6.790</td>
<td>0.0000</td>
</tr>
<tr>
<td>TunnelEntrance_SumOfVolume</td>
<td>0.00924660</td>
<td>0.00798825</td>
<td>1.158</td>
<td>0.2471</td>
</tr>
<tr>
<td>UpstreamEntrance_SumOfVolume</td>
<td>0.0260559</td>
<td>0.00706183</td>
<td>3.690</td>
<td>0.0002</td>
</tr>
<tr>
<td>UpstreamRamp_SumOfVolume</td>
<td>0.0491263</td>
<td>0.0102426</td>
<td>4.796</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 3.6 presents the result from the introduction of the post-Portland right lane traffic conditions variables. The model in the table is the final fit. Actually to comprehend better the behavior of the variables each was also individually introduced to the model of the previous section. The following observations were made:

- With the introduction of the right lane occupancy the tunnel volume variable drops out. Confounding is noted between the right lane occupancy and the Hennepin exit variable, hinting at an interaction between these two variables. This does not happen with the individual introduction of any other variable.
- The volume variable produces the smallest improvement to the deviance.
- The speed variable when introduced alone was not significant and the final estimate had a positive sign. The same variable in the flow breakdown model had a negative sign. This generates the suspicion that the speed in the right lane might be the result of two separate causal factors.
- The removal of the last tunnel variable minimizes the possibility of an association with crash causes. This is consistent with the flow breakdown analysis, which also displayed no association with the tunnel traffic on the condition of the right lane flow variables. It also suggests that the majority of the lane changes positioning vehicles for the I-394 exit have been accomplished upstream of the measurement point.

**Post-Portland Middle Lane Conditions**

Considering the observations made regarding the merge difficulty, and the influence it might have on driver distraction, the effect of the middle lane traffic was analyzed although crashes on that lane were very rare. The process was conducted in the same fashion as in the earlier steps.
Table 3.7 Logit model #4: addition of measurements on the middle lane downstream of Portland Ave.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-14.4170</td>
<td>0.780680</td>
<td>-18.467</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_middle_lane_Speed</td>
<td>0.428721</td>
<td>0.0183226</td>
<td>23.399</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_middle_lane_Headway</td>
<td>-0.00374991</td>
<td>0.0000838603</td>
<td>-4.472</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Occupancy</td>
<td>0.128520</td>
<td>0.00718370</td>
<td>17.890</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Volume</td>
<td>0.0235551</td>
<td>0.0230772</td>
<td>1.021</td>
<td>0.3074</td>
</tr>
<tr>
<td>Portland_right_lane_Speed</td>
<td>-0.276883</td>
<td>0.0147072</td>
<td>-18.826</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Headway</td>
<td>-0.00170944</td>
<td>0.000118363</td>
<td>-14.442</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.0781365</td>
<td>0.00525221</td>
<td>14.877</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Volume</td>
<td>0.0711244</td>
<td>0.0193425</td>
<td>3.677</td>
<td>0.0002</td>
</tr>
<tr>
<td>Hennepin_SumOfVolume</td>
<td>0.0597246</td>
<td>0.0144270</td>
<td>4.140</td>
<td>0.0000</td>
</tr>
<tr>
<td>UpstreamEntrance_SumOfVolume</td>
<td>0.0295274</td>
<td>0.00813814</td>
<td>3.628</td>
<td>0.0003</td>
</tr>
<tr>
<td>UpstreamRamp_SumOfVolume</td>
<td>0.0307633</td>
<td>0.0118585</td>
<td>2.594</td>
<td>0.0095</td>
</tr>
</tbody>
</table>

From the step-by-step introduction of the new variables we observed:

- With the introduction of the speed on the middle lane the right lane volume drops out. When only the headway is introduced the significance of the volume is reduced but remains still in the acceptable region. This result does not contradict the proposed flow breakdown casual model since speed, not volume, was the main element.

- A strong confounding effect was observed between the middle lane speed and the Hennepin volume. Very strong confounding was observed between the right and middle lanes speed variables, to the point of changing the sign of the estimate for the right lane. Such an effect prompts further analysis. The introduction of an interaction term between the two variables is warranted. Actually, according to the preliminary causal model, the difference in speeds should be also important.

In the next step the speed difference between the right and middle lanes will be introduced as well as an interaction term.

**Combined Effects**

Table 3.8 presents the results from the introduction of the speed difference and the interaction term.
## Table 3.8 Logit model #4: addition of measurements on the middle lane downstream of Portland Ave.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Est/SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-65.2753</td>
<td>3.23560</td>
<td>-20.174</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_MvsR_SpdDifference</td>
<td>1.96277</td>
<td>0.0992345</td>
<td>19.779</td>
<td>0.0000</td>
</tr>
<tr>
<td>Right X Middle Interaction</td>
<td>-0.0466831</td>
<td>0.00236934</td>
<td>-19.703</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_middle_lane_Speed</td>
<td>3.36069</td>
<td>0.170221</td>
<td>19.743</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_middle_lane_Headway</td>
<td>0.000012157</td>
<td>0.000122049</td>
<td>0.100</td>
<td>0.9207</td>
</tr>
<tr>
<td>Portland_right_lane_Occupancy</td>
<td>0.285724</td>
<td>0.0181119</td>
<td>15.776</td>
<td>0.0000</td>
</tr>
<tr>
<td>Portland_right_lane_Speed</td>
<td>Aliased</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portland_right_lane_Headway</td>
<td>-0.00177521</td>
<td>0.000179204</td>
<td>-9.906</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Occupancy</td>
<td>0.164504</td>
<td>0.0121112</td>
<td>13.583</td>
<td>0.0000</td>
</tr>
<tr>
<td>RI94_flyover_Volume</td>
<td>-0.0133278</td>
<td>0.0345410</td>
<td>-0.386</td>
<td>0.6996</td>
</tr>
<tr>
<td>Hennepin_SumOfVolume</td>
<td>-0.0259219</td>
<td>0.0270642</td>
<td>-0.958</td>
<td>0.3382</td>
</tr>
<tr>
<td>UpstreamEntrance_SumOfVolume</td>
<td>0.0108587</td>
<td>0.0143796</td>
<td>0.755</td>
<td>0.4502</td>
</tr>
<tr>
<td>UpstreamRamp_SumOfVolume</td>
<td>-0.0331361</td>
<td>0.0210695</td>
<td>-1.573</td>
<td>0.1158</td>
</tr>
</tbody>
</table>

The introduction of the interaction term resulted in major changes in the model structure. A step-wise analysis based on deviance showed that the key term is the middle lane speed. The effect of the interaction, or effect modification as it is called in epidemiology, suggests that the effect of middle lane speed on crash likelihood depends on the speed of the right lane. Specifically, from the plots of the model response Figure 3.13 we can see that the critical value for the speed on the middle lane is 37 mph (59.54 km/h), on the right lane 36 mph (58 km/h), and for speed difference greater than 10 mph (16.1 km/h).

- What the graphs suggest is that the likelihood of a crash is reduced as the speed in the middle lane gets either higher or smaller than 37 mph (59.54 km/h). If the speed of the middle lane is much lower than 37 mph (59.54 km/h) then lane changing would be easy, demanding little attention from the drivers who are free to look ahead and also complete their maneuver before the Portland overpass. On the other hand, if speed on the middle lane is much higher than 37 mph (59.54 km/h) lane changing is so difficult that most of the drivers opt to stay in the right lane.

- When the speed on the right lane is lower than 36 mph the crash likelihood is reduced since attitude changes can be handled easier. When the right lane speed is much higher than 36 mph then drivers are paying more attention on the task and as suggested from the flow breakdown model the likelihood of a compression wave is reduced.

- The likelihood of a crash increases more sharply when the difference between lanes is higher than 10 mph (16.1 km/h).

The above findings are at the moment based on the response of the crash likelihood model. Further research of lane changes and gap-acceptance in the study area might reveal additional information.
The I-94 flyover ramp volume dropped out with the introduction of the new terms. This is in agreement with the flow breakdown model which did not consider the volume levels on the ramp as critical to the flow breakdown at the merge point. Additionally, the average headway on the middle lane also dropped down suggesting that the speed difference and level are the only relevant causal factors.

**Graphical Causal Model Refinement**

Based on the results from the step-wise logistic regression analysis the preliminary causal model can be refined to better reflect the data suggestions. Figure 3.14 shows the new crash structure flow chart. Figure 3.15 shows the final causal model as it is supported by the data, the observations, and traffic flow theory principles. Both in the flow chart and the model there are some elements still unknown. According to the previous chapter analysis the position of the sun and other environmental conditions also affect the likelihood of a crash. The exact causal mechanism of this interaction is yet to be discovered. Additionally, the human factors related to the drivers’ speed selection and gap acceptance also warrants further analysis. For completeness these elements were also included in the figures but with dashed lines since their causal chain is not yet clear.

**Conclusions**

In this chapter an attempt was made to uncover the causal factors responsible for the crashes in the I-94 high crash section. Effort was made to follow an organized sequence of steps in the hope of both reducing the risk of missing vital clues and for increasing the possibility of replicating this process in other high crash areas. The ultimate goal of this study was to identify the crash causal mechanism. Similar to other efforts in causality investigations, this has not been fully accomplished. Regardless, we believe that the uncovered causal relationships can greatly assist engineers in search of crash prevention interventions. For example, the Hennepin exit conditions persevered throughout the analysis and only became independent of the crash likelihood when we conditioned on the speed difference/level. In contrast, the tunnel conditions dropped out early on, suggesting that the traffic patterns are not relevant to the crash risk, discharging currently popular but costly road geometry modifications.
Figure 3.13 Critical values of speeds: middle, right, and difference.
Figure 3.14 Flow chart of crash causal structure as supported by the analysis results.
Figure 3.15 Crash causal model as supported by the data.
4 Immersive 4-D Visual Simulation of Driver Warning Systems

The original motivation behind developing the immersive real-time 4D visualization of actual driver crash scenarios was to allow practitioners and stake-holders the ability to observe the effectiveness of several proposed candidate crash mitigation warning systems as what would be ‘experienced’ by the driver (road side or ‘in-vehicle’). Second, the observations can be made within the scenarios from different vehicles and different in-vehicle vantage points. The Deployment issues and insight to the efficacy of a proposed driver warning systems would be uncovered early on in the design to deployment pipe-line, thus saving time and reducing costly design cycles to achieve optimal designs. The intent of the utilizing the system is not to replace a driving simulator. Rather, the immersive visualization process fits at the beginning of the design pipe-line, followed by subsequent testing using driving simulation for human factors evaluation for the most promising candidate systems, and eventual deployment in the field.

Introduction

One assumption at the onset of the project, is that the spatial perception of what is observed in the Digital immersive ENvironment (DEN) is similar to the real world. There is a body of literature that has examined the validity of spatial perception within virtual environments produced produced by large screen display systems similar to the DEN. For example, Plumert, Kearney, and Cremer (2004) compared subjects perceived action-based distance estimates (they imagined walking speed travel times) to landmarks placed 20 to 120 feet away. Their data showed very good correspondence between the real and virtual environment. Note that their immersive system, although very similar configuration used herein, did not utilize stereo perception in their experiments. Recently Kenyon et. al (2007) found scale and size consistency in a CAVE environment very closely resembled results to similar perceptions done in real-world experiments. Interestingly, subjects did not appear to adopt motion parallax strategies – even though their virtual environment would allow it—to ascertain object size; thus the dominant spatial cues were scale and stereopsis. The aforementioned studies do did consider the effects of self-motion—as is the case for virtual driver’s perspective observation. Harris et. al (1998) found that adding coupled vestibular cues to rendered optical flow within a virtual environment actually resulted subject’s perceived distance to be less accurate than by just utilizing vision-based self-motion alone. Only linear forward moving accelerations were considered in their experiments, however. Bicycle safety has been evaluated by observing perceived crossing gap acceptances in a virtual environment representation of urban intersections using an immersive large-screen display system very similar to the DEN (Kerney, et al., 2006, Plumert, et al., 2004). The primary aim of their studies were to explore differences crossing behavior between children and adults for a myriad of known, a-priori traffic condition exposures generated within the virtual environment.

The sections are organized first, by explaining the scene environment modeling methodology, vehicle trajectory integration, the limitations of the current development state of the aforementioned visualization system, and finally a detailed description of the system configuration.
**DEN 3D Environment and Object Geometry Modeling Process**

The 3D environment model consists of 41,600 polygons utilizing over 100 texture images. The working format of the 3D environment data base is OpenFlight, which is the same format utilized by the HumanFIRST lab at the University of Minnesota. The 3D modeling process consists of (1) building a high fidelity 3D model of the roadway and surrounding environment, (2) registering the microsimulation software roadway network with the 3D roadway model, (3) Implementing a library of 3D vehicles optimized for 3D simulation (with accurate interior geometry and ‘out-of-windsheild’ representations), and (4) building a 3D road surface elevation look-up table.

In order to create the 3D model suitable for real-time aforementioned immersive roadway visual simulation, the MTO laboratory forged a partnership agreement with SimWright Inc. specifically to utilize two of their commercial photogrametric feature extraction software packages -- SpiDAR and StereoGIS. Others have reported results from road alignment studies that indicate absolute accuracies within +/-3cm along all three axes for flight altitudes of 2,000 ft. Above Ground Level (AGL), and +/-8 cm at for flight altitudes of 12,000 feet AGL. The data were verified by FDOT surveyors office, (Jones et al., 2003). Both software tools required stereo-pair aerial flight data (aerial photos with at least sixty percent overlap) of the region of interest and the associated Auto Triangulation bundle (AT) calibration data. The flight data containing the imagery of the freeway and surrounding area were from a two flight missions performed on April 20, 2002 and May 3, 2002 by Horizons Inc, for the city of Minneapolis and Hennepin County. The flight altitudes were approximately 5000 feet. Each resulting aerial image is approximately a 14K x 14K diapositive image with a resolution of approximately 0.46 feet per pixel (14 microns in the diapositive). The accuracy of the SimWright methodology reported in the literature is highly dependent on the absolute AT accuracy (accuracy being defined as the difference between measured points on the terrain surface and the same points reproduced by image triangulation between two camera aerial image views of the same point). The AT bundle adjustment for the flight data used for this study produced an accuracy no greater than .434 feet RMSE over all twelve control points. Note that the AT production given to the University of MN by Horizons required a documented data release agreement by the City of Minneapolis.

![Figure 4.1 Stereo pair coverage tiles covering the Interstate 35W and Interstate 94 commons freeway section.](image)

The high crash rate area is within MnnW_3. Source: Google Maps.
a) general epipolar image pairs from two camera positions

b) two overlapping rectified aerial images along flight path

Figure 4.2 Epipolar geometry, the base line, CC’, and CX, C’X, all lie in the same plane (shaded). For the flight path data, overlapping aerial image pairs are reprojected as in (b) onto common plane parallel to a line between optical centers (in direction of flight path).

The images and AT report were then sent to SimWright for further image processing in order to create georectified epipolar image pairs. In essence, any image pixel which is visible in both images can be intersected along the same directional epipolar line (fig 2). In this case the direction of the epipolar line follows (but is not equivalent) to the flight path nadir of the aerial imagery. The processed images can then be used for the feature extraction algorithms available in the software using stereographic techniques (the lab purchased a low-cost 3D OpenGL compatible stereo-capable graphics card for this purpose). All features are expressed in Minnesota South NAD83 State Plane coordinates (the SpiDAR plugin allows the user to specify a different model origin). The sight for this study required four image pairs to be processed (Figure 4.1). Since none of the twelve surveyed ground control points utilized for the flight mission were visible from the image pairs purchased by the lab, the final absolute XYZ accuracy analysis by SimWright was not available. However, relative accuracy and consistency checks were performed by observing the difference between measured 3D points along image pair boundaries. Table 4.1 summarizes results of this test performed by SimWright. The results from...
Table 4.1 indicate all but one elevation check was within an absolute value of 6 inches. The source of error for the larger elevation discrepancy between stereo pair image sets MinnW_3 and MinnW_1 is unknown.

Table 4.1 Point elevation errors along image pair boundaries.

<table>
<thead>
<tr>
<th>Images</th>
<th>1st image Z value</th>
<th>2nd image Z value</th>
<th>Difference in Z values</th>
</tr>
</thead>
<tbody>
<tr>
<td>MinW_1 &amp; MinW_2</td>
<td>256.4141128</td>
<td>256.563945</td>
<td>-0.15</td>
</tr>
<tr>
<td>MinW_1 &amp; MinW_3</td>
<td>256.6710046</td>
<td>257.241047</td>
<td>-0.57</td>
</tr>
<tr>
<td>MinW_4 &amp; MinW_2</td>
<td>259.3335577</td>
<td>259.3486711</td>
<td>-0.02</td>
</tr>
<tr>
<td>MinW_4 &amp; MinW_3</td>
<td>259.3335577</td>
<td>259.2672566</td>
<td>0.07</td>
</tr>
<tr>
<td>MinE_1 &amp; MinE_2</td>
<td>272.572127</td>
<td>272.4960424</td>
<td>0.08</td>
</tr>
</tbody>
</table>

The next step requires the extraction of road edges. The first attempt involved automated boundary edge detection and extraction of the 3D edges using the StereoGIS software. Although the edges appeared to be extracted reasonably well, we were unsuccessful at importing the edges into MultiGEN Creator version 3.0. Instead, SpiDAR was used to extract arbitrary points along the road edges and then b-spline curves passed through the points to create the final road and overpass edges (Figure 4.3). The edges formed the path basis for extruding surface polygons representing the roadway.

![Figure 4.3 Extracted road and overpass edges of the Interstate 35W/94 commons from east of Cedar Avenue up to the lowery tunnel.](image)

SpiDAR was also used to embed the lane markings since lane and road geometries vary considerably along the road model. This was achieved by using the image texture projection tool within SpiDAR to capture 2^n sized textures from the aerial image pairs onto the roadway. The location and spacing of lane markings on the projected texture are then used directly to inlay lane strip polygons onto the road surface (Figure 4.4).
Finally, the 3D location and dimensions of surrounding road signage, lighting, buildings, and other infrastructure were digitized with SpiDAR. Signage content was created in PhotoShop CS using FHWA upper and lower case series B-E and font emulation and background template FHWA color schemes available from www.triskele.com, and saving them out as a 2^n dimensioned Portable Network Graphics (png) texture maps. Detailed imagery of various signs where obtained by drive-by photographs and video. Other building textures were obtained by extracting oblique views from MicroSoft (http://maps.live.com) or still photos taken on-site. The images in Figure 4.5 are road level views of the resulting model as projected on the front display. The location of each image as the observer travels westbound, is indicated.
E. fly-over entrance ramp
G.: View point of entrance merge onto I94 West Bound from 35W fly-over

Figures 4.5 (a – g) Final 3D road model, progressing west bound as indicated on map. The critical accident area is just beyond D and F.

3D Vehicle Models

The goal was to create detailed 3D vehicle models which accurately represent interior and exterior geometry with a minimal number of polygons. The process started with a single occupancy vehicle library set from purchased from Marlin Studios. In order to utilize each model for efficient real-time 3D rendering, each vehicle needed to be modified by creating four view eye distance Level Of Detail (LOD) nodes. The closest view LOD nodes contain the vehicle’s interior. Most of the LOD creation was done with the surface polygon reduction tools within MultiGEN Creator, using trial and error visualization within the cave to tune switching distances and the general percent reduction as show in Table 4.2.

Table 4.2 Vehicle model Level Of Detail (LOD) node structure.

<table>
<thead>
<tr>
<th>Distance (meters)</th>
<th>LOD Level</th>
<th>Polygon count</th>
<th>Percent Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 1.5</td>
<td>HIGH-with interior</td>
<td>2200</td>
<td>0%</td>
</tr>
<tr>
<td>1.5 to 25.0</td>
<td>HIGH</td>
<td>1800</td>
<td>20%</td>
</tr>
<tr>
<td>25.0 to 75.0</td>
<td>MED</td>
<td>700</td>
<td>55%</td>
</tr>
<tr>
<td>&gt; 75.0</td>
<td>LOW</td>
<td>450</td>
<td>80%</td>
</tr>
</tbody>
</table>

The final 3D models can either be directly imported into the application software developed for the project (described in detail next), or, as done for the application described herein, pre-
processed into an efficient format for rapid loading using the OpenSceneGraph (OSG) \textit{osgconv} utility.

\textbf{DEN Microsimulation Vehicle Trajectory Emulation}

An application utilizing OpenSceneGraph (Martz, 2007) was developed with the objective to ‘replay’ vehicle trajectories and observe candidate crash warning systems in the DEN. The latest version used to test and develop the software is OpenSceneGraph 1.2. compiled and installed under the Debian/Etch 2.6.18-4-686 SMP Linux operating system. The software architecture developed for this project is shown in Figure 4.6. The base libraries below the dotted line in the figure were developed previously and are described in the appendix.

\textbf{3D Road and Microsimulation Road Geometry Matching}

The microsimulation road network was created before the 3D model was built, with the 2D road section geometries being constructed by tracing over 6 inch resolution geo-referenced orthorectified high resolution photo tiles previously supplied by the MTO (GlobXplorer, June 2000). Due to the fact that different orthorectified image products of the same coverage area can vary based on the process used in their creation, and also due to the fact that traffic simulation typically does not need to consider the road geometry accuracy at the same level of detail as that required by detailed 3D visual simulation, the road network geometry needed to be adjusted using the actual 3D model geometry. This was accomplished by creating orthorectified image tiles of the 3D model and importing them as a background layer in the microsimulation software (AIMSUN). The microsimulation road network was then adjusted to match the road and lane boundaries from the 3D model.
Real-Time 3D Visualization of Vehicle Trajectories

Once the road geometries are properly matched, vehicle trajectories at desired sample intervals (a minimum output sample period of 0.5 seconds for AIMSUN) are extracted from the microsimulation and stored to a database utilizing the AIMSUN extension c++ library API. A mysql database is used to store all vehicle trajectory data (time, vehicle type, x, y, heading). The DEN application connects to the database via the ODBC protocol, and queries the database to buffer trajectories over several time samples. The returned data are post processed to clip trajectories within the road network geometry (the actual traffic simulation boundary was much larger) in order to improve computational performance on the display host (TM_UpdateCallback_SQLquery_vehicles and SQL_Vehicles objects in Figure 4.6).

Lastly, a state machine, TM_UpdateCallback_VehicleDispatch in Figure 4.6, keeps track of vehicles ‘entering’ and ‘exiting’ the displayed road network. As vehicles enter the network, they are dynamically assigned to a pool of ‘inactive’ (their OSG traversal nodemask is set to zero-state) vehicle objects pre-loaded in memory and dispatched within the visual simulation; when a vehicle ‘leaves’ the visualized road network, the state machine for the particular vehicle ID resets to the zero state. The vehicle position and orientation are calculated by the callback node object, TM_UpdateCallback_MoveVehicle, which is attached to every vehicle. The vehicle trajectory calculations involve interpolating vehicle position and orientation between the time...
steps in the database. Note that elevation is used within the microsimulation road network. Typically, the extra effort of precisely matching of the elevation to the actual road provides little or no benefit in order for the microsimulation model to reproduce the real-world traffic flow behaviors. However, as elevation is critical to the visual simulation, a methodology was developed which will be described next.

As mentioned previously elevation was not utilized in the microsimulation. Instead a road surface Z-elevation look-up table (ZeLUT) was created in order to match any point on the road surface to an elevation value (Figure 4.7). The advantage of building the ZeLUT is that computationally expensive intersection testing amongst many hundreds, if not thousands (neither case uncommon for large-scale microsimulation of real urban road networks), of vehicles in the simulation are avoided, while still representing subtle changes in road elevation. The creation of the ZeLUT was achieved by casting downward rays at one meter spacing along the Northerling and Easterling directions through all road objects in the database (via an Intersector object available in OpenSceneGraph). For each grid cell where there is an intersection, the elevation value, $z_{hit}$, and intersection hit number, $h_i$, for each cell $C_i$ are stored. The hit number attribute, $h$, is used to indicate roads crossing on top of each other by incrementing each time the ray pierces through each surface. The utilization of the attributes to impute vehicle elevation and pitch are described next.

As seen in Figure 4.7, overpasses must be recognized by the moving vehicles (colored in green for first layer, red for second overpass layer, etc.). Consider a vehicle position on the roadway surface, $(x_n, y_n)$, at time $t_n$, and a future position $(x_{n+1}, y_{n+1})$ at the next time interval $t_{n+1} = t_n + \Delta T$. The ZeLUT maps $(x_n, y_n)$ and $(x_{n+1}, y_{n+1})$ into existing cells, $C$, by common integer look up, i.e.:

$$C_i = \text{INT} \left( \frac{y_n - Y_{min}}{\Delta Y} + N_y \cdot \frac{x_n - X_{min}}{\Delta X} \right), \quad \text{for } C_i \in \mathbb{C}$$

$$C_j = \text{INT} \left( \frac{y_{n+1} - Y_{min}}{\Delta Y} + N_y \cdot \frac{x_{n+1} - X_{min}}{\Delta X} \right), \quad \text{for } C_j \in \mathbb{C}$$

where

$\Delta X$, $\Delta Y$ = ZeLUT cell spacing along the Easterling and Northling axes,

$N_y$ = number of cell rows in the ZeLUT, and

$X_{min}, Y_{min}$ = minimum boundary lines surrounding the visualized road network model.

Then the normative approximation to grade, $\|G\|$, is used to determine the elevation $z_{n+1}$ at $(x_{n+1}, y_{n+1})$, i.e.:

$$\| G \| = \text{MIN} \left\{ \frac{\left( z_{h_{j,i}} - z_{h_i} \right)^2}{\left( x_{n+1} - x_n \right)^2 + \left( y_{n+1} - y_n \right)^2} \right\}_{h_{j,i}=1...m}$$

56
To add robustness into the calculation, $\|G\| < G_{tol}^2$. If $\|G\| \geq G_{tol}^2$, then the previous elevation is used for $n+1$. Finally, vehicle pitch angle, $\beta$, is determined utilizing $\beta = G$.
Figure 4.7 Z-elevation LUT ‘voxel’ (a x,y,z cell point in the table) rendering of roadway. The different colors represent the ‘layer’ attribute, h, in the elevation LUT which in this case is due to road overpasses.
Limitations of Visual Simulation of Vehicle Motion Trajectories

Horizontal trajectories are created at a given point in time using linear interpolation between the stored time samples in the database. Although the computation is the least expensive, the algorithm is somewhat problematic when either the trajectory time interval becomes large, or when there is an abrupt change in the vehicle velocity. This is because under linear interpolation of position, the velocity vector is a step function which is not an accurate representation of actual corresponding vehicle following model and other driving behavior constraints. Other numerical methods such as k-3 kernel splines, or Lagrange polynomial interpolation kernels (Shoup, 1979), might be considered for vehicle trajectory imputation since they preserve vehicle velocity and acceleration from between trajectory sample points. Presently, the current data sets do not present any actual crashes, although traffic buildup leading to possible crash-prone conditions can be generated. In short, optimal trajectory ‘fitting’ functions which best represent actual vehicle dynamics during crash and non-crash simulation modeling remains an open issue for research and experimentation as to such differences between are actually perceived by the observer.

For this study, simulation trajectories are only considered for the west-bound direction, since the crash site of interest involved traffic flow in this direction. The software and database are not limited to this direction. However, rendering and computational performance of the current hardware places a practical limit on the traffic densities which can be visualized in real-time. After some experimentation with the micro-simulation trajectories, approximately 500 cars traveling within the road environment maintains a frame rate averaging 20 to 30 frames per second. Note that the frame rate improves as more vehicles are culled from the observer’s viewpoint. The process utilized to produce and render vehicle trajectories are described next, followed by the current limitations of the software.

Current Limitations of Simulation Observation

Limitations of the current hardware to render more dense traffic patterns was previously discussed. At present, there are several limitations in the software for what has currently been developed. These limitations will be three sections capitulated by subject observation and field of view, and driver warning systems. A solution path to overcome each limitation will be proposed for each.

In Vehicle Observation

The aim of the software is to allow observers to ‘sit’ inside vehicles during crash prone conditions. This involves ‘attaching’ the observer within the DEN to a moving vehicle. This task has not been completed to date. Currently, observers can utilize the wand to ‘drive’ or ‘fly’ down the roadway while the simulation is running. The wand can also be used to position the observer along the roadside or from any arbitrary perspective while the simulation is in progress. A mouse can also be utilized in absence of a 3D tracking device for similar functionality (most useful for single screen viewing by the experimenter, etc.). The method ‘sitting’ the observer within a vehicle entails attaching the DEN coordinate frame (as opposed to the camera viewpoint) to a desired moving vehicle object. In OpenSceneGraph, the simplest method to
implement this is to nest another Application Callback node underneath the moving vehicle callback node.

**Simulated Visual Driver Field of View**

As the screen sizes do no encompass the entire field of view of the observer, 3D stereoscopic clipping artifacts affects the feeling of being completely immersed. The affect will be strongest in the vertical field of view coverage. Figure 4.8 presents an example of an observer ‘sitting’ within one of the vehicle models. The observer is positioned in the driver’s seat such that the eye center is given. Given the current display screen dimensions of approximately 77 inches x 61 inches, suppose the observer is positioned in the driver’s seat such that the eye center is 26 inches from the screen front surface and 27 inches the left screen surface, and right screen surfaces (e.g., $\frac{77}{2} = 38.5”$), and approximately 30 inches from the top edges of the displays (and thus $61” – 30” = 31”$ from the bottom edges), then the effective vertical field of view, $\theta_v$ is:

$$\begin{align*}
\theta_{vf} &= \tan^{-1}\left(\frac{31.0}{26.0}\right) + \tan^{-1}\left(\frac{30.0}{26.0}\right) = 50.0° + 49.1° = 99.1° \text{ for the front screen} \\
\theta_{vsl} &= \tan^{-1}\left(\frac{31.0}{27.0}\right) + \tan^{-1}\left(\frac{30.0}{27.0}\right) = 48.9° + 48.0° = 96.9° \text{ for left driver side} \\
\theta_{vsp} &= \tan^{-1}\left(\frac{31.0}{49.8}\right) + \tan^{-1}\left(\frac{30.0}{49.8}\right) = 31.1° + 31.1° = 62.2° \text{ for right side}
\end{align*}$$

where, $\theta_{vf}$ = vertical angle for the front screen surface \\
$\theta_{vsl}$ = vertical angle for the left and right screen surface

Thus, the view of the inside cab interior will typically be clipped below bottom of the dashboard and doors. With current stereoscopic display technology used for this system, this limitation is difficult to mitigate completely. As a final note, it is recommended to position ‘sit’ the observer asymmetrically (Figure 4.8). The reason for this is to increase the left over-the-shoulder ‘blind spots’ field of view. In reference to Figure 4.8, a horizontal FOV angle of $\tan^{-1}\left(\frac{77}{27} - 26\right) = 62°$ degrees behind the driver’s center eye point is achieved. Note that, head tracking enables visual 3D simulation of look-over left-shoulder blind spot coverage to be increased by moving the eye point closer to the left display.
Figure 4.8 Scaled DEN placed `inside` a Chevrolet Blazer SUV. Note field of view (indicated by shaded frustums) will be affected by observer’s head movement.

Obviously it is not possible to turn around and look behind the vehicle. This can be mitigated by adding a fourth display. Even still, to remove 3D stereo clipping artifacts, a floor projection is needed which is difficult to achieve with the current 3D stereo implementation. Secondly, the current application does not simulate mirror reflected views. It is now possible to leverage OpenSceneGraph to implement a virtual mirror (the mirror itself and the reflected image are accurately rendered within the scene) however this requires a higher performance rendering platform than what is currently being utilized in the lab. Finally, some subjects are not able to fuse computer generated stereo imagery (while still others lack stereoscopic depth perception completely).

Lastly, a word on the software implementation is in order. Dual monitor output graphic cards are now ubiquitous. The performance of graphics processors and CPUs are far above the capabilities of the current hardware (description of current hardware configuration given in the appendix). Although OpenSceneGraph is designed to enable multi-pass rendering of the scene graph to
multiple viewports, the base software libraries previously developed and utilized for this project render to a single display device.

**Driver warning subsystems**

The original intent was to build in candidate driver warning systems for observational evaluations. Essentially, such warning systems are additional objects with programmed behaviors governed by an OpenSceneGraph callback node object attached to the visual object. For example, variable message signs, ‘programmed’ lane lights, in-vehicle visual indicators, etc. Such warning systems have yet to be implemented into the software. Writing any text can be achieved by changing the display state of OsgText objects. Road-side or lane-lights can be ‘turned-on’ simply by hiding/unhiding ‘lit’ and none-lit representations of the object. Lastly, the simulated road-side traffic sensing system which would then ‘trigger’ any of the warning subsystems objects has not been implemented.

Adding audio warnings could be achieved by utilizing the Simple DirectMedial Layer (SDL) audio functions to play back sound clips over stereo headphones (http://www.libsdl.org). If spatial localization is required (i.e., ‘outside the head’ sound), the OpenSceneGraph OpenAL (open Audio Library) can be used as a first approximation. More accurate 3D sound rendering systems typically dedicate a ‘sound server’ or high-end sound card (Creative Labs) to render one or more sound sources with respect to the user’s position and head orientation. This is because transforming the left and right ear sound signal as the observer turns and moves their head must be done in real-time. Perhaps only the advent of the sound and not its apparent source location is only relevant, and if such is the case, utilizing SDL with a reasonable quality sound card would suffice, simplifying the immersion display architecture significantly.

**Conclusions**

Although the original objectives were not met completely over the designated project period, several valuable insights were gained regarding the processes necessary to achieve high fidelity immersive real-time visualization of crash scenarios and warning prevention systems. First, building highly accurate models of complex test sights—especially to reproduce the complex geometry which may affect visual perception and plausible warning system deployments—involved considerable manual effort. It is fair to mention that recently there have been improvements in the tools from SimWright and MultiGen that may stream-line the aforementioned processes. For example, it may now be possible to transfer 2D or 3D automated edge extraction data directly into MultiGen—a process previously which could not be realized to the extent needed for this project. Secondly, the method develop so far is more generalized approach than what others have developed for exploring traffic safety issues using immersive virtual environments (Kearney et al, 2006), since the trajectory interface to data can be more seamlessly tied to real-time (see Pack et al., 2005) or a-priori generated traffic data. Finally, we believe a suitable framework has been developed which can now be used to explore a myriad of problems related to traffic safety.
5 Car-Following Models: Historical Review and Implications on Freeway Traffic Safety Modeling

Issues and Current Solution

Freeway rear-end crashes are a major roadway safety problem. Currently up to 25% of all crashes in the U.S. are rear-end collisions involving two or more vehicles. Though usually non-fatal, rear-end collisions more often than not induce non-recurrent congestions, causing substantial vehicle-delays while greatly degrading freeway performance. To be sure, approximately 157 million vehicle-hours of delay annually, or one-third of all crash-related delays in the United States are caused by rear-end collisions (Brown et al. 2000).

Over the past ten years, the focus on technological solutions to rear-end collision problem has continuously intensified, while various safety countermeasures and ITS innovations have been developed and/or deployed in the field aiming to reduce collision occurrences and improve safety performance. In spite of some successful implementations, it has been found that expected safety improvements from a particular ITS technique or safety countermeasure might turn out to diminished, sometimes even negative safety gain due to behavioral adaptation or unforeseen factors (Ledoux and Archer 1999). Clearly, in order to achieve optimal safety benefits, any safety treatment should be extensively (perhaps iteratively) assessed during its design stage and prior to field deployment. This calls for an effective and reliable tool that can assist testing, verification, and improving innovative safety concepts and technical solutions against different environmental, situational, and/or driver-vehicle conditions, without resorting to prohibitive and potentially hazardous field experimentations.

Microscopic traffic simulation features its modeling on individual vehicular interactions and detailed information about acceleration, deceleration, velocity, and position etc for each and every vehicle in the network under study. Due to its level of resolution and modeling realism, micro-simulation is potentially the most viable tool of choice for safety performance evaluation. However, most existing micro-simulation models by design target only normal driver behavior in typical traffic conditions. This means, either the functional structure of the underlying car-following models or parameters distributions of these models are deliberately constrained to outlaw unsafe behavior, hence excluding occurrence of vehicle crashes. Because of this, the use of micro-simulation has long been limited to collision-free scenarios. Some recent research efforts have focused on enhancing existing micro-simulation models by developing conflict statistics and safety surrogates to enable safety-related assessments (Hunuenin et al. 2005; Kosonen 1999; Kosonen and Ree 2000; Ledoux and Archer 1999; Mehmood et al. 2001; Torday et al. 2003). Yet due to the basic fact that crash occurrence is deliberately excluded in (current) micro-simulation models, it has been difficult to test and validate the correlation between extracted conflict statistics/surrogate measures with actual crash potential. This has inevitably tainted the credibility of safety-performance conclusions obtained by use of micro-simulation.
Review of Car-Following Models

Car-following models which describe the behaviors of vehicles following each other in longitudinal dimension without overtaking have been the subject of numerous research efforts in the past 50 years. The motivation of developing a good car-following model arises from the need to analyze the effects on traffic flow of proposed control measures or designs to the road network. Car following modeling forms the cornerstone of all microscopic simulation models and makes a major part in the development of modern traffic flow theory. In recent years, car-following models have become of increased importance as a clear understanding of car-following behavior is critical for new development, implementation and evaluation of various ITS techniques. The following sections give a comprehensive examination of major existing car-following models including the most recent developments.

Deterministic Stimulus-Response Models

Pipes-Gazis-Herman-Rothery Family

Deterministic stimulus-response models are defined essentially by their acceleration (response) function with respect to relative velocity (stimulus). The first dynamical equation describing this relationship was proposed by Pipes (Pipes, 1953), starting from an assumption that for every 10 mph increment of speed the following vehicle should leave one car length, i.e.,

\[ x_{n-1} = x_n + s_0 + L_{n-1} + \tau v_n \]  

(1)

where, \( s_0 \) is the legal distance between the vehicles at rest; \( x_{n-1} \) and \( x_n \) are the position for the preceding and following vehicles, respectively; \( L_{n-1} \) is the length of the preceding vehicle; \( \tau \) is constant in time unit; \( v_n \) is the velocity of following vehicle.

Rewrite Equation (1) as

\[ v_n = \frac{x_{n-1}(t) - x_n(t) - s_0 - L_{n-1}}{\tau} \]  

(2)

Differentiating both sides of Equation (2), one arrives at the dynamical equation proposed by Pipes:

\[ \frac{dv_n(t)}{dt} = \frac{v_{n-1}(t) - v_n(t)}{\tau} \]  

(3)

where \( v_n(t) \) is the velocity of following vehicle, \( v_{n-1}(t) \) preceding vehicle.

Solving Equation (3) one immediately arrive the conclusion that steady state solution, where all vehicles move at the same speed, is always stable so no clustering effects are described.

Chandler et al. (1958) introduced a finite time delay \( T \) into Equation (3), leading to

\[ \frac{dv_n(t + T)}{dt} = \frac{v_{n-1}(t) - v_n(t)}{\tau} \]  

(4)

Analytical analysis of Equation (4) indicates that the system reaches its stability limit when \( \frac{T}{\tau} = 0.5 \). Beyond this limit, i.e., when \( \frac{T}{\tau} > 0.5 \), the system is destabilized leading to spontaneous
clustering. Although capable of capturing the spontaneous clustering effect, this model is independent of the distance between vehicles which greatly restricted the model’s applicability in at least the following aspects. First, the fundamental relationship between speed and density cannot be derived from the model as there is no dependence on space headway in the model dynamics (Krauss, 1998); second, it can not describe the acceleration of a single vehicle in freeway flow.

Gazis et al (1959) introduced into the model a dependence on the distance between vehicles by defining:

$$\frac{dv_n(t+T)}{dt} = \alpha \cdot \frac{v_{n-1}(t) - v_n(t)}{x_{n-1} - x_n}$$

Integrating Equation (5) leads to the steady state speed-density relationship which agreed with Greenberg’s macroscopic theory (Greenberg, 1959):

$$\rho \propto \ln \left( \frac{\rho_{jam}}{\rho} \right)$$  \hspace{1cm} (6)

Edie (1961) supposed that the acceleration function should also be dependent on the momentary velocity of the following vehicle and recognizing the discontinuity in the fundamental diagram between freeway flow and congested flow, he proposed the dynamical equation:

$$\frac{dv_n(t+T)}{dt} = \alpha v_n(t) \frac{v_{n-1}(t) - v_n(t)}{(x_{n-1} - x_n)^2}$$  \hspace{1cm} (7)

where the sensitivity coefficient $$\alpha$$ has two different values for free-flow and congestion situation.

Gazis, Herman and Rothery (1961) unified and generalized the above formulations by allowing the sensitivity to dependent on arbitrary powers of spacing and momentary velocity:

$$\frac{dv_n(t+T)}{dt} = \alpha v_n(t)^m \frac{v_{n-1}(t) - v_n(t)}{(x_{n-1} - x_n)^l}$$  \hspace{1cm} (8)

where the exponents $$l$$ and $$m$$ are free parameters. When $$l = 2$$ and $$m = 0$$, linear steady-state speed–density can be derived which coincides with the results of Greenshields (Greenshields, 1935)

Similar critiques also apply to Gazis-Herman-Rothery model. Again, this model is not suitable for describing very low traffic where drivers accelerate to their desired velocity. The car-following behavior in dense traffic is also somewhat unrealistic. For example, the acceleration rate remains zero as long as velocity difference is zero regardless of the distance gap between vehicles. In addition, symmetric acceleration/deceleration (e.g., other conditions identical, and the relative velocity differs only in direction, then the respective absolute value of acceleration rate and deceleration rate would be the same) behavior is also unrealistic.

In order to capture more complex traffic flow patterns that exhibit capacity drops and hysteresis loops, on top of Pipes car following theory, Zhang (Zhang et al, 2004) recently proposed a new car following theory for multiphase vehicular flow:
\[ v_n(t + T) = \frac{d_n(t)}{\gamma_n(t)} \]  

(9)

where \( d_n \) is the distance separation between vehicles; \( \gamma_n \) is the desired gap time for the follower to travel the distance separation. Rather than using a constant gap-time for \( \gamma_n \), Zhang et al supposed \( \gamma_n(t) \) is dependent on both gap-distance and the traffic phase. According to Zhang et al (2004), one of the possible definitions for \( \gamma_n \) is

\[
\gamma_n = \frac{d_n(t)}{v_f} \quad \text{if} \quad d_1 \leq d_n(t)
\]

\[
= \gamma_1, \quad \text{if} \quad 0 \leq d_n(t) \leq d_0
\]

\[
= \frac{d_n(t)}{v_f} \quad \text{if} \quad d_0 \leq d_n(t) < d_1 \quad \text{and} \quad v_{n-1}(t) = v_f
\]

\[
= \gamma_1, \quad \text{if} \quad d_0 \leq d_n(t) < d_1 \quad \text{and} \quad v_{n-1}(t) < v_f
\]

Zhang’s model successfully replicated complex traffic patterns such as capacity drop and traffic hysteresis, however, this model could result in unrealistically high accelerations in some cases. For example, assuming a typical free flow speed \( v_f = 30 \text{ m/s} \) (67.5 mph), and a reaction time \( T = 1.8 \text{ s} \). According to this model, a free vehicle (initial speed 0) can achieve its desired speed in 1.8 seconds. This would require an unrealistically high acceleration of \( 30/1.8 = 16.67 \text{ m/s}^2 \). This critique was also confirmed by simulations results (Wang, 2004)

**Helly Family**

Based on Pipes original dynamical equation (i.e., Equ (3)), Helly (1959) added an additional linear term to allow for the adaptation of the acceleration according to the gap distance between vehicles:

\[
a_n(t + T) = \frac{dv_n(t + T)}{dt} = C_1[v_{n-1}(t) - v_n(t)] + C_2[x_{n-1}(t) - x_n(t) - D_n(t)] \]  

(10)

where \( C_1, C_2 \) are constant coefficients, \( D_n(t) \) is desired following distance given as:

\[
D_n(t) = \alpha + \beta v(t) + \gamma a_n(t) \]  

(11)

A major strength of Helly model is the implementation of an error element (Brackstone and McDonald, 1999). This means, Helly model may be implemented such that once an acceleration or deceleration rate is determined, the driver will not update the value till relative velocity or gap distance deviates substantially from the predicted value using this acceleration value. Major developments of Helly-like models include a complex model proposed by Xing (1995), which combines Gazis-Herman-Rothery model with Helly model allowing for different states such as standard driving, acceleration from standing queue, acceleration in free flow etc.

**Safety Distance Car Following Models**

The underlying logic for safety distance car following models starts from the assumption that drivers try to keep a safe following distance from preceding car in order to avoid potential
collisions. The first major development of this type of models was made by Gipps (1981). Recent new developments include Krauss (1998) and the so-called Intelligent Driver Model (Treiber et al, 2000)

**Gipps Model**

Gipps model is based on two velocity updating rules:

\[
\begin{align*}
v_n^a(t + \tau) &= \frac{v_n(t) + 2.5 \times a_{\text{max}} \times \tau \times \left[ 1 - \frac{v_n(t)}{V_n} \right] \times \sqrt{0.025 + \frac{v_n(t)}{V_n}}}{v_n(t) + 2.5 \times a_{\text{max}} \times \tau \times \left[ 1 - \frac{v_n(t)}{V_n} \right] \times \sqrt{0.025 + \frac{v_n(t)}{V_n}}} \quad (12) \\
v_n^b(t + \tau) &= b_{\text{max}} \times \tau + \sqrt{\left( b_{\text{max}} \tau \right)^2 - b_{\text{max}} \left( 2 \times [x_{n-1}(t) - L_{n-1} - x_n(t)] - v_n(t) \times \tau - \frac{v_{n-1}^2(t)}{b'} \right)} \quad (13) \\
v_n(t + \tau) &= \min \left\{ v_n^a(t + \tau), v_n^b(t + \tau) \right\} \quad (14)
\end{align*}
\]

where \( \tau \) is reaction time;

- \( a_{\text{max}} \) is the maximum desired acceleration rate for vehicle \( n \);
- \( b_{\text{max}} \) is the maximum desired braking rate for vehicle \( n \);
- \( x_n, x_{n-1} \) are the position of vehicle \( n \) and \( n-1 \), respectively;
- \( b' \) is driver \( n \)'s estimation for vehicle \( n-1 \)'s maximum desired braking rate;
- \( V_n \) is the desired speed of vehicle \( n \)

In essence, Eq (12) describes vehicle’s acceleration behavior in free flow when there is no constraints from leading vehicles, i.e., vehicle \( n \) is free to accelerate to its desired speed; while Eq (13) describes drivers car following behavior in dense traffic, that is, vehicle \( n \) (the follower) chooses a speed that ensures a **safe stopping distance** from preceding vehicle \( n-1 \), such that in the worst case a collision would not occur.

Gipps proved that this model is collision free as long as drivers don’t underestimate the desired braking rate \( b' \) for the preceding vehicle. Gipps offered no calibration of his parameters, but instead performed simulations using realistic values from normal distribution, finding that the model produced realistic behaviors such as disturbance propagation, asymmetric accelerations and decelerations. Further theoretical analysis (Wilson, 2001) indicated that Gipps model produces realistic fundamental relationship between steady state speed and density. In addition, instability of traffic flow under Gipps model is found to be related to the distribution of \( b' \) (i.e., drivers’ estimation of preceding vehicle deceleration capability) and drivers reaction time.

The main attractiveness of Gipps model is that all its parameters have realistic physical meanings, obviating the need for elaborate calibration process. However, it seems somewhat unreasonable to assume a driving strategy that always anticipates the preceding vehicle to brake to a complete stop, which has been subject to critiques in literature (Krause, 1998; Brackstone and McDonald, 1999).
**Krauss Model**

Krauss (Krauss, 1999) proposed a similar car following model parallel to Gipps model. The velocity updating rules is given as:

\[
\begin{align*}
    v_{n}^{\text{safe}}(t) &= v_{n-1}(t) + \frac{d_n(t) - d^{\text{des}}(t)}{\tau_b + \tau} \\
    v_{n}^{\text{des}}(t + \Delta t) &= \min [v_{n}^{\text{max}}, v(t) + a(v)\Delta t, v_{n}^{\text{safe}}(t)] \\
    v_{n}(t + \Delta t) &= \max [0, v_{n}^{\text{des}}(t + \Delta t) - \eta]
\end{align*}
\]  

(15)

(16)

(17)

where \(d_n(t)\) is the gap distance for vehicle \(n\); \(d^{\text{des}}(t)\) is the desired gap distance; \(\tau\) is driver reaction time; time scale \(\tau_b\) is defined as \(\frac{(v_n + v_{n-1})}{b}\), \(b\) is the deceleration rate; \(\eta\) is stochastic perturbation which is assumed to be \(\delta\) correlated in time.

Krauss model can address the transient traffic flow behavior such as capacity drop and the stability of so-called wide jams (jams with almost zero traffic speed propagating upstream with constant shock wave speed). He showed in this work that in order for microscopic models to exhibit correct description of capacity drop and thus of wide jams, the models need to inhibit mechanisms that yield higher queue inflows than queue discharge rates. This applies to all microscopic models for which reaction time (or mean headways during free-flow) is smaller than the so-called jam escape time. Car following models without this feature therefore cannot correctly describe congested traffic flow.

**Intelligent Driver Model**

Intelligent Driver Model (IDM) defines an acceleration function in relation to momentary vehicle velocity, gap distance and relative velocity (Treiber et al, 2000):

\[
\begin{align*}
    \frac{dv_n(t)}{dt} &= a[1 - \left(\frac{v_n(t)}{v_0}\right)\delta - \left(\frac{s^*}{s}\right)^2] \\
    s^*(t) &= s_0 + v_n(t)T + \frac{\frac{v_n(t)(v_{n-1} - v_n)}{2\sqrt{a.b}}}{\delta}
\end{align*}
\]  

(18)

(19)

where

\(s^*\) is the desired gap distance

\(T\) is safe time headway

\(a\) maximum acceleration

\(b\) desired deceleration

\(s_0\) jam space headway

\(v_0\) desired velocity

\(\delta\) acceleration exponent
IDM is accident free and can successfully capture complex traffic patterns including the self-organized characteristic traffic constants (Kerner, 1998), hysteresis effects and states transition; all the model parameters have meaningful interpretation and empirically measurable. In addition, the fundamental diagram and stability property can be easily calibrated.

**Optimal Velocity Model**

**Bando Model**

Optimal velocity models assume that drivers select their velocity according to an “optimal” velocity function, i.e.,

\[
\frac{dv_n(t)}{dt} = \frac{V^{opt} - v_n(t)}{\tau} \tag{20}
\]

Bando et al (Bando et al 1995) assumed the optimal velocity function is as follows:

\[
V(\Delta x) = \tanh(\Delta x) \tag{21}
\]

where \(\Delta x\) is gap distance.

This optimal velocity function \(V(\Delta x)\) has the desired property, i.e., as \(\Delta x \to 0\), \(V(\Delta x) \to 0\) and bounded as \(\Delta x \to \infty\). In this approach the speed-density relation can not be derived from the equation but are prescribed through \(V(\Delta x)\).

Bando’s model is widely used by physicists due to its simplicity and analytical properties. Qualitative properties of this model are almost exactly the same as those of the macroscopic Kerner-Konhauser model (Kerner and Konhauser, 1993, 1994). However, realistic velocity relaxation time (\(\tau\)) are of the order of 10 s for city traffic and 40 s for freeway traffic, therefore much larger than reaction delay time (Treiber et al, 2000). Crash are avoided only if \(\tau < 0.9\) s, leading to unrealistically high value of acceleration rate. Moreover, the acceleration and deceleration rate are symmetric with respect to the deviation from the optimal velocity, which is unrealistic as the absolute value of deceleration rate is usually stronger than that of acceleration (Treiber et al, 2000).

In order to overcome the problems, Helbing and Tilch (Helbing and Tilch 1998) proposed a generalized force model (GFM). GFM incorporated reactions to both relative velocity and gap distance, and agrees with empirical data fairly well. However, GFM may produce low congestion wave propagation speed (12.11 km/h compared to empirically observed 17~23 km); this problem was further address by Jiang et at (Jiang et al, 2001).

**Newell Model**

The model proposed by Newell (Newell, 1961) can be considered as an optimal velocity model, which is defined as:

\[
v_n(t + T) = V^{opt}(\Delta x) = v_0 \{1 - e^{-[(s_n - 1 - s_0)/v_0T]}\} \tag{22}
\]

This model incorporates a desired velocity \(v_0\) for vanishing interactions (\(\Delta x \to \infty\)) and a safe time headway \(T\) for describing car-following behavior in dense traffic. The model is collision
free, but the immediate dependence on density leads to unrealistically high acceleration rate of the order $\frac{v_0}{T}$ (Helbing and Tilch, 1998).

**Cellular Automata Model**

Cellular automata models are characterized by discretizing with respect to both time and space. The whole traffic dynamics is obtained through integer operations of cellular automata unit. These types of models are particularly widely used by physicists due to its computational efficiency and simple updating rules. Even though macroscopic traffic dynamics can be produced with this type of model, however, as no explicit driver behaviors are incorporated into the updating rules it is not suitable for behavior-intensive simulation and evaluations.

The first cellular automaton models applied in traffic simulation were proposed by Cremer and Ludwig (Cremer and Ludwig, 1986). However, the most widely-used cellular automaton traffic model is Nagel-Schreckenberg model (Nagel and Paczuski, 1995; Nagel and Schreckenberg, 1992). Specifically, in Nagel-Schreckenberg model the road is discretized into cells corresponding to segment of 7.5 meters in length. The cell has two states: empty or occupied, depending on the presence of a vehicle. Vehicle velocity is characterized by the number of cells it moves every time step. Each time step a vehicle’s velocity and position is updated based on its maximum desired velocity and whether there is a vehicle in front blocking its movement. Stochastic perturbations are introduced into Nagel-Schreckenberg model as a noise term in the updating rules. The model can produce empirical speed-density relationship and spontaneous clustering fairly well (Krauss, 1999).

As mentioned earlier, cellular automaton model is computationally efficient to be applied simulating large road networks. However, it seems insufficient to help further the understanding of complex traffic dynamics as the underlying rules are over-simplified without behavioral implications.

**Psycho-Physical Model (Action Point Model)**

Action point model describes drivers car following behaviors on the basis that the following vehicle attempts to maintain a desired gap distance and alters its acceleration response only when certain perceptual thresholds (action points) are exceeded. This approach was initiated by Todosiev (Todosiev, 1963) and expanded by Michaels and Cozan (Michaels and Cozon, 1962). Several existing microscopic traffic simulation programs including *Paramics* (Fritzche 1994) and *VISSIM* (Wiedemann, 1974) incorporated this approach.

Specifically, four thresholds for the following behavior are identified in Action Point Model (Wiedemann, 1971):

1. Minimum desired following distance $S_{\text{min}}$

   $$S_{\text{min}}(v) = S_0 + \alpha \sqrt{v}$$

   where $S_0$ is the desired spacing at rest, $\alpha$ is a coefficient. Both $S_0$ and $\alpha$ are probabilistically distributed; $v$ is velocity.
(2). Maximum desired following distance $S_{\text{max}}$

$$S_{\text{max}}(v) = S_0 + \alpha \sqrt{\beta \cdot v},$$

where $\beta$ is a coefficient which is also probabilistically distributed.

(3). Threshold for recognizing small negative (i.e., closing) relative speed:

$$\nu^{\text{cf}}(\Delta x) = -\frac{(\Delta x)^2}{\kappa^2},$$

where $\Delta x$ is the gap distance, $\kappa$ is a coefficient randomly distributed among driver population.

(4). Threshold for the perception of small positive (i.e., opening) relative speed:

$$\nu^{\text{op}}(\Delta x) = \frac{(\Delta x)^2}{\mu^2},$$

where $\Delta x$ is the gap distance, $\mu$ is a coefficient randomly distributed among driver population.

On crossing these thresholds, a driver may perceive an unacceptable change in relative speed or gap distance, and will initiate a change in the sign of his acceleration in the order of $0.2 \text{ m/s}^2$.

The difficulty with this approach is the lack of objective calibration of the individual parameters and thresholds, and consequently the models as a whole.

Fuzzy Logic Models

Fuzzy logic car following models describe the car following phenomenon through a set of “IF…THEN” rules developed though common sense and experience. For example, IF “not close” AND “opening” THEN “accelerate”. Chakroborty and Kikuchi (1999) first proposed fuzzy-logic based car following model. Other fuzzy logic models include Yihai et al (1993) and Rekersbrink (1995). Unfortunately, these type of models are computationally expensive and consequently not applicable to large network simulation (Krauss, 1998).

Implications of Car-Following on Traffic Safety Modeling

Traffic flow instabilities entail drastic fluctuations in space headway and relative speed that could lead to vehicle collisions. This issue has long been brought to the attention of traffic researchers and considerable research efforts have be devoted to investigate the microscopic foundations of individual vehicle dynamics and its correlation to traffic flow instabilities (Lubashevsky et al. 2003). This is usually done through analyzing fixed points stability of the dynamical equation describing car-following process. For example, it has been shown that the fixed point $\Delta \nu = 0$ for the GM family of car-following models:

$$a(t) = \alpha \frac{[v(t)]^\prime}{[\Delta x(t-\tau)]^\mu} \Delta v(t-\tau)$$

(23)
become locally unstable when \( a \tau \frac{v_0}{g_0} > \frac{\pi}{2} \) (Gazis et al. 1961); or asymptotically unstable for vehicle platoons if \( a \tau \frac{v_0}{g_0} > \frac{1}{2} \) (Gazis et al. 1961; Zhang and Jarrett 1997). Usually the instability is attributed to the collective motion of car ensemble; and considered to be the cause of non-damped oscillations in the relative motion of vehicles (See Figure 5.1). Moreover, linear stability analysis has indicated that when the magnitude of disturbances become sufficient, negative space headway could take place somewhere in between a platoon of vehicles indicating occurrence of vehicle crashes (Gazis et al. 1959; Gazis et al. 1961). Similar line of reasoning has been employed to help explain identified common freeway accident mechanism as collective effects (Davis and Swenson 2004).

![Figure 5.1 Measured car-following behavior. Data were collected from a Japan test track using GPS. This type of non-damped oscillations is typical of car-following process, sometimes referred to as “closing following spirals.”](image)

In fact, if the instability is indeed solely collective effect, then any dynamical equations similar to Equation (1) can hardly predict the instability of a lag vehicle when the lead vehicle is driving at a constant speed, i.e., no oscillation spirals will appear. However, this is not in accordance to empirical observations as shown in Figure 5.1. Indeed, the pattern in Figure 5.1 cannot be simply explained as a fixed point dynamical process disturbed with white noise, either; as in this case the oscillating spirals should distribute evenly around fixed points. Clearly, in addition to the (collective) structural instability as specified by dynamical equations for vehicle ensemble, there is an instability mechanism inherent to individual vehicle-driver-unit not related to collective effect (Lubashevsky et al. 2003). Such instability mechanism is unlikely a result of pure physical regularities; rather, driver-performance factors such as human visual perception process and decision errors may be vital. In order to realistically simulate real-life collision occurrences rather than generating artifacts, it is essential that a high fidelity car-following model takes into account both structural and the human-performance related instability and validates involving parameters against in-depth crash data.
Research Directions

As indicated above, in order for a car-following model to realistically replicate instability hence “generate” vehicle collisions, two types of instability mechanisms need to be captured, one is structural instability; the other is non-structural human performance related instability. So far, there seem no existing car-following models that are successful in combining these two mechanisms.

♦ Structural Instability

This type of instability is determined by physical regularities governing the mechanics of vehicle’s acceleration, velocity and displacement, i.e., the instability stipulated by the dynamical equations.

♦ Non-structural Instability

This type of instability arises from “non-physical” elements involved in the driving process. Such instability is rooted in the non-linear feature of driver perceptual-response process which is characterized by series neural-perceptual thresholds.

To recap, in order to generate vehicle collisions, or in order for the generated collision to be realistic, an **ideal car-following model should first be “reasonably” structurally non-stable.** Structurally non-stable means there exists a subspace of the entire parameters space for which the dynamics becomes unstable (i.e., it is realizable that \( \Delta x = 0 \) while \( \Delta v > 0 \)). Second, it is very important that the unstable parameter subspaces are physically meaningful. For example, if a car following model becomes unstable only when reaction time lies 0.1~0.3 s, this is considered not physically meaningful as such reaction time range is well beyond common distributions of human reaction times, while with such a small value, it is very unlikely vehicle collisions are mostly preventable. Furthermore, **the car-following model should incorporate non-structural instability by considering driver performance factors involved in driver’s perception and cognition process.** This needs to establish a series of perceptual thresholds that reflects the non-linear characteristics of human cognitive and perceptive process.
6 A Vehicle Trajectory Collection and Processing Methodology and Its Implementation to Crash Data

Introduction

Collision-inclusive vehicle trajectories, i.e., trajectories collected from real life crashes, are essential for safety-related microscopic traffic modeling. However, reliable collision-inclusive trajectories are hard to obtain. This is because crashes are rare and random events difficult to capture live; furthermore, there is still a lack of efficient, accurate and reliable post-processing algorithm that can effectively filter raw position errors and derive consistent speed and acceleration profiles. To be sure, data consistency is generally critical for microscopic modeling, but has not been emphasized in earlier studies. This requires positions, speeds and accelerations contained in the dataset must satisfy the respective integral and differential relationships throughout the relevant period of time. In earlier studies, raw trajectories were at best only processed to filter measurement errors in positions without resolving any data inconsistencies explicitly. This always resulted in inconsistent trajectories dataset, where accelerations when integrated over time don’t match the speeds profile; or speeds after integration don’t match the positions data.

In this paper, a methodology for collecting and processing collision-inclusive trajectories is presented. This includes a wireless-based video collection system, automated trajectories extraction procedures, and a post-processing algorithm specifically for filtering errors and developing consistent speed and acceleration profiles. Most importantly, this post-processing algorithm employs a bi-level optimization structure attempting to minimize measurement errors, and resolve internal inconsistency with positions, speeds and accelerations data. The proposed methodology is actually implemented to a high crash-rate freeway section in the Twin Cities, Minnesota. Over 700 hours of video recordings were collected, while 54 trajectories of relevant vehicles from 10 crashes/near-crashes were extracted. The extracted raw trajectories were then post-processed to smooth out errors and develop speed and acceleration profiles. It was found that the post-processing algorithm is very effective in eliminating both measurement and inconsistency errors from the extracted raw trajectories. Moreover, the proposed post-processing algorithm is further compared to Locally Weighted Regression, an approach used in earlier data collection studies (Toledo et al. 2007), by conducting a sensitivity analysis where the magnitude of measurement errors is intentionally varied with different values. The comparison results suggest that the proposed algorithm is not only more robust with respect to varying measurement errors, but also effective in removing data inconsistency from speed and acceleration profiles. This implies that the proposed data processing algorithm generally can generate more accurate and reliable vehicle trajectories data than the Locally Weighted Regression approach.

Background

Conventionally, vehicle trajectories (collision-free) are extracted from serial aerial photos taken from a certain altitude (Hall 1999; Herd 1968). One of the first photography-based data collection efforts was initiated by FHWA in the early 1980s, where vehicle trajectories datasets were developed from time-lapse aerial photos of six freeway sites (FHWA 1985). Recent
photography-based data collection efforts include the one conducted by Ossen et al. (Ossen et al. 2006; Ossen and Hoogendoorn 2005). Parallel to aerial photography, GPS and radar sensing technologies have also been applied to collecting vehicle trajectories. One major data collection effort using GPS technology was conducted by Grursinghe et al. (Grursinghe et al. 2002; Ranjitkar 2004). In their experiments, trajectories of 10 vehicles were collected from a number of runs on a test track using Real-time Kinematics GPS. Similarly, Ma (Ma 2005; Ma 2006) used one single probe vehicle equipped with GPS to collect trajectories during multiple test runs on Stockholm highways. Notably, under the auspice of the NHTSA, a large-scale GPS-based naturalistic driving study collected trajectories data from 100 probe vehicles (Dingus et al. 2006). Besides GPS technology, radar sensing was also commonly employed to collect trajectories data. For example, the Institute of Transport Research, German Aerospace Center has developed trajectories dataset containing 120 minutes trajectories at 0.1 second resolution using radar sensing technique (Hidas and Wagner 2004).

When compared to aerial photography, GPS and radar sensing-based approaches are generally more flexible and provide more detailed and accurate information on vehicle dynamics. However, the output data are only available for the instrumented vehicles; this usually results in relatively small sample size unfavorable for statistical analysis. In recent years, video-based data collection is becoming popular and considered as one of the most cost-effective ways of collecting large-scale trajectories from the field. Briefly, this involves extracting trajectories from successive video frames either manually, or through automated vehicle tracking software. Among others, the on-going NGSIM project sponsored by the FHWA is one of the largest data collection efforts attempting collecting (collision-free) trajectories from multiple field videos (Kovvali et al. 2007).

Regardless of the specific data collection approaches employed, once collected from the field, raw trajectories data generally are non-smooth and/or non-monotone due to measurement errors. These errors must be filtered out. In addition, instantaneous speeds and accelerations also need to be estimated so that the dataset can be eventually used in microscopic modeling. This means the raw trajectory data need to be post-processed. In earlier studies, the issue of post-processing trajectories has received little attention. More often than not, raw position data were not smoothed at all or only coarsely processed to filter errors, while speeds and accelerations are directly inferred from the position observations by simple finite differencing (i.e., subtraction of the position measurements in consecutive observations). For example, Smith (Smith 1985) collected position measurements using aerial photography. No post-processing was conducted even though the report characterized vehicle movements as jerky. Instantaneous speeds were directly inferred from the difference between two consecutive position observations. Wei et al. (Wei et al. 2005) developed a tool called VEVID that automates the trajectory collection. However, they do not perform any post-processing of position measurements. Instantaneous speeds and accelerations are again directly inferred by simple subtractions.

In effect, the type of coarse processing described above, particularly finite differencing actually introduces even more errors rather than reducing them. For example, it can be seen immediately from the following Taylor expansion in Equation 1 that finite differencing tends to introduce additional errors in the order of \( \sum_{n=2}^{\infty} \frac{(At)^{n-1}}{n!} f^{(n-1)}(t) \), resulting in unreliable speed and acceleration estimations:
\[ \frac{f(t + \Delta t) - f(t)}{\Delta t} = f'(t) + \sum_{n=2}^{\infty} \frac{(\Delta t)^{n-1}}{n!} f^{(n-1)}(t) \]  

where \( f(t) \) is the position profile, \( t \) is the time index, \( \Delta t \) is data collection interval, \( f'(t) \) is the actual speed at time \( t \), \( f^{(n-1)}(t) \) is the \((n-1)\)th order derivative of \( f(t) \), while \( \frac{f(t + \Delta t) - f(t)}{\Delta t} \) is the speed estimation.

In addition to measurement errors, Punzo et al. (Punzo et al. 2005) first explicitly raised the issue of consistency of the positions, speeds and accelerations estimated from the trajectory data. They pointed out that the time-series observations should satisfy basic equations of motion. This means, the estimated speed profile when integrated over time should match the position data, and similarly, the estimated acceleration profile after integration should match the estimated speed time series. Equivalently, this can be expressed as:

\[ \hat{x}(t) = \hat{x}(0) + \int_0^t \dot{v}(s)ds \]  

\[ \dot{v}(t) = \dot{v}(0) + \int_0^t \ddot{a}(s)ds \]

where \( \hat{x}(t), \dot{v}(t) \) and \( \ddot{a}(t) \) are estimated time series position, speed and acceleration data respectively. It should be stressed that the consistency in position, speed and acceleration data are generally critical for valid microscopic modeling, especially car-following studies. This is because generally the independent variables (output) of car-following models are accelerations while these models are practically validated based on position data. Any inconsistency in position, speed and/or acceleration data will be coupled with modeling errors leading to unreliable results. So far, there have been no post-processing algorithms that can effectively resolve data inconsistencies when processing trajectory data.

In the next sections, a methodology for collecting and processing trajectory data is presented. Particularly, this methodology includes an integrated crash video collection system, automated trajectory extraction, and most importantly, a new post-processing algorithm that seeks to minimize not only measurement errors in raw position data, but also resolve data inconsistencies in speed and acceleration estimations. The methodology is general, but particularly suitable for extracting trajectories of vehicles involved in crashes due to their sudden accelerating and decelerating characteristics and the need for accuracy in determining collision dynamics.
Figure 6.1 Proposed 3-stage data collection and processing methodology.
**Proposed Methodology**

The proposed methodology consists of three stages, i.e., crash video collection, raw trajectory extraction, and post-processing. This is best illustrated in Figure 6.1. As depicted in the figure, at the first stage crash videos are collected using an integrated wireless surveillance system developed at the University of Minnesota (Hourdos et al. 2004). This system captures live crashes from one selected high crash rate freeway section, then digitizes, compresses and transmits the collected video recordings to a central supervision station via wireless communication. At the second stage, the collected crash recordings are processed to extract raw trajectory data using video-based vehicle tracking techniques. Finally, the extracted raw trajectories are post-processed to filter measurement errors and generate speed and acceleration profiles. This methodology is presented next.

**Crash Video Collection**

Inasmuch as crashes are rare and random events, in order to maximize the number of crashes collected, an integrated video collection system was developed in the Minnesota Traffic Observatory (MTO) and deployed taking advantage of wireless communication technology. The purpose of this system is to facilitate capturing live crashes from high crash rate locations, then digitizing and transmitting the collected recordings to a remote supervising station for inspection and further processing. The architecture of the system is shown in Figure 6.1 (see Stage 1 in the figure) while a more detailed description of the system can be found in (Hourdos et al. 2004). As can be seen from the figure, this integrated system includes CCTV surveillance cameras, on-site small-factor PCs, and wireless communication infrastructure. Generally, the CCTV cameras can be deployed on the roof of a high-rise near the selected data collection site, monitoring and recording the traffic for a predetermined period of time during each day, e.g., from 8:00 am till 20:00pm. Analog videos from these cameras are digitized and compressed at optimized frame rates (e.g., 10 fps) by the small-factor PCs that are deployed on-site with the cameras. Digitized recordings are transmitted to a remote supervising station through the wireless communication infrastructure. At the supervision station, crash recordings are inspected to identify any crashes or near-crashes that have occurred during the day. Identified crash recordings are then stored separately to be processed in the next stage.

**Raw Trajectories Extraction**

In this stage, the collected crash recordings are processed to extract raw trajectories. Specifically, this is accomplished by using a video-based vehicle tracking program called *NG-VIDEO*, standing for Next Generation Vehicle Interaction and Detection Environment for Operations. *NG-VIDEO* is a C++ open-source program developed by the FHWA for extracting vehicle trajectories from videos captured by multiple cameras (Zhang et al. 2006). It has been successfully applied in the *NGSIM* project to extract (collision-free) vehicle trajectories from two freeway sections in California (Kovvali et al. 2007).

The working flow of applying *NG-VIDEO* to extracting trajectories is illustrated in Figure 6.1 (see Stage 2 in the figure). It should be stressed that this phase involves several non-trivial subtasks that have to be accomplished manually prior to applying the *NG-VIDEO* program.
These subtasks include Camera Calibration, World Coordinate Matching, and Video Orthorectification. Camera Calibration means obtaining the intrinsic parameters (focal length, aspect ratio, radial distortion etc.) of the surveillance cameras. Within the context of this study, this involves using the subject camera to snap a planar pattern shown at various orientations, and estimating the camera’s intrinsic parameters from the snapshots. Moreover, World Coordinate Matching is the process of mapping pixel coordinates from the crash recordings to world coordinates in real life. This is accomplished with the aid of the ArcGIS software and high-resolution (0.5 foot pixel image resolution) ortho-aerial images. Finally, Video Orthorectification requires registering the images of crash videos to the coordinate system in the aerial image. After orthorectification, terrain features in the video images are removed and vehicles appear to be moving on a plane surface. Once these subtasks are accomplished, the outputs are fed to NG-VIDEO. The latter generates raw trajectories data and stores them in a MySQL database.

Worthwhile to mention is that the underlying tracking algorithm used by NG-VIDEO is based on Maximum Correlation Matching (MCM) of intensity pixels (Kim and Malik 2003). This tracking algorithm overcomes the limitations of previous machine vision algorithms that have difficulty to obtain accurate vehicle positions in the presence of shadows and occlusions. This is very important as sometimes the selected data collection site is occluded by bridges and flyovers, and/or the traffic composition has large proportion of heavy duty vehicles that could shadow smaller vehicles passing by.

Data Post-Processing

In this stage, extracted raw position data are post-processed to filter potential measurement errors and develop consistent speed and velocity profiles. Considering that the trajectories of the vehicles involved in crashes usually contain mixed regimes (accelerating, decelerating, and cruising), instead of using one single polynomial, a more realistic fitting strategy is to divide the total time interval in smaller intervals and use lower degree polynomials in each of these subintervals. The ensemble of these pieces of polynomials is termed spline in the field of numerical analysis. In this study, the problem of filtering measurement errors and resolving data inconsistencies is formulated as a spline based bi-level optimization problem. The upper level optimization tries to minimize the inconsistency errors by looking for optimal number of polynomial pieces, while the lower level seeks the optimal coefficients matrix of the spline to minimize suitable measures of roughness subject to interpolation constraints. Speeds and accelerations are directly available from the coefficients matrix. The formulation is explained in detail below.

Mathematical Foundation

Mathematically, a piecewise polynomial, also termed spline, can be expressed as:

\[ S : [a,b] \rightarrow \mathbb{R}^1, \text{ and } S \text{ is composed of successive polynomial pieces } P_i \]

\[ P_i : [k_i, k_{i+1}) \rightarrow \mathbb{R}^1, \quad i = 0, \ldots, m - 2 \]

where

\[ a = k_0 < k_1 < \ldots < k_{m-1} = b \]
The given \( m \) points of \( k_i \) are called knots. The vector \( K = (k_0, k_1, \ldots, k_{m-1}) \) is called knot vector for the spline. Stated differently, the knots are break points of each local polynomial on successive subintervals \( [k_i, k_{i+1}) \), \( i = 0, \ldots, m - 2 \). Further, if any of the polynomial pieces on the respective subintervals has a degree up to \( n \), then the spline is said to be of degree \( n \) or of order \( n + 1 \). If two adjacent piecewise polynomials \( P_{i-1}(t) \) and \( P_i(t) \) at the break point \( k_i \) shares common derivative values from order 0 (i.e., the function value) up through the derivative of order \( r_i \) \( (r_i \leq n) \), then the spline is said to be of smoothness at least \( C^{r_i} \) at \( k_i \). In other words, when a spline is said to be of smoothness at least \( C^{r_i} \) at \( k_i \), the following mathematical conditions must be satisfied:

\[
P_{i-1}(k_i) = P_i(k_i)
\]
\[
\frac{dP_{i-1}(t)}{dt} \bigg|_{t=k_i} = \frac{dP_i(t)}{dt} \bigg|_{t=k_i}
\]
\[
\frac{d^2P_{i-1}(t)}{dt^2} \bigg|_{t=k_i} = \frac{d^2P_i(t)}{dt^2} \bigg|_{t=k_i}
\]
\[
\vdots
\]
\[
\frac{d^rP_{i-1}(t)}{dt^r} \bigg|_{t=k_i} = \frac{d^rP_i(t)}{dt^r} \bigg|_{t=k_i}
\]

A vector \( R = (r_1, r_2, \ldots, r_{k-2}) \) such that the spline has smoothness \( C^{r_i} \) at \( k_i \) for \( 0 < i < k - 1 \) is called smoothness vector for the spline. Given a knot vector \( K \), a smoothness vector \( R \), and \( n \times (m - 1) \) matrix \( M \) of its local polynomial coefficients, the set of all splines of degree \( \leq n \) comprises a spline vector space that can be denoted by \( S_n^R(K) \).

**Formulation of the Bi-Level Programming**

**Lower Level Optimization**

According Newton’s physical law, the (longitudinal) movement of a vehicle can be described using polynomials of degree 2 if the vehicle acceleration is constant; or polynomials of degree 3 when the acceleration is also time dependent. Considering the fact that in congested traffic, and particularly in crash situations, a driver may frequently alternate between braking and accelerating, a spline of degree 3 (order 4) is suitable for interpolating the actual trajectory. Also the spine should have smoothness \( C^{2} \) at each knot, meaning that for two adjacent local polynomials, the 0 order function value (position), and the first order derivative (velocity) should be the same at the break point. This ensures that both the position trajectory and speed profile are continuous at the break points. Note \( C^{3} \) smoothness is not necessary; this means acceleration profile doesn’t have to be continuous over time. This is because a driver may brake immediately after accelerating in real life driving causing abrupt changes in the acceleration profile.

In short, the objective of lower level optimization is to filter (smooth) potential measurement errors while generating velocity and acceleration profiles for each vehicle. At this level, the number of breaking points, i.e., knots, is fixed. This can expressed as:
where

\[ K = (k_0, k_1, \ldots, k_{m-1}) \] is the knot vector; \( m \) is the number of knots;

\( n = 3 \) is the degree of the spline;

\( m \) is the total number of knots;

\( m - 1 \) is the total number of local polynomial pieces;

\( S(t) \) is the spline to be optimized;

\( TR(t) \) is the raw trajectory time series;

\( t_j \) is the time stamp of raw trajectory data;

\( J \) is the total number of data points contained in the raw trajectory;

\( P_i(t) \) is the local polynomial that composes the spline \( S(t) \); \( i = 0, 1, \ldots, m - 1 \);

\( A_{\min}, A_{\max} \) are lower and upper bounds for acceleration rate, respectively.

As mentioned earlier, the objective of this lower level optimization is to find an optimal spline that minimizes the least square deviation between the actual trajectory time series and the points predicted by this spline. In other words, the purpose is to find a best piecewise polynomial that is composed of \( m-1 \) local polynomials. Each local polynomial corresponds to the driver’s acceleration or braking action, or mixed. This reflects the motivation of using a spline rather than a single polynomial, i.e., in congested traffic or crash situations, vehicle movements are characterized by frequent changes in acceleration or deceleration. Also this lower level optimization is subject to a constraint of \( C^2 \) smoothness. This ensures that position trajectory as well as speed profile at the break points are both continuous and smooth. Note, the number of break points (knots) is fixed at this level; optimizing this variable is the subject of upper level optimization.

**Upper Level Optimization**

The upper level optimization seeks solution to minimize inconsistency errors. Specifically, this is formulated as:
Figure 6.2 Schematic work flow of the spline based bi-level programming process.
As shown in (2.7), the objective function is the squared difference between the optimal spline found in the lower level optimization, and the position series reconstructed using acceleration profiles estimated from the spline. The aim is to find optimal number of break points, or equivalently, optimal number of polynomial pieces that minimize the inconsistency error. Figure 6.2 illustrates the schematic work flow for this bi-level programming process.

**Implementation**

The proposed methodology has been implemented to a high crash freeway section in the Twin Cities, Minnesota. Figure 6.3 (a) gives an aerial view of this section. Specifically, the section is in Interstate-94 westbound in downtown Minneapolis between Park Ave and 3rd Ave. It is a three-lane section of approximately 1900 feet in length. Average daily volume is approximately 80,000 vehicles per direction. Figure 6.3 (b) provides a summary of the 2002 crash statistics both for this section as well as other high crash-rate locations in the Twin Cities area (Hourdos 2005). As shown in Figure 6.3 (b), the regional average crash rate in the Twin Cities area is 0.96 million-vehicle-miles (MVM) in 2002, whereas the selected section has the highest 4.81 mvm compared to other locations. This is almost four times higher than the regional average level and roughly equivalent to one crash every two days.

The integrated video data collection system was deployed on the roof of a high-rise building near the site. Figure 6.3 (c) shows the picture of two surveillance cameras of the system. The data collection started from May 2006 and ended in August 2006. By the end of the four-month data collection period, a total of over 700-hours video recordings (digitized and compressed at 10fps) were collected. Even though a number of crashes and near-crashes had been identified from the recordings, eventually only 6 crashes and 4 near-crashes were selected for further processing. These crash cases were selected because they were captured with good image quality, also the collision location in each case permits extracting trajectories of at least 5 vehicles in the platoon for up to 90 seconds. Finally, identified video recordings were saved in two formats. The first format is one-hour video for each crash/near-crash case, covering traffic situation 45 minutes before and 15 minutes after the time the crash/near-crash occurred, while the second is 90-second video clip that only covers 1.5 minutes of traffic situation before the crash/near-miss occurred. The first type of videos provides comprehensive visual information regarding traffic dynamics before and after crashes, while the second type is used to extract crash trajectories.

Figure 6.4 demonstrates the extracted trajectories of 8 vehicles that had been involved in a crash. The crash occurred on the rightmost lane between Portland Ave and the flyover to I-94 Westbound at 18:27:40 pm, August 8, 2006. At the time the weather conditions were dry. A total of 54 trajectories had been extracted from the 10 crashes/near-crashes. Similar plots for other crashes/near-crashes are not presented here due to space limitations. In Figure 6.4, it can be clearly seen that the lead vehicle, denoted as Veh09, started braking hard at approximately 28 sec.
Figure 6.3 (a) Data collection site: a high crash-rate freeway section of Interstate-94 WB.

Figure 6.3 (b) Ten highest crash sections in 2002. (Source: Mn/DOT Crash Facts 2002.)

Figure 6.3 (c) Surveillance cameras that monitor the site.
from a speed of about 51 mph. In response, the following vehicles, i.e., Veh10 and Veh11 decelerated and adjusted their speeds accordingly. However, Veh14 did not apply braking; instead it switched to the adjacent middle lane while keeping its initial speed. Veh06 seemed to have extraordinary long reaction time that leaves the last vehicle in the platoon, i.e., Veh08 no time to react till the crash occurred. Figure 6.5 illustrates post-processed positions, speeds and accelerations data of the lead vehicle (Veh09). As can be seen from Figure 6.5, the post-processing algorithm removed the spikes in the raw position data while the overall patterns are well retained. Figure 6.5 also plots the differences (errors) between the raw and filtered positions. The mean of the errors are found to be at the magnitude of $-2.46 \times 10^{-5}$ ft, and standard deviation 0.66 ft. This means the post-processing algorithm is effective in filtering out errors without distorting the original trend.
Table 6.1 Error statistics and optimal number of local polynomials when post-processing vehicle trajectories from the crash occurred on Aug 8, 2006.

<table>
<thead>
<tr>
<th>Vehicle ID</th>
<th>Error Measures</th>
<th>Position Measurement Errors</th>
<th>Position Inconsistency Errors</th>
<th>Speed Inconsistency Errors</th>
<th>Optimal Number of Local Polynomials ($m^*$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MAE (ft)</td>
<td>RMSE (ft)</td>
<td>MAE (ft)</td>
<td>MAE (ft/s)</td>
</tr>
<tr>
<td>Veh09</td>
<td>Position</td>
<td>0.263</td>
<td>0.667</td>
<td>1.752</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td>Measurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Veh10</td>
<td>Errors</td>
<td>0.158</td>
<td>0.346</td>
<td>1.476</td>
<td>1.780</td>
</tr>
<tr>
<td>Veh11</td>
<td>Error</td>
<td>0.229</td>
<td>0.530</td>
<td>1.497</td>
<td>1.610</td>
</tr>
<tr>
<td>Veh14</td>
<td>Measures</td>
<td>0.212</td>
<td>0.533</td>
<td>0.902</td>
<td>1.073</td>
</tr>
<tr>
<td>Veh15</td>
<td></td>
<td>0.276</td>
<td>0.569</td>
<td>0.479</td>
<td>0.637</td>
</tr>
<tr>
<td>Veh06</td>
<td></td>
<td>0.439</td>
<td>0.877</td>
<td>0.847</td>
<td>1.059</td>
</tr>
<tr>
<td>Veh07</td>
<td></td>
<td>0.307</td>
<td>0.627</td>
<td>0.00012</td>
<td>0.000146</td>
</tr>
<tr>
<td>Veh08</td>
<td></td>
<td>0.513</td>
<td>0.939</td>
<td>0.964</td>
<td>1.225</td>
</tr>
</tbody>
</table>
Figure 6.5 Post-processing the trajectory of Veh09 of the crash occurred on Aug 8, 2006.
Table 6.1 summarizes the error statistics when post-processing the eight vehicles in the crash on Aug 8, 2006. Statistics for other crash/near-crash cases exhibit similar trends hence omitted here. Specifically, three types of errors are inspected in the table, i.e., measurement error, position inconsistency error, and speed inconsistency error. For each type of error, three measures are evaluated:

1) Mean Absolute Error (MAE)

\[
MAE = \frac{1}{n} \sum_{i} |x_i - y_i|
\]

2) Root Mean Square Error (RMSE)

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i} (x_i - y_i)^2}
\]

3) Mean Absolute Percentage Error (MAPE)

\[
MAPE = \frac{1}{n} \sum_{i} \frac{|x_i - y_i|}{x_i}
\]

where \( n \) is the number of total observations; \( x_i \) is the raw position data; \( y_i \) is the filtered position data when evaluating measurement errors; or, \( x_i \) is the filtered position data, \( y_i \) is the position data reconstructed by integrating twice the estimated acceleration profile when evaluating position inconsistency errors. Likewise, when evaluating speed inconsistency errors, \( x_i \) is the estimated speed profile, \( y_i \) is the speed data reconstructed by integrating the estimated acceleration profile. As shown in Table 6.1, the proposed post-processing algorithm is effective in removing measurement errors while retaining internal data consistency between position, speed and acceleration data. For example, when filtering raw positions data, the MAE ranges from 0.158 feet to 0.513 feet, RMAE ranges from 0.346 to 0.939, while MAPE ranges from 0.0296% to 0.0982%. These indicate that the post-processing algorithm succeeded in removing errors without distorting the original trends in the data. Also the inconsistencies in position data are well below 1.752 ft (MAE), and speed inconsistencies are constrained under 0.536 ft/s. These low error statistics suggest the post-processed trajectories have relatively high accuracy and reliability suitable for microscopic modeling. Table 6.1 also presents the optimal number of local polynomials (i.e., \( m - 1 \)) for each vehicle trajectory. This variable is optimized in the upper level optimization. From the table, it can be seen that the optimized number is different for each individual vehicle. This is because different drivers have different characteristics thus the respective trajectory has different regimes of vehicle dynamics that cannot be described by a fixed number of local polynomials.

**Sensitivity Test**

In order to best evaluate the effectiveness of the proposed post-processing algorithm, a sensitivity test was performed as part of this study. In the test, filtered position data of one previously processed vehicle are taken as ground truth, and then zero-mean random noises are added to it, with the standard deviation of the noise intentionally incremented from 1 ft to 14 ft.
In Figure 6.6 the ground truth data and the data with noise are presented, where the left chart shows the noise with a standard deviation of 6 ft, the right chart 8 ft (other cases with standard deviations ranging from 1 to 14 ft are not presented due to space limitations). The proposed post-processing algorithm was applied to the noisy data to filter position errors and derive speed/acceleration profiles. Meanwhile, Locally Weighted Regression (LWR), an approach that has been employed in earlier studies (e.g., the NGSIM project), was also applied to post-process the noisy data. Results obtained with the post-processing algorithm proposed in this study (denoted as BOA for simplicity, meaning Bi-Level Optimization Algorithm) are then compared to those obtained with Locally Weighted Regression (LWR). Note, a description of Locally Weighted Regression is not presented here but interested readers are recommended to refer to (Toledo et. al. 2007). Figure 6.7 and Figure 6.8 summarize the comparison results.

In Figure 6.7, the MAE measure that evaluates the difference between ground truth positions and filtered positions are presented. As indicated in this figure, when the standard deviation of the noise is small (< 2 ft), both BOA and LWR can achieve similar MAE below 1.0 ft. However, when the magnitude of the noise increases, the MAEs effected by LWR increase rapidly, reaching as much as 10 ft. By contrast, BOA is very robust with increased noise. From the figure, it can be seen that, regardless of the magnitude of the noise, the MAEs effected by BOA are always below 3.5 ft. Additionally, it can be observed that the window size influences the effectiveness of LWR. Generally the wider the window size, the better the filtering results (see Figure 6.6). This is because wider widow sizes mean more information is taken into account when conducting the regression.

Figure 6.8 compares BOA to LWR in terms of their respective effectiveness in resolving data inconsistencies. Figure 6.8 (a) illustrates the MAEs for position inconsistencies, while Figure 6.8 (b) depicts the MAEs for speed inconsistencies. In both figures, it can be seen that the BOA results in almost zero inconsistency errors in both position and speed data. LWR performs fairly well when the magnitude of the noise is less than 1 ft. However, when the standard deviation increases, LWR fails to resolve inconsistency errors and the latter increased rapidly. For example, when the standard deviation of the noise is 4 ft, BOA resulted in position inconsistency error of 1.399 ft (MAE measure); by contrast, under the same noise level, LWR with window size 7 resulted in position inconsistency error of 202.3 ft, with window size 9, 163.1 ft, and with window size 15, 20.79 ft. Even though with increased window size, LWR can reduce the inconsistency error from over 200 ft to 20.79 ft, but still, this number is almost 20 times higher than BOA results. Similarly, with 4 ft standard deviation of the noise, BOA achieved a speed inconsistency error of only 0.515 ft/sec; while LWR with window size 7 resulted in speed inconsistency error of 37.9 ft/sec, with window size 9, 27.74 ft/sec, and with window size 15, 5.85 ft/sec. Again, the speed inconsistency error under LWR at best is almost 10 times higher than BOA. Clearly, this suggests that LWR is not effective in resolving data inconsistencies when the data is noisy. This is because LWR uses a moving window to fit a local polynomial for every point in the middle of the window. The noise distorts the resulted polynomial for the subject data point due to amplified discontinuities. When the window moves to the next point, the inconsistency is further accumulated. As a result, the internal consistency for the dataset is broken, leading to significant inconsistency errors.
Concluding Remarks

In this paper, a comprehensive methodology for collecting and processing collision-inclusive vehicle trajectories is developed. This methodology includes a wireless-based video collection system, automated trajectories extraction procedures, and a post-processing algorithm specifically for filtering errors and developing consistent speed and acceleration profiles. This post-processing algorithm employs a bi-level optimization structure trying to minimize not only measurement errors, but also resolve internal inconsistency with positions, speeds and accelerations data. The proposed methodology is actually implemented to the highest crash-rate freeway section in the Twin Cities, Minnesota. Over 700 hours of video recordings were collected, while 54 trajectories of relevant vehicles were extracted from 10 crashes/near-crashes. The extracted raw trajectories were post-processed using the proposed algorithm. It was found that the post-processing algorithm is very effective in eliminating both measurement and inconsistency errors from the extracted raw trajectories. Moreover, the proposed post-processing algorithm is further compared to Locally Weighted Regression (LWR), in a sensitivity analysis where the magnitude of measurement errors is intentionally varied with different values. The comparison results show that when the standard deviation of the noise increases from 1 ft to 14 ft, the proposed algorithm can reduced the magnitude of position errors to less than 3.5 ft (MAE measure), while LWR resulted much higher position errors in the order of 10 ft (MAE measure). Also the proposed algorithm was found to be able to achieve much lower position and speed inconsistency errors than LWR. The improvement to the error statistics was in the order of at least 10~20 times lower when comparing the new algorithm to LWR approach. This suggests that the new algorithm is not only better at filtering position errors, but also more effective in removing data inconsistency from speed and acceleration profiles. In other words, the proposed data processing algorithm generally can generate more accurate and reliable vehicle trajectories data than the Locally Weighted Regression approach. Finally, before closing it should be stressed that the methodology presented in this paper is particularly suitable for extracting trajectories of vehicles involved in crashes due to their sudden accelerating and decelerating characteristics. However, the methodology by itself is a general one, meaning it is also applicable to collecting and processing collision-free trajectories.
Figure 6.6 Data used in sensitivity test: ground truth vs. ground truth added with noise.  
Left: Noise Mean 0, Noise Standard Deviation 6 ft; Right: Noise Mean 0, Noise Standard Deviation 8 ft
Figure 6.7 Comparing the bi-level optimization algorithm to locally weighted regression: mean absolute errors between ground truth position data and filtered position data.
Figure 6.8 (a) Comparing the bi-level optimization algorithm to locally weighted regression: mean absolute errors for position inconsistencies.
Figure 6.8 (b) Comparing the bi-level optimization algorithm to locally weighted regression: mean absolute errors for speed inconsistencies.
7 Literature Review on Safety Surrogate Measures for Micro-simulator

Traffic safety has been regarded as one of the top concerns in society. Many innovative safety improvements have been proposed to reduce the number of crashes and crash potential. An effective evaluation of such improvements is very important. Intuitively, an ideal evaluation involves analysis of the number of crashes collected at a project site over a long observation period. However, directly measuring the number of seldom occurring crashes or heavily relying on limited historical data is insufficient and undesirable even impossible in some cases. For example, a safety evaluation of a proposed road geometric improvement should be done before construction starts. In response to the shortcomings of relying solely on historical crashes data, safety surrogate measures, which are defined as quantifiable observations that replace or be a supplement to crash records, have been the focus of considerable studies. For years now, micro-simulation has been widely used as a cost-effective evaluation tool in road planning and traffic management. However, it is controversial and challenging to evaluate safety alternatives in a micro-simulator, a perfectly safe environment where no crash can occur. The underlying car-following, lane change, and gap acceptance models always assure a safe situation. Nevertheless, micro-simulators are capable of extracting vehicles’ position, speed, acceleration/deceleration profiles, etc; these profiles could be used to calculate safety surrogate measures. Thus, the safety surrogate measures derived from a micro-simulation hold a great potential to assess effectiveness of safety solutions. A variety of safety surrogate measures has been applied in the different traffic situations and will be discussed in the following sections.

Safety Surrogate Measures for Intersections

Safety surrogate measures at an intersection have been developed and refined since the Traffic Conflict Technique (TCT) was initiated by Perkins and Harris at the Detroit General Motors Laboratory. Subsequently, considerable research has been conducted to introduce additional safety surrogate measures as well as to enhance corresponding approaches (Perkins and Harris 1967; Hayward 1972; Hyden et al. 1982; Perkins and Bowman 1986; McCoy and Peterson 1988; Archer 2000; Gettman and Head 2003). Hyden et al. (1982) first gave the conflict definition and made a significant contribution to TCT at an applicable level. One project, called SINDI (acronym of Safety Indicator) focused on the safety problems associated with different types of urban intersections and prompted efforts to develop a more detailed driver behavior model for the Helsinki Urban Traffic Simulator (HUTSIM)(Kosonen, I. 1991). Gettman and Head (2003) investigated the potential of deriving surrogate measures based on the capabilities of current micro-simulation models. In their report, a set of safety surrogate measures related to TCT have been identified. The most prevalent safety surrogate measures are summarized in Table 7.1. Most of them were primarily used to indicate the degree of conflicts at an intersection. Our objective is to evaluate the effectiveness of safety improvements at a high crash site in an urban freeway. Thus, most of them may be suitable for conflicts in the ramp junction while the last three surrogates (DR, GT, and PSD) may be applied in the freeway section.
<table>
<thead>
<tr>
<th>Surrogate Measures</th>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to Collision</td>
<td>TTC</td>
<td>The time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained.</td>
</tr>
<tr>
<td>Encroachment Time</td>
<td>ET</td>
<td>Time duration during which the turning vehicle infringes upon the right-of-way of the vehicle going straight.</td>
</tr>
<tr>
<td>Initially Attempted Post-Encroachment</td>
<td>IAPT</td>
<td>Time lapse between commencement of encroachment by a turning vehicle plus the expected time for the through vehicle to reach the point of collision and the completion time of encroachment by the turning vehicle.</td>
</tr>
<tr>
<td>Post-Encroachment Time</td>
<td>PET</td>
<td>Time lapse between end of encroachment of turning vehicle and the time that the through vehicle actually arrives at the potential point of collision.</td>
</tr>
<tr>
<td>Deceleration Rate</td>
<td>DR</td>
<td>This measure is quite simply a measure of the highest rate at which a vehicle must decelerate to avoid a collision.</td>
</tr>
<tr>
<td>Gap Time</td>
<td>GT</td>
<td>Time lapse between completion of encroachment by turning vehicle and the arrival time of the crossing vehicle if they continue with same speed and path.</td>
</tr>
<tr>
<td>Proportion of Stopping Distance</td>
<td>PSD</td>
<td>Ratio of distance available to maneuver to the distance remaining to the projected location of collision.</td>
</tr>
</tbody>
</table>

**Safety Surrogate Measures for Work-zones**

It is generally accepted that a higher crash rate is more likely if speed variance increases. Zhu and Saccomanno adopted two safety surrogate measures: uncomfortable deceleration and speed variance to explore the relative collision risks under different work zone lane closure layouts. These two parameters were estimated from the microscopic simulator “Integration”. In their model, a total of 70 virtual detectors were placed along a 1500 meter testing segment to measure traffic conditions. The 20 seconds average deceleration rate can be calculated from the following equation:
\[ D = \frac{(S_u / 3.6)^2 - (S_d / 3.6)^2}{2 \times \text{Space}} \]  

(1)

\( D \) represents deceleration rate \((m/\text{sec}^2)\)

\( S_u \) is average speed on upstream loop detectors \((km/h)\)

\( S_d \) is the average speed on downstream loop detectors \((km/h)\)

Space is the distance between adjacent loop detectors

After obtaining the deceleration rate profile, they set 9.8 ft/sec\(^2\) as the threshold value of uncomfortable deceleration and identified the exceeding values to indicate the relative crash risk. The average speed variances at the beginning of the work zone were also derived from the simulator. The evaluation of proposed different layouts was accomplished based on these two surrogate measures. These surrogate measures are potentially applicable to represent the traffic conditions prior to rear-end collisions caused by stop/slow traffic conditions.

**Safety Surrogate Measures for Tunnels**

Safety concerns have been raised for tunnels due to fires and crashes. To evaluate preventive solutions, it is necessary to identify surrogate measures which describe unsafe traffic conditions. In Europe, early on, spatial headway was considered as a safety standard for tunnels while later time headway was referred by the European Directive. Domenichini et al. (2004) conducted a safety analysis in a 500 meter long virtual tunnel using the Advanced Interactive Microscopic Simulator for Urban and Non-urban Networks (AIMSUN). They first examined the relationships between flow, density, speed, time headway and acceleration under various traffic conditions from Level of Service (LOS) A to LOS E. With a reduction in LOS, driving freedom will decrease as vehicle interaction increases. This could contribute to potential crashes. The vehicles’ average speed, speed variance, acceleration/deceleration were proposed as surrogate measures to describe the degree of this increasing interaction and indicate an unsafe situation. However, an unexpected result showed that the average speed and speed variance both are reduced as LOS decreases. This may imply an increase in safety. Therefore, the average speed and speed variance were not sufficient to indicate unsafe traffic conditions. Finally, it was found that the vehicle’s acceleration/deceleration can reflect the actual level of interaction among vehicles. Thus it is a proper surrogate measure in the tunnel studies.

**Safety Surrogate Measures for Freeway Entrance Ramps**

A surrogate measure: unsafe density \((UD)\) was presented in a freeway ramp case study (Barcelo et al. 2003; Torday et al. 2003; Huguenin et al. 2005). Before discussing \(UD\), a local variable: “unsafe parameter” was first defined to determine the possibility of a hypothetical collision between two consecutive vehicles. It can be expressed as the products of three variables:

\[ U = \Delta S \cdot S \cdot R_d \]  

(2)

here \( U \left(\frac{m^2}{s^2}\right) \) is an unsafe parameter.
\( \Delta S \left( \frac{m}{s} \right) \) represents the speed differential between leader and follower.

\( S \left( \frac{m}{s} \right) \) is follower’s speed.

\[ R_d = \frac{R}{R_{\text{Max}}} \] is a ratio between the deceleration of the leader and its maximum deceleration capacity.

The above equation implies that the probability of a hypothetical crash between two consecutive vehicles is proportional to both the follower’s speed and the speed differential between leader and follower when the leader starts braking. A higher value of \( U \) indicates a higher unsafe level.

At each simulation step, \( U \) is calculated at a specific location and for each consecutive pair of vehicles. \( UD \) is an average value of \( U \) over every link of the road network at each aggregation period. It can be expressed as:

\[
UD = \frac{\sum_{s=1}^{S_t} \sum_{v=1}^{V_v} U_{v,s} \cdot d}{T \cdot L} \left( \frac{m^2}{s^2} \right)
\]

where \( UD \) -- unsafe density
\( U_{v,s} \left( \frac{m^2}{s^2} \right) \) -- unsafe variable of vehicle \( v \) in simulation step \( s \)
\( V_v \)-- number of vehicles in the link
\( S_t \)-- number of simulation steps within aggregation period
\( d (s) \)-- simulation step duration
\( T (s) \)-- aggregation period duration
\( L (m) \)-- section length

To test \( UD \), a case study was conducted in a 7 kilometer long section which includes two entrance ramps in a peripheral motorway. This section was identified as a high crash location. The morning peak hour (7am-8am) was simulated in AIMSUN and \( UD \) was calculated via an extension module of AIMSUN. It was found that the parameter \( UD \) has several advantages. For instance, it is able to compare the safety level between different links, on a specific section over time, or even different simulation scenarios. On the other hand, there are several limitations for this new measure: first, it only accounts for rear-end collisions. Secondly, it doesn’t consider road environment and human factors such as driver behavior.

The authors also attempted to relate \( UD \) to the number of crashes. The results showed that the evolution of \( UD \) is consistent with the spatial distribution of crashes over three years. However,
a spatial shift was found in the comparison between the evolution of $UD$ and the actual crashes. What causes this spatial shift is still an unsolved question.

**Literature Summary**

The various surrogate measures extracted from existing simulators have been reviewed for analyzing safety problems in the different situations such as intersections, work zones, tunnels, and freeway ramps. However, current micro-simulators bring several difficulties in using such surrogate measures for safety evaluation. First, “crash free” in the simulator breaks the consistency between traffic performance and safety. The performance will decrease as a crash occurs. On the other hand, the bad performance may contribute to a high probability of crashes. Currently, “crash free” simulators cannot fully represent these two-way casual relationships between performance and safety. Secondly, driver behavior models in all simulators are generally based on the same critical parameter: reaction time. Reaction time is either a constant value or allowed limited variation in current models. This isn’t able to account for the always changing reaction time in reality. For instance, the same driver may have a slow reaction due to inattention or tiredness. Furthermore, driver decision-making is a much more complicated process beyond current capabilities of micro-simulators. They cannot interpret aggressive driving, misjudgment, etc. Finally, it is extremely hard to establish a quantitative relationship between surrogate measurements derived from “crash free” simulators and real-world crash rates. Consequently, limitations on safety evaluation are not only due to the surrogate measures but also caused by simulators. A more sophisticated simulation model is needed for safety evaluation.

Our research objective is to use some appropriate surrogate measures to evaluate the effectiveness of proposed geometric improvements at a high crash site in an urban freeway. Modified models or external control of simulated driver behavior may be relevant for our research. No direct reference on these subjects can be found at this moment. Several surrogate measures such as speed variance, speed differential between adjacent lanes, speed differential between leader and follower, deceleration rate, distribution of merge points, distribution of gap acceptance, braking distribution, and unsafe density will be considered. It is also possible to propose and test new surrogate measures for freeway in our research. Some traffic metrics defined in an early study “Identification of Crash Prone Traffic Condition.” (Hourdos, 2004), such as: Traffic Pressure (PT), which is the product of density and speed variance, could be tested in this project. The potential surrogate measures are summarized in the following table.
Table 7.2 Potential surrogate measures in a micro-simulator in our project.

<table>
<thead>
<tr>
<th>Surrogate measures</th>
<th>Applicable situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed variance</td>
<td>Compression waves and weaving area.</td>
</tr>
<tr>
<td>Deceleration rate</td>
<td>Compression wave and stop/slow traffic</td>
</tr>
<tr>
<td>Unsafe density</td>
<td>Merging in the freeway junction</td>
</tr>
<tr>
<td>Speed differential between adjacent lane</td>
<td>Large differential in weaving area</td>
</tr>
<tr>
<td>Speed differential between leader and follower</td>
<td>Compression waves and weaving area</td>
</tr>
<tr>
<td>Braking distribution</td>
<td>Compression waves and weaving area</td>
</tr>
<tr>
<td>Distribution of gap acceptance</td>
<td>Merging area</td>
</tr>
<tr>
<td>Traffic Pressure</td>
<td>Compression waves</td>
</tr>
</tbody>
</table>
References


Kim, Z. and Malik, J. “Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking,” 9th IEEE International Conference on Computer Vision (ICCV '03), 524-531, Nice, France.


Ma, X. and Anderson, I. (2005), "Dynamic car following data collection and noise cancellation based on the Kalman smoothing." IEEE International Conference in Vehicular Electronics and Safety (IEEE ICVES'05), Xi'an, China.

Ma, X. Driver Modeling Based on Computational Intelligence Approaches - Exploration and Modeling of Driver-following Data Collected Using an Instrumented Vehicle, Royal Institute of Technology (KTH) (2006), Stockholm, Sweden.


Ranjitkar, P. (2004), "Experimental analysis of car following dynamics based on RTK GPS data," Proceedings of Infrastructure Planning 2004, Japan Society of Civil Engineers (JSCE), Tokyo, Japan.


Todosiev, E.P., The action point model of the driver vehicle system, Report 202 A-3, Ohio State University, Columbus, OH, 1963.


