# APPLICATION OF COMPUTER VISION TECHNIQUES FOR AUTOMATED ROAD SAFETY ANALYSIS AND TRAFFIC DATA COLLECTION 

by<br>KARIM ALDIN ISMAIL<br>B.Sc., Ain Shams University, 2002<br>M.Sc., Ain Shams University, 2005<br>M.A.Sc., the University of British Columbia, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
in

THE FACULTY OF GRADUATE STUDIES
(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver)

October 2010
© Karim Aldin Ismail, 2010


#### Abstract

Safety and sustainability are the two main themes of this thesis. They are also the two main pillars of a functional transportation system. Recent studies showed that the cost of road collisions in Canada exceeds the cost of traffic congestion by almost tenfold. The reliance on collision statistics alone to enhance road safety is challenged by qualitative and quantitative limitations of collision data. Traffic conflict techniques have been advocated as a proactive and supplementary approach to collision-based road safety analysis. However, the cost of field observation of traffic conflicts coupled with observer subjectivity have inhibited the widespread acceptance of these techniques. This thesis advocates the use of computer vision for conducting automated, resource-efficient, and objective traffic conflict analysis. Video data in this thesis was collected at several national and international locations. Real-world coordinates of road users' positions were extracted by tracking moving features visible on road users from a calibrated camera. Subsequently, road users were classified into pedestrians and non-pedestrians, not differentiating between other road users' classes. Classification was based on automatically-learned and manually-annotated motion patterns. Subsequent to road user tracking, various spatiotemporal proximity measures were implemented to measure the severity of traffic events. The following contributions were achieved in this thesis: $i$ ) codevelopment of a methodology for tracking and classifying road users, ii) development of a methodology for measuring real-world coordinates of road users' positions which appear in video sequences, iii) automated measurement of pedestrian walking speed, $i v$ ) investigation of the effect of different factors on pedestrian walking speed, $v$ ) development and validation of a methodology for automated detection of pedestrian-vehicle conflicts, $v i$ ) investigation of the application of the developed methodology in a before-and-after evaluation of a pedestrian scramble treatment, vii) development of a methodology for aggregating event-level severity measurements into a safety index, viii) development and validation of two methodologies for automated detection of spatial traffic violations. Another contribution of this thesis was the creation of a video library collected from several locations around the world which can significantly aid in future developments in this field.


## Table of Contents

Abstract ..... ii
Table of Contents ..... iii
List of Tables ..... viii
List of Figures ..... X
Acknowledgements ..... XV
Chapter One: Introduction ..... 1
1.1 Challenges ..... 1
1.2 Motivation ..... 7
1.2.1 Growing Importance of Pedestrian Studies ..... 7
1.2.2 Environmental Concerns ..... 8
1.2.3 Demographic Changes ..... 10
1.2.4 Traffic Conflict Techniques ..... 11
1.2.5 Developments in Computer Vision ..... 13
1.3 Problem Statement ..... 14
1.3.1 Problem One: Recovery of Real-world Coordinates ..... 16
1.3.2 Problem Two: Measurement of Walking Speed ..... 16
1.3.3 Problem Three: Severity of Pedestrian-vehicle Conflicts ..... 17
1.3.4 Problem Four: A Before-and-After Context ..... 17
1.3.5 Problem Five: Aggregation of Severity Measurements ..... 18
1.3.6 Problem Six: Automated Detection of Traffic Violations ..... 18
1.4. Contributions ..... 19
1.4.1 High-accuracy Pedestrian Data Collection ..... 20
1.4.2 Automated Analysis of Pedestrian-vehicle Conflicts ..... 22
1.4.3 Video Library ..... 23
1.5. Thesis Structure ..... 24
Chapter Two: Literature Review ..... 26
2.1 Background ..... 26
2.2 Traffic Conflict Techniques ..... 28
2.2.1 Important Milestones ..... 28
2.2.2 Objective Conflict Indicators ..... 34
2.2.3 Challenges to Traffic Conflict Techniques ..... 38
2.3 Computer Vision Developments ..... 43
2.3.1 Computer Vision Developments for Pedestrian Detection ..... 43
2.3.2 Computer Vision Developments for Vehicle Detection ..... 51
2.4 Selected Applications in Transportation Engineering ..... 57
2.4.1 Road Safety Analysis ..... 57
2.4.2 Behavioural Analysis ..... 60
2.4.3 Performance Evaluation. ..... 62
2.4.4 Traffic Performance Monitoring. ..... 63
2.4.5 The Next Generation Simulation (NGSIM) Program ..... 66
2.4.6 Traffic Simulation ..... 68
2.4.7 Other Applications ..... 72
2.5 Privacy Issues ..... 73
Chapter Three: Recovering Real-world Road User Positions ..... 78
3.1 Background ..... 78
3.2 Previous Work ..... 88
3.3 Methodology ..... 89
3.3.1 Camera Model ..... 89
3.3.2 Cost Function ..... 91
3.3.3 Implementation Details ..... 95
3.4 Case Studies ..... 97
3.4.1 Annotation of Calibration Data ..... 97
3.4.2 Validation ..... 98
3.4.3 Effect of Different Cost Function Components ..... 101
3.4.4 Visualization of Results ..... 105
3.5 Conclusions ..... 109
Chapter Four: Automated Pedestrian Data Collection using Computer Vision Techniques ..... 111
4.1 Background ..... 111
4.2 Issues with Pedestrian Data ..... 113
4.3 Previous Work ..... 117
4.3.1 Studies on Walking Speed ..... 117
4.3.2 Techniques of Measuring Walking Speed ..... 120
4.3.3 Challenges of Pedestrian Tracking in Computer Vision ..... 120
4.4 Methodology ..... 121
4.4.1 Camera Parameters ..... 123
4.4.2 Feature Tracking and Grouping ..... 125
4.4.3 High-level Object Processing ..... 126
4.4.4 Manual Input to the Video Analysis System ..... 127
4.5 Case Study ..... 128
4.5.1 Data Collection ..... 128
4.5.2 Data Analysis ..... 129
4.5.3 Validation ..... 133
4.5.4 Discussion ..... 136
4.6 Conclusions ..... 139
Chapter Five: Automated Detection of Pedestrian-vehicle Conflicts. ..... 141
5.1 Background ..... 141
5.2 Previous Work ..... 146
5.2.1 Pedestrian-vehicle Conflicts ..... 146
5.2.2 Severity Conflict Indicators ..... 146
5.2.3 Pedestrian Detection and Tracking ..... 147
5.3 Methodology ..... 148
5.3.1 Camera Calibration ..... 148
5.3.2 Video Formatting ..... 150
5.3.3 Object Tracking ..... 152
5.4 Case Study ..... 153
5.4.1 Site Description and Data Collection ..... 153
5.4.2 Calculation of Conflict Indicators ..... 154
5.5 Validation ..... 157
5.6 Discussion ..... 163
5.7 Conclusions ..... 164
Chapter Six: Automated Analysis of Pedestrian-vehicle Conflicts: A Context for Before-and-After Studies ..... 166
6.1 Background ..... 166
6.2 Previous Work ..... 170
6.2.1 Conflict-based Before-and-After Studies ..... 170
6.2.2 Video-based Road User Detection and Tracking ..... 171
6.3 Methodology ..... 172
6.3.1 Road User Classification. ..... 172
6.3.2 Validation of Tracking Performance ..... 177
6.3.3 Camera Calibration ..... 181
6.3.4 Conflict Indicators ..... 182
6.4 Discussion ..... 187
6.5 Conclusions ..... 190
Chapter Seven: Methodologies for Aggregating Traffic Conflict Indicators ..... 195
7.1 Background ..... 195
7.1.1 Theoretical Preliminaries ..... 195
7.1.2 The Severity Dimension ..... 198
7.1.3 Conflict Indicators as Partial Images ..... 202
7.1.4 Aggregation of Road Safety Cues ..... 209
7.2 Methodology ..... 210
7.2.1 Integration of Conflict Indicators. ..... 210
7.2.2 Mapping to Severity Indices ..... 214
7.2.3 Aggregation of Severity Measurements ..... 220
7.3 Case Study ..... 224
7.3.1 Empirical Independence of Conflict Indicators ..... 224
7.3.2 Results of Different Aggregation Approaches ..... 226
7.3.3 Accounting for both Severity and Frequency ..... 235
7.4 Conclusions ..... 244
Chapter Eight: Automated Detection of Traffic Violation Events ..... 249
8.1 Background ..... 249
8.2 Methodology ..... 252
8.2.1 K-means Clustering using Linear Piecewise Parameterization ..... 253
8.2.2 Violation Detection using LCSS matching ..... 259
8.3 Case Study ..... 262
8.4 Conclusions ..... 273
Chapter Nine: Summary Conclusions and Future Work ..... 275
9.1 Background ..... 275
9.1.1 Chapter One: Introduction ..... 279
9.1.2 Chapter Two: Literature Review ..... 279
9.1.3 Chapter Three: Recovering Real-world Coordinates for Points thatAppear in Video Observations......................................................................... 280
9.1.4 Chapter Four: Automated Measurement of Pedestrian Walking
Speed ..... 282
9.1.5 Chapter Five: Automated Detection of Pedestrian-vehicle Conflicts ..... 284
9.1.6 Chapter Six: Automated Safety Analysis in a Before-and-After
Context. ..... 286
9.1.7 Chapter Seven: Development of an Aggregate Safety Index ..... 288
9.1.8 Chapter Eight: Automated Detection of Traffic Violations ..... 290
9.2 Recommendation for Future Work ..... 292
9.2.1 Validation of Automated Traffic Conflict Analysis ..... 292
9.2.2 Potential Developments of Conflict Indicators ..... 294
9.2.3 Other Extensions to Work Presented in this Thesis ..... 296
References ..... 298
Appendices ..... 313
Appendix A ..... 314
Appendix B ..... 315
Appendix C ..... 321

## List of Tables

Table 1.1 Thesis structure ..... 25
Table 2.1 Drawbacks of conflict indicators ..... 36
Table 2.2 Traffic conflict definitions in the literature ..... 42
Table 3.1 Summary of case studies of camera calibration ..... 85
Table 3.2 RMSE calibration error using Tsai Algorithm (Tsai 1987) and different cost function compositions. The numbers of point correspondence, distance, and angular constraints are in columns $\mathrm{CD}_{p}, \mathrm{CD}_{d}$, and $\mathrm{CD}_{a}$ respectively ..... 98
Table 4.1 Sample of previous studies on pedestrian walking speed ..... 119
Table 4.2 Summary of tracking parameters ..... 131
Table 4.3 Summary of walking speed statistics ..... 136
Table 5.1 Summary of validation results ..... 161
Table 6.1 A comparison between the speed and prototype classifiers. ..... 175
Table 7.1 Severity benchmark values for constructing mapping functions ..... 216
Table 7.2 Correlation coefficients for pairs of conflict indicators with at least one calculable value ..... 228
Table7.3 Correlation coefficients for only pairs of commonly calculable conflict indicators ..... 228
Table 7.4 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user ..... 232
Table 7.5 Summary results for different aggregation strategies for after conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user ..... 233
Table 7.6 Summary results for different aggregation strategies for all conditions. Representative statistics are drawn only from calculable values of positive DST values ..... 234
Table 7.7 Summary results for the two-sample $t$-test for the difference in meanbetween conflict indicators and indices for before and after time conditions. " 1 "means the test for before>after was significant at the 0.05 significance level andconversely after>before for "- 1 ". " 0 " means no significant difference was found.... 235
Table 7.8 Summary results for before and after index values normalized by the total number of tracked road users. Indices representing an event are the maximum of all mapped conflict indicators238

Table 7.9 Summary results for before and after index values normalized by the total number of tracked road users. Indices representing an event are the average of all mapped conflict indicators238

Table 7.10 Summary results for before and after index values normalized by the
product of the volumes of pedestrians and vehicles in millions. Indices for every
event are the maxima of all mapped conflict indicators ..... 243

Table 7.11 Summary results for before and after index values normalized by the
product of the volumes of pedestrians and vehicles in millions. Indices for every
event are the averages of all mapped conflict indicators ..... 243

Table 8.1 Summary of peak performance of different violation detection approaches
273
Table A. 1 Summary of video data ..... 314
Table B. 1 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index at each frame ..... 315
Table B. 2 Summary results for different aggregation strategies for after conditions.Representative statistics are drawn only from calculable values of each indicator orindex at each frame316
Table B. 3 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index at each frame taking into account frequency of observation ..... 317
Table B. 4 Summary results for different aggregation strategies for after conditions.Representative statistics are drawn only from calculable values of each indicator orindex at each frame taking into account frequency of observation318

Table B. 5 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user taking into account frequency of observation 319
Table B. 6 Summary results for different aggregation strategies for after conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user taking into account frequency of observation ..... 320

## List of Figures

Figure 3.1 The difficulty of relying on the automated extraction of road user tracks. Figure a) shows the motion patterns of vehicles at a busy intersection in Chinatown, Oakland-California (sequence OK). Figure b) shows pedestrian motion patterns. 83

Figure 3.2 An illustration of camera calibration issues that arise in urban traffic scenes. Figure a) shows a frame taken from video sequence BR-1 shot at VancouverBritish Columbia. Figure b) shows a sample frame from video sequence K1 of traffic conflicts shot in Kentucky. 86

Figure 3.3 Calibration data for video sequence BR-2. Point correspondences are annotated with their serial numbers. Points marked with red are calculated and points in blue are annotated. The segments in red define the distance conditions. The segments in blue define pairs of lines for angular conditions. Figure a) shows the calibration data (points, and lines) in the image space. Figure b) shows the backprojection of the calibration data to world-space.
Figure 3.4 Examples of reduced camera calibration error due to the inclusion of various cost function components. Figure a) shows the RMSE error of test sets BR1:4 and PG. Figure b) shows the back-projection error in terms of the difference between the true and calculated lengths of 12 line segments in sequence OK. The 12 segments were not used in the calibration. The length difference is normalized by the segments length: Back projection error $=\mathrm{L}_{\text {true }}-\mathrm{L}_{\text {calculated }} / \mathrm{L}_{\text {true }}$ 103
Figure 3.5 The enhancement in calibration by including different cost function components for video sequences K1 and K2. Figure a) shows the back-projection error measured as the difference between the real-world lengths of a total of 20 line segments calculated from two camera settings at K1 and K2. The discrepancy in the lengths of the validation line segments were normalized by each line segment length (average 12.57 m ). Figure b) shows the lengths of the validation line segments for case 5. Refer to Figure 4 for the indication of cases 1:5. 104
Figure 3.6 Reference grid for video sequences BR-2, PG, and OK, overlaid on frames of the video sequence and orthographic images. The grid spacing is 1 m and the height of the vertical reference lines (depicted in blue) is 4.0 m . Sequences BR-1 and BR-3:4 are recorded at the same site (BR) with different fields of view. 106
Figure 3.7 Reference grids for video sequences K1 K2. The non-linear calibration parameters could capture the distortions at the closer sidewalk of sequences K1 and K 2 . The grid spacing is 2.0 m and the height of the displayed vertical line segment (depicted in blue) is 4.0 m .
Figure 3.8 In this traffic safety application, accurate road user tracks are required to measure their temporal and spatial proximity. Left are the back-projected pedestrian and motorist tracks. Right are the CV-based tracks of the interacting road users.

Figures a) and b) show the world and image space of video sequence PG. Figures c) and d) show the world and image space of video sequence OK.
Figure 4.1 Layout of the pedestrian detection and tracking system. The figure shows the five main layers of the system. Depicted also is the data flow among system modules from low-level video data to a database of detected, tracked, and classified road user. 122

Figure 4.2 Pedestrian tracks at site BR-2. Figure a) shows road user tracks in the image space. Figure b) Right figure shows the same tracks projected on an orthographic image. The trajectories are classified by object type (vehicles or pedestrians) and direction. Trajectory clusters 1,2 , and 3 are for pedestrians moving Southeast-Northwest, Northwest- Southeast and Crossing respectively, while cluster 4 is for vehicles. 125
Figure 4.3 Road user trajectories (tracks) transformed to world coordinates. Tracks
of motorized road users are depicted in red. Remaining tracks are color-coded based
on a k-means clustering of pedestrian tracks..................................................... 130
Figure 4.4 The horizontal axis shows the frame number (surrogate for time) and the vertical axis shows the observed speed in $\mathrm{m} / \mathrm{s}$. Pedestrian profiles (in green) exhibit a characteristic rhythm. 132
Figure 4.5 A sample frame from night-time video analysis. Displayed are red bounding boxes around pedestrian objects and their walking speed134

Figure 4.6 The figure shows pedestrian trajectories that crossed through the marked
data collection area. Trajectories are collated and projected to the world image from
different videos with different fields of view and hence may be truncated in different
regions
135

Figure 4.7 Two sets of check-lines for collecting manual observations of average walking speeds. The spacing between the upper set of check-lines (crosswalk) is 3.61 m and the spacing between the bottom set of check-lines (long segment) is 11.58 m .

Figure 4.8 Figure a) shows validation of walking speed measurements at day-time. Horizontal axis depicts walking speed based on the time interval required to walk between two check lines. Vertical axis depicts the average walking speed within the same time interval based on automated pedestrian tracking. Figure b) shows validation of walking speed measurements at night-time conditions.
Figure 4.9 Figure a) Walking speed frequency distribution for pedestrians moving through the data collection area shown in Figure 4.4 across Robson St. Figure b) walking speed frequency distribution for pedestrians moving from Southeast to Northwest through corresponding data collection areas on both sidewalks of Robson Street. ....................................................................................................................... 138
Figure 5.1 The 22 points used to estimate the camera calibration are displayed on a video frame in Figure a) and on an orthographic satellite image of the traffic scene in

Figure b). Bulleted points ( $\bullet$ ) are manually annotated and $x$-shaped points ( x ) are projections of annotated points using the estimated camera parameters.
Figure 5.2 A sample of pedestrian tracks is projected on an orthographic satellite image of the traffic scene. Vehicle tracks are depicted in red and pedestrian tracks are in black.

Figure 5.3 Conflict indicators for a sample traffic event. The left figure describes the traffic event shown in figure 5.4a. The right figure describes the traffic event shown in figure 5.4b.
Figure 5.4 Sample of automatically detected important events along with road users' trajectories. The numbers under each image are respectively the min TTC (seconds), PET (seconds), maximum $\operatorname{DST}\left(\mathrm{m} / \mathrm{s}^{2}\right)$, and $\min$ GT ( s ). In the images, the road user speed is indicated in $\mathrm{m} / \mathrm{s}$. 162
Figure 6.1 Road user prototypes for the before-and-after scramble phase. Figure a) shows the pre-scramble vehicle prototypes(pre-scramble-veh). Figures b), c), and d) show pre-scramble-ped, post-scramble-veh, and post-scramble-ped, respectively. The color coding is the result of a k-means clustering in 4 classes based on the prevalent prototype direction. 176
Figure 6.2 Receiver Operating Characteristic (ROC) Curve for the speed and prototype classifier (for the smoothed max speed classifier, the road user speed is smoothed with a moving average filter). The threshold for the speed classifiers is $3 \mathrm{~m} / \mathrm{s}$. 177

Figure 6.3 Plot of the cost function with respect to ( $D_{\text {connection }}, D_{\text {segmentation }}$ )........... 179

$$
\begin{aligned}
& \text { Figure 6.4 Sample frames from validation results. The number of missed detections } \\
& \text { is } 3 / 32 \text { with } 29 \text { false detections mainly due to over-segmentation. Figure a) shows a } \\
& \text { sample frame from a post-scramble sequence with labelled pedestrians. Figure b) } \\
& \text { shows the pedestrians tracked in the same frame using the optimized tracking } \\
& \text { parameters. The bicyclist annotated with a box in Figure b) is correctly identified as a } \\
& \text { non-pedestrian (given a screen label 'ca')................................................... } 180
\end{aligned}
$$

Figure 6.5 Calibration of the video camera. Figures a) and b) show the calibration features. Points are labelled. Segments in red are distance constraints. Segments in blue constitute angular constraints. The inferred camera location is marked. Figures c) and d) show the projection of a reference grid from the world space in c) to image space in d). World images are taken from Google Maps.
Figure 6.6 Conflict clustering. Figure a) shows an interaction between a pedestrian and an over-segmented vehicle (tracked twice, object 5638 on the front side and the other 5639 encompasses its horizontal projection). The spacing between these vehicle objects and the pedestrian at minimum TTC and GT are 2.18 m and 1.53 m respectively. Both are below a spacing threshold of 3 m and are therefore grouped. Figure b) shows an illustration of the graph implementation.

## Figure 6.7 Sample frames with automated road user tracks. The captions display "Event" the event order in the list of potential interactions, "objects" the numbers of the interacting objects, and the indicated conflict indicators. <br> 192

Figure 6.8 Before-and-after spatial distribution of traffic conflicts. A conflict
positions is selected as the position at which the motorist was separated by either a
minimum Gap Time (GT) or minimum Time to Collision (TTC). Figure a) shows the
before spatial distribution of conflict locations based on min GT. Figure b) shows the
after distribution of conflict positions based on min GT. Figure c) shows the before
distribution of motorist position at min TTC. Figure d) shows the after distribution of
conflict positions based on min TTC.

Figure 6.9 Distribution of different conflict indicators values for before and after scramble phase. Analyzed video durations are 2 hours before and 2 hours after. |PET| and $|\mathrm{GT}|$ are the moduli (unsigned) values of the Post Encroachment Time and Gap Time conflict indicator. 194

Figure 7.1 Two events of only subtle difference in context and comparable values of minimum Time to Collision (min TTC). The subtle difference in context however entails significantly different severity.206
Figure 7.2 A depiction of two mappings from conflict indicators to severity index. Shown also are the parameters for function mapping 1. Mapping parameters (p1:p5) are shown in the legend that were collected from benchmarks in the literature. For example $\mathrm{p} 1=8.0$. ..... 219

Figure 7.3 A schematic for different aggregation approaches. Note that datapoints are deinfed over ordered pair of road users to avoid redundant recording of the same traffic event.221
Figure 7.4 Severity index distributions for before and after conditions. Function mapping was used. Maximum indices were selected for every frame (upper row) and road user (bottom row) ..... 239
Figure 7.5 Severity index distributions for before and after conditions. Function mapping was used. Average indices were selected for every frame (upper row) and road user (bottom row) ..... 240

Figure 7.6 Conflict indicator and index distributions for before and after conditions. Maximum indicators and indices were selected for every road user. Distributions are normalized by the total number of exposure events241

Figure 7.7 Conflict indicator and index distributions for before and after conditions. Average indicators and indices were selected for every road user. Distributions are normalized by the total number of exposure events 242

Figure 8.1 A vehicular track approximated by three different piecewise linear models. The vertical axis shows the cosine of the instantaneous azimuth of road users
and the horizontal axis shows the moment of measurement relative to the total life
span of the track. .................................................................................................. 256
Figure 8.2 Sample tracks approximated to a fourth-degree piecewise linear model (a sequence of four linear segments). Vertical axes show the cosine of the instantaneous azimuth of road users and horizontal axes show the moment of measurement relative to the total life span of each track. 257
Figure 8.3 Sample grids project from world space (right figure) to the image space (left figure). ..... 263
Figure 8.4 Violation tracks (left figure) and a sample of normal tracks (right figure) during 2 hour observations at Darwaza intersection ..... 264
Figure 8.5 A superimposition of the 183 normal movement prototypes used for LCSS-based violation detection ..... 267
Figure 8.6 Performance of LCSS-based classification using a range of values for matching distance $\epsilon$ in meters and similarity threshold $\sigma$. The directional cosine threshold was set to 0.9 . ..... 268
Figure 8.7 Evident instability of detection results of k-means clustering ..... 271
Figure 8.8 The receiver operating characteristic curve for the two violation detection approaches presented in this chapter. K-mean clustering was conducted on a full sample size of 986 tracks. ..... 272

Figure 8.9 The receiver operating characteristic curve for the two violation detection approaches presented in this chapter. K-mean clustering was conducted on a reduced sample size of 448 tracks.272
Figure 9.1 A hypothetical case of two road users $U_{1}$ and $U_{2}$ predicted to follow two trajectories $\mathrm{H}_{12}$ and $\mathrm{H}_{22}$ that lead to a potential collision. Predicted positions of road users are approximated using a bivariate Gaussian. ..... 296

## Acknowledgements

I am greatly thankful to Dr. Tarek Sayed for being the dedicated supervisor and the keen visionary during the six years I spent at the University of British Columbia. I owe much of my scholarship development and successful activities to his inspiration, motivation, and timely guidance. I would also like to thank my supervising committee, Dr. Robert Millar, Dr. Thomas Froese, and Dr. Mohamed Wahba for their review of the thesis. I especially note Dr. Millar for his dedication to the cause of this thesis. I am especially grateful to Dr. Wahba for the thorough review of this thesis and for poring over its technical details. I would like to thank Dr. Yinhai Wang, Dr. John Meech, and Dr. Gregory Lawrence for their insightful comments.

I was blessed by the professional mentorship of Clark Lim whose wisdom, principled approach to engineering practice, unwavering belief in this research, and assistance with video data collection played a key role in the development and the completion of this thesis. I was also very fortunate to work with Dr. Nicolas Saunier. I credit an important part of the development of my academic character to the period spanning three years during which I worked and co-authored with Dr. Saunier on the subject of this thesis.

The video analysis system used in this thesis is based on the commendable work of the hundreds of dedicated open-source developers of the OpenCV project (Bradski \& Pisarevsky 2000). Dr. Saunier, then a post-doctoral fellow at the University of British Columbia, is valuably credited for the development of several modules additional to libraries available in the OpenCV project. Notable of which are the C++ implementations of the following algorithms: Feature Grouping (Beymer et al. 1997), Longest Common Subsequence (LCSS) similarity measure (Vlachos, Kollios \& Gunopulos 2005) (recoded from a Matlab implementation obtained from the previous algorithm authors), Prototype Learning (Saunier, Sayed \& Lim 2007), and Tracking Performance Evaluation (Saunier, Sayed \& Ismail 2009).

I would like to thank Dr. Mohamed Zaki for the whole-hearted technical and editorial assistance at the final stage of this thesis. In addition, the vehicle tracks analyzed in Chapter 9 were produced with help from Dr. Zaki, then a post-doctoral fellow at the University of British Columbia. Dr. Mohamed Zaki also implemented in the C++ language the technique proposed in Chapter 9 by the author for LCSS-based violation detection.

I am grateful to the contribution of the following persons in regard to video data collection and manual annotation of road user data: Sam Young and staff of the MMM Group (Vancouver office) for supporting the video survey, Hsu Hua Lu for assisting in the manual annotation of the video sequences, Varun Ramakrishna for video annotation, Jenna Hua and Dr. David Ragland for granting us access to the pedestrian scramble video data. I am thankful for Jarvis Autey for proof-reading parts of this thesis. I am thankful for the anonymous reviewers of the published work on which this thesis is based.

I would like to acknowledge my indebtedness to my late father Samir and to my mother. Their unabated hard work, their adherence to a standard of excellence, and their utmost dedication to my upbringing have a more profound effect on my personal and professional development than will ever be apparent. I am deeply thankful to my wife Mai and my daughter Hana for being the dream I have always had, for forgiving the singular selfishness behind the completion of my thesis, and for giving me new dreams to pursue.

I am ultimately grateful to God for granting me the knowledge, the intellect, the time, and the providence that I often times take for granted.

## Introduction

### 1.1 Challenges

In the mobility-oriented transportation systems, road users are in an eternal state of motion, seeking one trip destination after another. For this mobility to be affordable, road users need to be ensured an adequate level of safety. Mobility and safety are therefore widely regarded as valid performance measures of a transportation system.

The incidence of road collision is aptly described as a global epidemic of staggering yet often overlooked consequences. Road collisions are predicted
to climb from the $10^{\text {th }}$ to the $8^{\text {th }}$ most common cause of death by the year 2030 (Mathers \& Loncar 2005). The global number of fatal road collisions was approximately 1.3 million in 2004 rising from 990,000 in 1990 (WHO 2004). Moreover, road collisions are the cause of tremendous social and economic losses. The global economic cost of road collisions is estimated as \$US 500 billion per year (UN 2003). A striking imbalance in the economic burden of road collisions occurs between low-income and high-income countries. On a worldwide level, the majority of fatal road collisions occur in low-income countries (90\%) (Suri \& Parr 2004). Conversely, for complex reasons, the larger share of economic cost of road collisions (approximately $85 \%$ ) is borne by high-income countries. In either case, road collisions afflict low-income countries with human losses and high-income countries with heightened cost of health care and insurance claims.

Canada is a highly developed state, which despite significant investment in road infrastructure, is still struggling with the incidence of road collision. A recent study by Transport Canada estimates that the annual cost of road collisions to the Canadian economy, including health care, environmental damage, lost productivity, and induced traffic congestion, is \$CDN 62.7 billion (Vodden et al. 2007). This represents an enormous $5 \%$ of the Canadian Gross Domestic Product. The cost of road collisions is far more than the approximate cost of \$CDN 3.7 billion/year due to the impact of traffic congestion on Canadian roads (Cannon 2006). The collective cost of road collision impacts all types of road users with various proportions. However, the impact on non-motorized and sustainable modes of travel is far more consequential than is apparent.

Road collisions represent an indirect and far reaching threat to meeting grassroots demand for a sustainable transportation system. While nonmotorized modes of travel are groomed as key alternatives to motor vehicles in order to improve the sustainability of the transportation system, they suffer from an elevated risk of collision involvement. This exceptional risk level stems from the fact that key non-motorized modes of travel such as biking and walking are engaged by the most physically vulnerable of road users. Road collisions involving non-motorized traffic are highly injurious and physically damaging. Ominously, the physical threat to the arguably most important mode of non-motorized travel, walking, is particularly severe.

In the developing world, the vulnerability of pedestrians and the little attention paid by policymakers to these modes of travel render the situation especially dangerous. In a review of 38 studies from developing countries, 27 studies ranked the frequency of pedestrian fatalities as the highest among all modes; accounting for $41 \%$ to $75 \%$ of all fatalities (Odero, Garner \& Zwi 1997). The problem of pedestrian vulnerability is also present in developed countries. Approximately $22 \%$ of fatal road collisions in Canada and $30 \%$ of fatal road collisions in British Columbia involve vulnerable road users; respectively $13 \%$ and $15 \%$ of which are pedestrians (ICBC 2006). As the global society is becoming more aware of the importance of non-motorized modes of travel, the threat that previous safety issues create to building a sustainable transportation system is expected to receive rightful attention from policy makers, researchers and practitioners. However, the growing awareness of road collisions has yet to foster an insightful understanding of the nature of road collisions and yet to create methods of analysis that meet the challenge.

Mainstream approaches to road safety analysis can be described as: reactive and collision-based. The traditional reactive approach represents the application of safety countermeasures to mitigate emergent safety problems or to evaluate new safety treatments (de-Leur \& Sayed 2003). Collision-based approaches to road safety rely on data drawn mainly from collision records, police reports, and insurance claims. The traditional reactive approach to safety based on collision data is challenged on several accounts.

The paradigm of reactive road safety analysis permits the analyst and the decision maker to wait for collisions to occur in order to conduct safety analysis or to devise countermeasures. According to reactive road safety analysis, there is little ability for predicting road safety issues and preventing their occurrence before the materialization of the induced social cost. Moreover, in order to evaluate the effectiveness of safety programs in remedying existing problems, adequate before-and-after observational periods of observation have to be allowed in order to prove a reduction trend, or the absence thereof. During extended observational periods, road collisions continue to occur. To address the shortcomings of reactive road safety analysis, the proactive approach to road safety analysis has been recently advocated (de-Leur \& Sayed 2003).

Proactive safety analysis seeks to diagnose safety problems and to identify suitable countermeasures before collisions take place. In achieving this objective, safety analysis draws on an array of historical safety record, evaluation studies, and performance records of control sites. There have been tremendous advances in proactive road safety analysis based on collision
prediction models (Harkey et al. 2008). This stream of research strives to predict collisions based on historical statistical associations with various attributes of the road system. In these models, road collisions are adopted as the main safety measure and also the main data type.

Despite the prominence of collisions-based proactive safety analysis, there are several limitations to the reliance on collision data in conducing road safety analysis (Chin \& Quek 1997):
a. Cost. Road collisions are by far the most costly and most dangerous events that take place on roads. However, in collision-based road safety analysis, road collisions are themselves the units of data. In that, conducting collision-based analysis, especially in experimental or trial-and-error settings, entails tremendous social and economic costs.
b. Attribution. The information obtained by police reports and interviews often does not allow the attribution of road collisions to a definitive set of causes. It is sometimes difficult to pinpoint the failure mechanism that leads to collision based on interviews, witnesses, and post hoc site investigations. In that, the safety analyst is often required to remedy or prevent a set of events whose causes are not precisely known.
c. Data Volume. Despite the enormous social burden of road collisions, the frequency of road collisions, especially in disaggregate data form, relative to other traffic events is low. Drawing statistically stable and significant inferences based on inherently noisy and rare data is challenging and sometime subject to divergent interpretations.
d. Data Quality. Collision reporting is mainly based on post hoc investigations such as witness accounts and site observations. The process is fundamentally deductive and subjective. Collision records are often incomplete and lack details. Furthermore, collision reporting is generally biased toward highly damaging incidents, while noninjurious collisions may go unreported.
e. Ethical Concern. In order to properly conduct fundamental tasks of road safety analysis, such as safety diagnosis, collisions have to have occurred and be have been recorded over an adequately long period. For example, in before-and-after studies, observation of road collisions may extend for a period of 1-3 years in order to draw stable inferences. For the identification of hazardous locations to be proper, several years of road collision observations have to be available. Previous limitations of collision data give rise to the paradoxical situation in which the safety analyst, for the sake of methodological correctness, strives to observe events that ought to be prevented.

Challenges of the reliance on road collision data are even more pronounced in the study of pedestrian safety. Pedestrian-involved collisions are more injurious and less frequent. Moreover, there is a traditional bias in transportation studies in favour of motorized modes of travel. This bias is most evident in the design and evaluation of transportation systems (Milam \& Mitchell 2008). Availability of pedestrian data has been commonly identified as a major challenge in the practice of pedestrian safety analysis (Hoogendorn, Daamen \& Bovy 2003). Examples of data needs include pedestrian volume and measures of exposure to collision risk which are often
expensive and time consuming to obtain (Pulugurtha \& Repaka 2008). Surrogates or statistical predictors of these types of data are often used in practice, e.g., (Shankar et al. 2003) (Greene-Roesel, Diógenes \& Ragland 2007). While in practice there are developed technologies and proven applications for motorized traffic counts, this is not the case for pedestrian traffic (GreeneRoesel et al. 2008). Pedestrians move in a less organized fashion, at higher densities, and in more complex and constrained spaces than vehicular traffic. Thus existing issues with data availability are, in the case of pedestrian safety, compounded due to the lack of reliable automated data collection methods.

### 1.2 Motivation

The growing momentum for building more sustainable transportation systems has provided an important focus of this thesis work. The second focus of this thesis was driven by the potential of computer vision techniques for solving well-entrenched problems in road safety analysis. Sustainability and safety constitute the main pillars of this thesis. The following sections provide a detailed description of the motivation behind this thesis.

### 1.2.1 Growing Importance of Pedestrian Studies

Walking is the most basic means of traveling and a key non-motorized mode of travel. Pedestrian movement is a critical component in a multimodal transport network that connects different modes of travel and interfaces with external activity areas. One of the desirable characteristics of walking is that building improvements to pedestrian facilities reflects on the overall connectivity and accessibility of the transport network. Creating pedestrian-
friendly environments is particularly beneficial to modes of travel whose level of functionality depends on attracting walking trips, e.g., public transit and cycling. Walking is a key activity of sustainable, healthy, clean, resourceefficient and liveable urban environment for which pedestrians are described as "lifeblood" (AASHTO 2001). New urban planning concepts redefine the function and mode-assignment of streets by emphasizing walkability and changing industry standards and professional practice in order to accommodate the pedestrian as a key road user (Greenberg 2005).

### 1.2.2 Environmental Concerns

The growing interest in pedestrian studies is driven also by popular and political awareness of environmental issues; especially the development of a sustainable transportation system. Despite the rhetoric, significant effort is yet to be expended by engineers and decision makers to build such system. Currently, the passenger car is the dominant mode of travel in Canada and the world. There is little evidence that it will be replaced by any other mode in the foreseen future. Approximately, one billion motor vehicles are operated in the global transportation system. In two decades, this number is expected to double (Sperling \& Gordon 2008). The current number of light motor vehicles registered in Canada is expected to double within less than two decades ${ }^{1}$; outpacing population growth by a wide margin (Doiron 2008). This unchecked growth in motorized modes of travel is accompanied by increasing release of air pollutants.

[^0]Despite the global awareness and the commitment to reducing global greenhouse gas (GHG) emissions, the reality belies expectations. In order to avoid global climate change, GHG emissions need to be reduced by $50-80 \%$ by 2050 (WHO-UNEP 2007). However, transportation-induced GHG emissions have more than doubled since 1970s, faster than any other energy sector. A recent review by Environment Canada shows that the outlook for Canada is not too promising (EnvCanada 2008).

Canada's GHG emissions are 29\% higher than Kyoto requirements. There was a $35 \%$ increase in GHG emission induced by road transportation from 19902006. Transportation is the second largest emission category in 2006, contributing approximately $27 \%$ of the total GHG emission; two thirds of which being from road transport (EnvCanada 2008). Emissions from energy consumed for personal transportation rose by $32 \%$ from 1990 to 2006. There was also a $37 \%$ increase in the size of the transport fleet during the period from 1990 to 2006 (EnvCanada 2008). Similarly, a review of the US National Transportation Statistics in 2008 shows that passenger vehicles produced approximately half of all carbon monoxide emissions and one third of all carbon dioxide emissions and nitrogen oxides in 2008 (USDOT 2008).

Despite the evidence in the literature that corroborates the importance of nonmotorized traffic, these modes of travel are in general overlooked, undervalued, and understudied in contrast with vehicular traffic. For example, current trip counts capture $16-33 \%$ of actual non-motorized trips (Litman 2003). Furthermore, collecting reliable non-motorized traffic information is especially challenging (Weinstein \& Schimek 2005). Planning
for pedestrian facilities and demand forecast of walking activities are areas of research that are yet to be developed to a level that matches vehicular traffic (Pulugurtha \& Repaka 2008). In general, there is poor integration of pedestrians into current transportation networks; especially what relates to interlinking of different activity areas (James \& Walton 2000). For example, vehicular traffic is traditionally the main focus of level of service improvements, with little attention to other modes that share the same segment of the transportation system (Milam \& Mitchell 2008). The trade-off between improving the level of service for motorized traffic and the impact on non-motorized transport is often either ignored or cursorily studied in the current practice. Significant challenges are poised to test the creativity of researchers and practitioners in order to better understand, analyze, and accommodate non-motorized modes of travel.

### 1.2.3 Demographic Changes

Aside from environmental and political motivations and aspirations for building a sustainable transportation system, the emerging research focus on pedestrians comes as a response to demographic changes in the general population. Along with other developed countries, the population of Canada is aging. The percentage of seniors (65+) in Canada increased from $13 \%$ in 2001 to $13.7 \%$ in 2006. This percentage is projected to reach $23-35 \%$ in 2031 and $25-30 \%$ by 2056 (Martel \& Malenfant 2006). Seniors in British Columbia represent $14.6 \%$ of the provincial population in 2006 making the province one the oldest in Canada (Martel \& Malenfant 2006). Similar national trends are observed in the US although the population is slightly younger (Shrestha
2006). The effect of demographic changes on the design of pedestrian facilities is palpable. Older pedestrians have different ambulatory characteristics, speed of reflexes, and sensory acuity from the general population.

Walking speed of older pedestrians has been the subject of a number of recent studies. Recent releases of standard design guides, e.g., MUTCD-Canada (Transportation Association of Canada 2002), are in the process of adopting slower standard walking speed in consideration of particularities of the elderly pedestrian. Further studies are still required to capture the differences among senior subgroups. Some studies suggest that this age group is not homogenous as assumed by past studies of walking speed (Stollof, McGee \& Eccles 2007). The development of methods of analysis capable of accurate and efficient measurement of individual pedestrian dynamics is critically needed to capture the changing characteristics of the population of system users.

The motivation presented thus far relates to the study of pedestrians from a broad perspective. An important focus in this thesis is on pedestrians from a road-safety perspective. The focus on road safety is driven by limitations in the state-of-the-art and recent developments in the realm of computer vision.

### 1.2.4 Traffic Conflict Techniques

Traffic conflict techniques have been advocated as an alternative or a supplementary approach to collision-based road safety analysis. Traffic conflicts or near-misses can be viewed as precursors to road collisions. The incidence of traffic conflicts can act as a surrogate safety measure. The definition of a traffic conflict has evolved since its first proposition by Perkins \& Harris (1968). A widely acceptable conceptual definition of a traffic conflict
is: " an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remained unchanged" (Amundsen \& Hydén 1977). Traffic conflicts possess important advantage over road collisions for the purpose of road safety analysis. Traffic conflicts are more frequent, much less costly than road collisions. Moreover, the observation and analysis of the positions of road users involved in traffic conflicts may provide insight into the failure mechanisms that leads to collisions.

Despite the well-recognized advantages of traffic conflict techniques, they suffer from: the inter- and intra-observer subjectivity and the costliness of conducting traffic conflict surveys. It is well recognized that trained observers perceive the severity of a traffic encounter in different ways. Consequently, they may disagree at whether a specific traffic event should be classified as a traffic conflict. Despite initial enthusiasm as well as decades of extensive practice, the subjectivity of human observers and the empirical evidence against the validity of traffic conflict techniques have discounted from their appeal. Furthermore, the significant cost required to train field observers and to institute traffic conflict surveys have been a major drawback.

Given the shortcomings of collision-based road safety analysis and observerbased traffic conflict techniques, the demand for a new paradigm for road safety analysis is building up. Road safety analysis may benefit if more frequent, less random, and less costly types of events are used in place of road collisions. In addition, by relying on more capable methods of analysis and technologies, the severity of traffic events may be measured in a
quantitative fashion. More sophisticated techniques of analysis are needed to enable cost- and resource-efficient processing of extended observational periods. As is demonstrated in several applications and methods of analysis in this thesis, traditional drawbacks of traffic conflict techniques can be eliminated by the informed adoption and application of computer vision techniques.

### 1.2.5 Developments in Computer Vision

Computer Vision is defined as "... the enterprise of automating and integrating a wide range of processes and representations used for vision perception ... such as image processing, statistical pattern classification, geometric modelling and cognitive processing" (Ballard \& Brown 1982). Computer vision techniques rely on video sensors as the main source of data. Video sensors are arguably one of the most powerful methods for the collection of road user positional data. Video data is rich in detail, recording devices are becoming less expensive, and automated analysis is possible using techniques developed in the field of computer vision. Furthermore, many jurisdictions are installing video cameras at traffic intersections for monitoring purpose. The ultimate goal of adopting computer vision techniques is the automated extraction of road users' positions as they navigate the field of view of video sensors.

Numerous road safety measures can be obtained from analyzing road user positions. Extracting road user tracks from video sequences enables positional analysis at a much higher spatial and temporal resolution than current techniques available in practice. Conducting manual positional analysis at a comparable precision using manual observations is
tremendously time consuming. By informed application of computer vision techniques, automated and precise positional measurement is possible in a time- and resource-efficient way. In an analogy to the well-established research stream that tries to confer "intelligence" to transportation systems, the use of computer vision techniques can be seen as an attempt to equip transportation systems with a "visual sense" 2 .

There has been accelerating development in computer vision techniques. It is becoming increasingly feasible to automatically detect, track, and classify road users by the automated analysis of video data. Many proven applications and commercial products have emerged in response to increasing demand from practitioners and researchers. The adoption of computer vision techniques in road safety applications is especially appealing. Automated road user detection and tracking can empower the traditional weaknesses of traffic conflict techniques. The judicious application of computer vision techniques enables the processing of video data in an automated and objective fashion.

### 1.3 Problem Statement

The study of pedestrians draws its importance from the physical vulnerability of this type of road users and the key role that walking activities play in a sustainable transportation system. Often, the road safety analyst is humbled by the challenges of studying pedestrian safety. One part of this challenge comes from intrinsic problems with the reliance on collision data. The other

[^1]part comes from the dearth and cost of pedestrian exposure data. Arguments that support the adoption of traffic conflicts techniques are of particular relevance and find more grounds in the study of pedestrian safety. Traffic conflict techniques are concerned with observing and evaluating the frequency and severity of traffic conflicts by a team of trained observers (Perkins \& Harris 1968). Issues of the current level of development of traffic conflict analysis arise as well in the study of pedestrian-vehicle conflicts. The study of pedestrian-vehicle conflicts cannot be carried out by direct adoption of existing techniques developed specifically for motorized modes. Rather, existing issues and general shortcomings in traffic conflict techniques have to be addressed and resolved before direct adaption is feasible.

Other streams of research include data collection of pedestrian movement to support behavioural studies, calibration and validation of simulation models, and general purpose surveys. It is common in pedestrian studies in the transportation literature to find explicit mention of problems or research challenges due to data limitations. However, the majority of existing techniques for automated data collection are developed for motorized traffic. To overcome these issues, video sensors have been used in practice to observe pedestrian movement. Subsequent to which manual in-office analysis is conducted by human observers. A key advantage of video sensors is that the collected data is rich enough to support automated analysis. The pursuance of this objective draws on the extensive literature of computer vision, in which computer systems are developed with the aim of automatically interpreting video data. A significant part of the research presented in this
thesis advocates the adoption of computer vision techniques for the detection and tracking of pedestrians and other interacting road users.

The issues and practical needs mentioned thus far drive the focus of this thesis. The research problems tackled in this thesis guide a number of developments toward a new paradigm of road safety analysis based on automated and objective positional measurements. This thesis presents a number of solutions for the following research problems:

### 1.3.1 Problem One: Recovery of Real-world Coordinates

Video sensors are adopted as the principal method of data collection in this thesis. Prerequisite to conducting precise analysis of road user positions is to recover the real-world coordinates of road users as they appear in video sequences. This practical need gives rise to the following research problem:

Develop a technique to map points in the image space to real-world coordinates. This technique should be accurate enough for positioning slow-moving road users such as pedestrians. The technique should be robust to measurement errors and to incomplete calibration data. Moreover, the technique should be reliable even if the video camera used for observation is inaccessible and its setting in the monitored scene is unknown. Finally, this technique should provide robust functionality when the video data is of low quality or suffers from excessive distortion.

### 1.3.2 Problem Two: Measurement of Walking Speed

The importance of developing new techniques for pedestrian data collection is paramount. Computer vision techniques offer an appealing solution to demands for more efficient and accurate data collection methods. Walking
speed has been the subject of recent studies in response to changes in the movement characteristics of pedestrians at signalized crosswalks. Despite the practical need that is likely to persist in the future, there was no evidence of the adoption of computer vision techniques to address this data need. This practical demand lays the ground for the following research problem:

Investigate the potential of applying computer vision techniques in accurately measuring pedestrian walking speed in open and crowded scenes. The accuracy of walking speed measurements should support subsequent developments in the thesis.

### 1.3.3 Problem Three: Severity of Pedestrian-vehicle Conflicts

By recovering real-world positions of pedestrians as well as vehicles, it is required to measure the severity of traffic events as described in the following research problem statement:

Develop necessary methods of analysis for the automated and objective measurement of the severity of pedestrian-vehicle conflicts. In demonstrating this application, video data, preferably needs to be collected for extended time, at a pedestrian facility that suffers from pedestrian safety concerns.

### 1.3.4 Problem Four: A Before-and-After Context

One of the main shortcomings of collision-based road safety analysis is the extended observational period required to observe stable trends. For example, typical observational periods for the before-and-after evaluation of safety treatments are 1-3 years. A key advantage of the method of analysis developed in addressing problem three is reducing this observational period
from a time span of 1-3 years to few weeks. This key practical advantage gives rise the following research problem:

Investigate the application of computer vision techniques coupled with various methods of analysis developed in earlier problems in evaluating pedestrian safety treatments. Problems such as automated pedestrian classification in wide-area pedestrian movement must be solved.

### 1.3.5 Problem Five: Aggregation of Severity Measurements

Various objective severity measures and conflict indicators are hypothesized to be of different and sometimes of independent nature. Each objective severity measure provides a cue for the underlying level of safety. In current practice, the integration of various road safety cues is more of an art than science. The following research problem is designed to address shortcomings in practice:

Develop a quantitative methodology for the integration of various objective measures of pedestrian-vehicle conflicts.

### 1.3.6 Problem Six: Automated Detection of Traffic Violations

A major focus of this thesis is on the use of traffic conflict data for measuring road safety. It can be argued that traffic conflicts may involve some degree of nonconformity or violation to traffic regulations. However, not all road user violations result in traffic conflicts. It is possible that the observation and analysis of road user violations may help probe the underlying and unobservable safety level ${ }^{3}$. Admittedly, video observation for the purpose of

[^2]traffic conflict analysis, although superior to the legacy observer-based field surveys, is limited in time. During this finite observational window, it may not be possible to observe a representative sample of traffic conflicts. Therefore, traffic violations may serve as a supplementary type of data for traffic conflicts. This gives rise to the following research problem:

Develop a technique for the automated observation of traffic violations. It is also required to contrast the developed technique with a standard approach in the literature for solving this research problem.

### 1.4. Contributions

This thesis documents a body of knowledge on the subjects of pedestrian data collection and road safety analysis. The contributions of this thesis consist of several advances to the state-of-the-art of road safety analysis and traffic data collection. The thesis also advocates an approach for road safety analysis that empowers the weaknesses of existing methods used both in research and in practice. The approach for safety analysis advocated and advanced in this thesis can be rightfully called a new paradigm of road safety analysis. The automated positional analysis of pedestrians and vehicles presented in this thesis gives deeper insight into road user interactions that constitutes an important advancement from the classical reliance on aggregate statistical associations. The type of data analyzed in this thesis, high-precision automatically extracted road user positions, has only been available in the
discipline of road safety for few years ${ }^{4}$. Besides a handful of other recent studies, the work presented in this thesis represents an attempt to advance the frontier of knowledge on road safety analysis. Furthermore, attempts undertaken in this thesis to tackle previous research problems provide benefits and implications that extend beyond the immediate research objectives of this thesis. An important byproduct of this thesis is an extensive video library of pedestrian movement. This video library created in this effort has provided service for research outside the scope of this thesis (Khanloo et al. 2010). The following sections provide a detailed description of the thesis contributions.

### 1.4.1 High-accuracy Pedestrian Data Collection

a. The development of a methodology for recovering real-world coordinates of points that appear in video sequences. The methodology is based on the idea of utilizing a set of geometric features that are typically available in the traffic scene. The methodology also involves a novel composition of an error term that reflects the discrepancy between features directly observed against their projection.
b. This methodology was implemented in the MATLAB language (Mathworks 2010) and was successfully used in all video analysis undertaken in this thesis (six different scenes) as well as other applications outside the scope of this thesis (nine different scenes). The

[^3]accuracy of estimates is superior to current practical requirements for road users tracking.
c. The successful application of computer vision techniques in measuring pedestrian walking speed in the Downtown area of Vancouver, British Columbia. The technique used is a type of feature-based tracking (as shorthand, it will be referred to as feature-based tracking). This study was to the best of the author's knowledge unique in terms of the large sample size, high pedestrian crowd density, and conducting automated walking speed measurements at night.
d. The video data analyzed featured the movement of pedestrian crowd drawn to a local event; a fireworks show. The automated measurements were positively validated. The accuracy of automated measurement of walking speed was contrasted with manual observations against which it yielded a satisfactory agreement.
e. Average walking speed was obtained from a relatively large sample of pedestrian walking speed measurements. To the best of the author's knowledge, this was the largest sample size analyzed among similar studies in the literature. The aggregate estimate of average walking speed serves as a key design variable in crowd management, traffic signal design, and design of pedestrian facilities.
f. Statistical analysis of the measurements was conducted in order to investigate the variance of walking speed under different conditions such as time of the day, type of pedestrian facility, movement direction, and longitudinal pavement slope. Results of this analysis provide
insight into the considerations required for the design of pedestrian facilities under different operational and physical conditions.

### 1.4.2 Automated Analysis of Pedestrian-vehicle Conflicts

a. Pedestrian tracking in open and mixed-use intersections is particularly challenging. The computer vision applications presented in this thesis prove the feasibility of adopting feature-based tracking in extracting tracks of pedestrians and vehicles involved in events of traffic conflicts.
b. The development of various methods of analysis for the automated and objective measurement of the severity of pedestrian-vehicle conflicts. In order to measure the accuracy of detecting pedestrianvehicle conflicts, system output was contrasted with observer-based traffic conflict analysis. The results demonstrated the effectiveness of the developed methodology for detecting pedestrian-vehicle conflicts.
c. The development of a new methodology for road user classification into pedestrian and non-pedestrians. The performance of the classification approach was superior to a maximum-speed-based classifier and provided a solid support for subsequent analysis.
d. The novel application of automated analysis of pedestrian-vehicle conflicts in the context of before-and-after evaluation of safety treatments. The pedestrian safety treatment analyzed is pedestrian scramble, a dedicated phase for pedestrians to cross from any curb to another. This contribution component represents the first attempt to probe the severity of all traffic events that involve a pedestrians and non-pedestrians in an automated fashion. The results of the automated
system were to a satisfactory degree consistent with findings in the technical literature.
e. The development of an objective methodology of integrating various severity aspects measured by different conflict indicators into a single severity index. The proposed methodology was tested on the video data used for developing the application on before-and-after safety evaluation. Important arguments were reached on two main strategies for aggregating safety measurements. It was demonstrated that the level of details at which the positions of road users were extracted provides superior exposure measures than simpler surrogates currently used in practice.
f. The development of a new technique for the automated detection of road user violations. The detection technique is based on identifying the anomaly of road user movements in contrast with previously learnt normal movement patterns. The superiority of the developed technique over standard solutions for this detection problem was successfully demonstrated.

### 1.4.3 Video Library

An important secondary contribution of this thesis is the creation of an extensive video library. The total length of video data observations is approximately 80 hours. Video observations were collected from intersections in cities of Vancouver and Edmonton Canada, Oakland California, Cairo Egypt, and Kuwait City Kuwait. The video observations have supported the several studies presented in this thesis as well as independent developments
in the field of computer vision by other parties (Khanloo et al. 2010). The video library is also being analyzed in a number of ongoing projects. Appendix A contains a more detailed description of the video library.

### 1.5. Thesis Structure

This chapter presents an introduction to the thesis, statement of the research problems, and research outline. Chapter 2 documents a comprehensive review of the literature. The review is broad in dealing with various applications of computer vision techniques in the realm of transportation engineering. A narrow focus of the review is on relevant work in the subject of pedestrian-vehicle conflict analysis. Chapter 3 presents the developed methodology for extracting real-world coordinates of features that appear in video sequences. Chapter 4 presents the application of computer vision techniques in the measurement of pedestrian walking speed. Chapter 5 and Chapter 6 outlines the details of automated pedestrian-vehicle conflict analysis in the context of conflict detection and before-and-after analysis, respectively. Chapter 7 discusses the development of a methodology for aggregating various conflict measures into a severity index. Chapter 8 contains the details of a new technique for the detection of road user movements in violation to traffic regulations. Finally, summary, conclusions and proposed future research are presented in Chapter 9. An outline of the research problems and contributions in each chapter is presented in Table 1.1.

Table 1.1 Thesis structure

| Chapter | Content | Related Problem | Related Contribution |
| :---: | :---: | :---: | :---: |
| One | Background materials, broad challenges, motivation for this work, research problems, list of contributions, and thesis structure. | - | - |
| Two | Review of the literature on the following subjects: detection and tracking of pedestrians and vehicles, traffic conflict techniques, and legal issues for video data collection. | - | - |
| Three | Description of a methodology for camera calibration, algorithmic details, and description of case studies. | 1 | 1.3.1a \& 1.3.1b |
| Four | Presentation of research work on the automated measurement and validation of pedestrian walking speed using feature-based tracking. | 2 | 1.3.1c-1.3.1f |
| Five | Presentation of a developed system for the automated detection of pedestrian-vehicle conflicts using feature-based pedestrian and vehicle tracking. Algorithmic details of objective conflict indicators are presented. | 3 | $\begin{gathered} 1.3 .2 \mathrm{a} \& \\ 1.3 .2 \mathrm{~b} \end{gathered}$ |
| Six | Description of an application for the automated before-andafter analysis for pedestrian-vehicle conflicts. Description of a novel methodology for road user classification into pedestrians and non-pedestrians. | 4 | $\begin{gathered} 1.3 .2 \mathrm{c} \& \\ 1.3 .2 \mathrm{~d} \end{gathered}$ |
| Seven | Description of methodologies for calculating aggregate severity measures based on observations of pedestrianvehicle conflicts. | 5 | 1.3.2e |
| Eight | Description of a methodology for automated detection of violating road user movement. | 6 | 1.3.2f |



## Literature Review

### 2.1 Background

The development of computer vision techniques for the purpose of automated detection and tracking of road users has been the subject of extensive study. In addition, the literature of transportation applications of computer vision techniques is steadily growing. This chapter presents a broad review covering the topics presented in this thesis and provides background for further developments in subsequent chapters. Specific
reviews of the literature are presented in subsequent chapters; each review covering their respective subject matter.

Two main topics are addressed in this chapter: traffic conflict techniques and relevant applications of computer vision techniques in transportation engineering. A broad and comprehensive review is not the primary objective of this chapter. Rather, representative studies are selected for a more focused and critical review. The two reviews have minor overlap represented by studies that adopt computer vision techniques for traffic conflict analysis.

The first review of traffic conflict techniques focuses on a number of key studies, also considered as milestones, on this subject area. The second review draws on an extensive body of work found in academic journals and key refereed conferences on the subject of computer vision techniques. After excluding unrelated applications such as pavement data collection, autonomous vehicle control, license plate recognition, and semi-automated analysis of video data, a total of 230 studies were obtained. The review presented in this chapter focuses on a reduced list of selected highly cited studies. At the end of this chapter, a discussion is provided on several legal issues concerning video data collection. The treatment of this subject is of special significance for future development of computer vision applications in transportation engineering.

### 2.2 Traffic Conflict Techniques

### 2.2.1 Important Milestones

The adoption of surrogate safety measures is an emerging subject of research and practice in transportation engineering. Surrogate safety measures are a class of techniques for road safety analysis which rely on data other than road collisions. It is arguable that road collisions occur due to some mechanism of failure in vehicle control. In some situation, the same mechanism of failure can be recovered to an extent that collision is avoided. In that, a traffic event may be classified into a number of different types depending on their safety consequences. In the context of road safety, a traffic event can be defined as the situation in which two road users navigating the same traffic facility come within reasonable temporal and spatial proximity from each other. As discussed before, in some traffic events, evasive action is undertaken, for example braking or swerving, with sufficient strength to avoid collision. These events are defined as near-misses or traffic conflicts ${ }^{1}$. If the strength of the evasive action is not capable of avoiding physical contact of the interacting road users, the involved road users collide or come to physical contact. This intuitive exposition applies to a wide range of disciplines in which nearmisses or injurious events are of similar nature to, and therefore act as precursors for, more catastrophic or fatal events.

One of the earliest sources that the author became aware of was the work by the industrial safety pioneer H. Heinrich (Hayhurst 1932). After extensive experience as an investigator of industrial accidents, Heinrich postulated that

[^4]for every one fatal accident or major injury, there were 29 minor and 300 nearmisses (Vanderbilt 2008). He later proposed an enduring idea that it is possible to create a hierarchical arrangement of fatal events, injuries, and near-misses, respectively. The intuition behind this work is that under the assumption that some structure of this arrangement is preserved, it is possible to draw inference on the incidence of top events (high in severity and consequences) by keeping a record of less severe, yet more frequent events, at the bottom.

The seminal work by Perkins and Harris was the first study on record that proposed a formal procedure to observe traffic conflicts (called near-collisions in the original work). They postulated that reducing hazardous traffic events may lead to reducing the frequency of road collisions (Perkins \& Harris 1968). They further argued that the same failure mechanism in the driving process leads to the occurrence of both traffic conflicts and road collisions. The authors followed a line of reasoning similar to that of Heinrich by postulating that the observation of traffic conflicts may provide sufficient data to draw inference on the occurrence of road collisions. Despite the limitations of the conceptual definition of traffic conflicts and the basic procedure for traffic conflict observations in this study, it can be regarded as a milestone in the course of development of traffic conflict techniques.

In an important step for formalizing traffic conflict techniques, the conceptual definition of a traffic conflict was modified so that it does not necessitate the occurrence of an evasive action. This redefinition of a traffic conflict eliminated the logical boundary between the nature of road collisions and traffic conflict
because the former may lack the occurrence of evasive action (Chin \& Quek 1997). Therefore, a traffic conflict was redefined to occur when, "two or more road users approach each other in space and time to such an extent that a collision is imminent if their movements remain unchanged" (Amundsen \& Hydén 1977).

The theory of severity hierarchy of traffic events was proposed by Hydén (1987) and was later investigated using real-world measurements and generalized to include the potential consequences of traffic conflicts by Svensson \& Hydén (2006). The theory postulates that there exists a severity dimension along which all traffic events can be arranged. At the one extremity of this dimension lie uninterrupted passages. The latter are traffic events with no conceivable chain of events leading to collision between the involved road users. At the other extremity of this dimension lie fatal collisions. It was originally postulated that the shape of the frequency distribution of traffic events at different severity levels would be a pyramidal (Hydén 1987). Later it was argued that with a particular selection of severity measurement, the safety hierarchy exhibits the shape of a diamond (Svensson \& Hydén 2006). The diamond shape of the severity hierarchy was corroborated by findings in this thesis.

Despite the wide recognition of the severity hierarchy theory, it suffers from an often overlooked drawback. There has not been any development of an objective mapping that interprets the spatial and temporal proximity of road users in a traffic event, including road collisions, into a unique severity dimension. That is, proposed objective mappings in the literature are unable to comprehend the severity of the entire set of possible traffic events. In fact,
the previous statement is only true due to the inclusion of road collisions to the set of all possible traffic events. There is no objective mapping from the spatial and temporal proximity of road users involved in a collision into the same severity dimension along which uninterrupted passages and traffic conflicts are arranged. A remedy for the previous drawback that constituted another milestone in the development of traffic conflict analysis was the extreme value formulation by Songchitruksa \& Tarko (2006). This approach represents the only quantitative attempt to create a unified theory or rather a common distribution representing both traffic conflicts and collisions along the same severity dimension.

The model proposed by Songchitruksa \& Tarko (2006) represents collisions as extreme realizations of an underlying distribution of the temporal proximities of road user. The model was plausible in selecting a boundary along the severity dimension between traffic conflict and collisions. A limitation of this model is that the proposed temporal proximity measure fails to comprehend the severity differential among road collisions. In that, the entire extreme value formulation faces the same question as does the theory of severity hierarchy. The theorized severity dimension remains to this date unobservable in its entirety. The different mappings available in the literature for the purpose of severity measurement capture only partial and in some cases independent aspects of severity. This shortcoming laid the ground for the work described in Chapter 7 which proposes an approach for integrating various safety cues in the context of pedestrian safety.

While important theoretical developments were being achieved to create logical and construct validity for traffic conflict techniques, the empirical validity of the techniques was put in question. Amid debate over the validity of traffic conflict techniques, Hauer \& Gårder (1986) presented one of few rigorous treatments of this subject. The intuition behind the validity of the traffic conflict techniques is whether the incidence of traffic conflicts constitutes a reliable predictor of road collisions. In that the validity of traffic conflict techniques can be regarded as a matter of degree. This degree can be measured by the variance of a statistical estimator of road collisions in terms of the frequency of traffic conflicts. This study was novel in the formalization of a statistical mechanism for measuring the validity of traffic conflict techniques. However, few subsequent studies that investigated the validity of traffic conflict techniques made informed use of the aforementioned statistical framework.

The last key study was conducted by Saunier and Sayed (2008). The conceptual definition of traffic conflicts contains the proposition "unchanged". Prior to the previous study, calculation of various spatial and temporal proximity measures had been conducted based on simplistic extrapolation of road user positions assuming constant velocity. While this extrapolation may be closer to reality in case of pedestrian-vehicle conflicts, it is not generally the case for other road users. The novelty in the aforementioned study is that road user trajectories are predicted based on the observation of the movement patterns of previous road users navigating the same traffic intersection. There are two key advantages of this approach:

1. Reliance on records of road user tracks to build extrapolation hypotheses. This data-driven methodology is favourable due to their transferability among different driving contexts and lesser reliance on rules, heuristics, and specific assumptions. It is also well suited to adapt automatically to changing conditions as can be expected for traffic patterns.
2. The extrapolation hypotheses are defined up to a probability. The latter is evaluated based on the frequency of observing a specific motion pattern. This is a logical representation of an uncertain course of action that significantly improves on current deterministic approaches.

While constituting a key methodological development, the following shortcomings are identified in the aforementioned study by Saunier and Sayed (2008):

1. The calculation of the probability of collision, based mainly on the previous work of Hu, Xie \& Tan (2004), is not based on a theoretical formulation of what constitutes a probability of collision and what its calculation entails. In essence, the calculated probability of collision is an index that maps to $[0,1]$ and its being referred to as a probability has to be more reasoned. Some conceptual refinements were proposed in subsequent work (Saunier, Sayed \& Ismail 2010).
2. This next shortcoming is mainly a consequence of the first one. The purpose of estimating road user trajectory is to predict how road users might have driven to end up in a collision. Two separate uncertainties exist. First, the destination of each road user and second, the deviation
from a precisely defined typical course of movement. The first source of uncertainty can be eliminated by observing a road user further in time. Because no formal definition of sample space was articulated, two types of uncertainties were indiscriminatingly used and produced erroneous probabilities of collision based on road user trajectories that would never be followed.
3. Finally, the proposed probabilistic conflict indicator requires the presence of a collision point between the conflicting road users. It is arguable that while the presence of a collision point is prerequisite for a collision to take place, the mere dangerous proximity between road users can be a genuine severity measure. For example, Songchitruksa \& Tarko (2006) demonstrated the validity ${ }^{2}$, i.e., relation to collisions, of a temporal proximity measure that does not require the presence of a collision course in predicting collision frequency. The severity differential between traffic events that include a collision and those which do not is virtually unknown. Recent work by (Laureshyn (2010) provides additional components to temporal proximity measures in order to create a continuum that includes traffic events with and without a collision course. Whether this synthetic continuum reflects the genuine severity of traffic events is yet to be proven.

### 2.2.2 Objective Conflict Indicators

Objective conflict indicators (simply conflict indicators, proximal safety indicators or proximity measures) are quantitative measures of the closeness

[^5]of a conflicting pair of road users, in space and time, in anticipation to a point of collision. The key advantages of conflict indicators are: 1) traffic events that contain calculable conflict indicators are more frequent than collisions, 2) conflict indicators are mainly quantitative measures, therefore they overcome some subjectivity limitations of traditional observer-based conflict indicators, 3) they measure genuine severity aspects of traffic conflicts, and 4) they have been adopted in numerous studies in the literature to measure safety, thus enabling validation and cross-comparisons of studies.

A number of fundamental problems in the use and the definition of conflict indicators have been identified (Chin \& Quek 1997). Many of these problems arise due to an inconsistent and a basic definition of what an "evasive action" is, when it is commenced, extrapolation hypotheses of road users, and validity of conflict indicators in measuring safety. Table 2.1 lists major drawbacks of conflict indicators commonly used in measuring the severity of pedestrian-vehicle conflicts.

It is evident that many of the issues related to the requirement of a collision course arise from the extrapolation of road users' positions, i.e., how road users would move "if their movements remain unchanged"; to quote the traffic conflict technique conceptual definition. Issues with the requirement of an evasive action are that it is sometimes difficult to explicate evasive actions from normal adaptations of the position and velocity of one road user while navigating a traffic stream. As discussed earlier, a more informed extrapolation of road user tracks has been developed by Saunier \& Sayed (2008). While conflict indicators have been lauded for their objective nature,
little work has been done on identifying the definitive severity aspects they individually measure. It can be argued that the reliance on road user positions provides partial image of the information required to capture the true severity of traffic events. For example, conflict indicators may report identical severity measurement for the following two cases: 1 ) a pedestrian is conflicted by a vehicle within some spatial and temporal proximity 2 ) the same encounter except with a larger vehicle. A human observer will intuitively rate the first case at higher severity level than the second.

Table 2.1 Drawbacks of conflict indicators

| Conflict Indicator | Drawbacks |
| :--- | :--- |
| Time to Accident (Hydén 1987) | Does not take into account reaction time. <br> Requires road users to be on a collision course. |
| Time to Collision (TTC) <br> (Hayward 1968) | Requires road users to be on a collision course. <br> Does not account for the velocity of impact. <br> Does not account for the length of the interaction. |
| Extended Time to Collision <br> (Minderhoud \& Bovy 2001) | Same issues as with TTC except that it accounts for the length of <br> the interaction. |
| Time to Zebra (Várhelyi 1996) | Is not based an underlying collision mechanism. |
| Post Encroachment Time |  |
| (Allen, Shin \& Cooper 1978) | Does not require speed and distance measurement, hence missing <br> many cues for conflict severity. <br> Lack of a clear definition of right of way infringement. Limited <br> ability to comprehend severity of interaction between motorists <br> and pedestrians, e.g., when the motorist accelerates past the <br> pedestrian. |
| Gap Time (Archer 2004) | Requires extrapolation of road user tracks. <br> Lack of a clear definition of right of way infringement. |
| Deceleration to Safety Time <br> (Topp 1998) | Requires extrapolation of road user tracks. |

In these hypothetical traffic events, it is possible that the human observer subjectively took into account the consequences of the potential collision that could transpire had no adequate evasive action taken place. Furthermore, the human observer subjectively compares the observer traffic conflict to the innumerable traffic events watched and experienced beforehand. To put it boldly, the subjective assessment of human observers which is widely seen as central weakness to traffic conflict techniques may in fact capture more sophisticated severity aspects than objective conflict indicators. Not surprisingly, Shinar (1984) found no significant agreement between human observers and objective conflict indicators. He concluded that it is likely human observers assess the severity of traffic conflicts through a different, but not necessarily deficient, mechanism than objective conflict indicators. To further prove the point, Svensson (1992) in validating the Swedish Traffic Conflict Technique found that serious traffic conflicts rated as such by subjective human assessment was in stronger correlation with collisions than serious conflicts rated by objective conflict indicators. The main shortcoming of subjective severity assessment of traffic conflicts is not precisely the ability to comprehend the severity of traffic conflicts, but rather the inconsistency in doing so. These facts support a hypothesis in this thesis that objective conflict indicators measure different and sometimes independent severity aspects that ought to be integrated to provide a better representation of the genuine and unobservable severity of traffic conflicts. This hypothesis is explained in more detail in Chapter 7.

### 2.2.3 Challenges to Traffic Conflict Techniques

Traffic conflicts are more frequent and much less costly ${ }^{3}$, if any, than road collision. Traffic conflicts can be observed in field and provide more information about the failure mechanism leading to road collisions. Despite the well-recognized advantages of the reliance on traffic conflicts over road collisions as the main data type, traffic conflict techniques suffer from several shortcomings. Following is a description of these shortcomings:

## a. Consistency

This shortcoming concerns the definition of a traffic conflict. The conceptual definition of traffic conflicts was originally based on evasive actions taken by one or more interacting road users (Perkins \& Harris 1968). This definition however has the logical shortcomings of not including road collisions as conflicts. This placed the approach at both perils of weak correlation with road collision, as evidenced by a number of subsequent studies, and conceptual difference. The other shortcoming of this conceptual definition is that it is often difficult to discriminate between an evasive action and regular precautionary actions or adaptations to road user movement. The latter events are irrelevant to safety analysis. A unified conceptual definition of a traffic conflict was therefore proposed, as presented earlier (Amundsen \& Hydén 1977). While the concept is well-defined, field observation requires the interpretation or codification of this definition into a set of rules. Several operational definitions have been proposed with the underlying strategy of developing the simplest procedure to capture the largest number of relevant events.

[^6]A review of the literature of traffic conflict analysis of pedestrian safety yielded a number of operational definitions as shown in Table 2.2. The definitions appear of specific nature and tailored to measure safety issues related to particular safety treatments. Moreover some key studies in the literature lack precise description of the operational definitions used to observer traffic conflicts, e.g., (Tiwari, Mohan \& Fazio 1998). As evidenced by Table 2.2, there was no common operational interpretation of traffic conflicts found in the literature. This shortcoming was evidenced by a number of comparative studies, e.g., (Grayson et al. 1984), in which different teams were asked to detect, rank and rate conflict severity from a common data set. There were considerable variations in the scores suggested by each team (Chin \& Quek 1997).

## b. Validity

The validity of traffic conflict techniques is often defined in terms of its ability to predict road collisions. Fundamentally, if observing traffic conflicts gives inference on the risk of collision, then reducing traffic conflicts can help safety agencies achieve their ultimate goal - reducing collisions. Proving the validity of traffic conflict techniques has been conducted using different approaches, e.g., methodology and citations in (Hauer \& Gårder 1986). Irrespective to the validation method, establishing a sound relationship between traffic conflicts and road collisions has been a persisting problem. Critiques of traffic conflict techniques argued that for every study that corroborates validity there is another that fails to find relationship with collisions (Williams 1981). As a response, proponents of traffic conflict techniques sought to improve the validity of the conflict techniques by redefining conflicts or explaining the
causes of poor correlation with collisions (Muhlrad 1993) (Hydén, Garder \& Linderholm 1982). Others argued that the main problem for proving the validity of traffic conflict techniques lies in the known issues with collision data (Chin \& Quek 1997). Others argued for the validity of traffic conflict techniques by construct, that is the approach is correct in its own right since traffic conflicts comprise a genuine danger to road users (Grayson \& Hakkert 1987). Comparative studies on the validity of traffic conflict techniques for predicting pedestrian-involved collisions provided mixed results (Lord 1996). A prime example of positive empirical evidence, against the previous argument, that proves the validity of traffic conflict techniques is the important study by Sayed \& Zein (1999). A statistically significant correlation between frequency of traffic conflicts and collisions was found at various severity levels.

## c. Reliability

From its beginning, traffic conflict techniques have been mainly based on observations collected in field by trained observers. Due to issues with operational definitions of traffic conflicts and the intrinsically demanding nature of the task for human observers, detection and severity rating of traffic conflicts have suffered from inter- and intra-observer variability. The first of these two shortcomings is classified as a variability or, lack of consistency, problem and the second as a repeatability problem (Glauz \& Migletz 1984). Conceivably, the rule details of an operational definition may be challenging to implement on site, especially under significant workload. In this situation, the field observer will become overwhelmed with detection and rating rules, thus increasing the chance of mistakes. To overcome some of these problems,
in-office analysis of video observations was advocated. However, fieldobserver collects higher-quality and first-hand assessment of traffic events with better judgement (Chin \& Quek 1997). Video observations, without dedicated video analysis tools, are restricted in field of view and dimensionality. A more promising alternative to the subjective and rule-based definitions of traffic conflict is objective conflict indicators, e.g., time to collision (TTC) (Hayward 1968), temporal proximity (Allen, Shin \& Cooper 1978), and other proximity measures. With advances in the field of computer vision, these conflict indicators can be calculated automatically, as demonstrated in this thesis, thus hopefully increasing accuracy and reducing the labour burden.

Table 2.2 Traffic conflict definitions in the literature

| Definition | Type |
| :---: | :---: |
| Undertaking of evasive action, e.g., brake-light indication or lane change, or a violation to traffic regulations (Perkins \& Harris 1968) (Williams 1981) | Generalconceptual |
| "....an observable situation in which two or more road-users approach each other in time and space to such an extent that there is a risk of collision if their movements remain unchanged" (Amundsen \& Hydén 1977). | Generalconceptual |
| Traffic conflict is an event in which a driver takes an evasive action to avoid a collision (Cynecki 1980). | Generalconceptual |
| An angle conflict occurs when a road user avoids striking a pedestrian. A traverseangle conflict occurs when a crossing pedestrian stops to avoid being struck by another road user (Tiwari, Mohan \& Fazio 1998). | Pedestrianoperational |
| A potentially unsafe interactive event that requires evasive action (braking, swerving or accelerating) to avoid collision (Archer 2004). | Generalconceptual |
| Pedestrian-vehicle conflicts defined and classified based on detailed event description (Medina, Benekohal \& Chitturi 2009): <br> Non-severe conflicts. <br> - A pedestrian crosses halfway and waits in the centre of the road as motorists pass. <br> - A pedestrian terminates crossing action and reverts to the curb. <br> - A pedestrian rushes to the exit curb due to an approaching motorist. <br> Severe conflicts: <br> - A pedestrian waits in the middle of a road with no median and motorists keep passing. <br> - A pedestrian runs to the exit curb and an approaching motorist does not seem to slow much. <br> - A motorist swerves around a pedestrian. <br> - A pedestrian forces a motorist to come to a sudden stop by stepping into the road. <br> - A pedestrian runs to cross a street forcing an approaching motorist to suddenly slow down. <br> - A pedestrian crosses a four-lane street right after getting off a bus and stops in front of the bus or causes motorists to suddenly stop. | Pedestrianoperational |

### 2.3 Computer Vision Developments

### 2.3.1 Computer Vision Developments for Pedestrian Detection

Pedestrian detection is a difficult task in computer vision. There is no explicit model for the appearance of a human body, which can assume innumerable shapes and poses. In general, machine learning techniques rely on the learning of an implicit pedestrian model from a set of pre-defined samples. Other difficulties arise from particularities of pedestrians such as: varying appearance, articulated pose, local non-rigidity, varying clothing, proneness to visual occlusion in crowd movement, and variable background within which pedestrians may be present. Research into automated pedestrian detection has recently intensified, driven mainly by a tremendous commercial and security application potential. From the immense diversity of pedestrian detection and tracking techniques, methodological patterns are noted. Typical sequence of analysis steps starts with 1) hypothesis generation, 2) classification, and 3) tracking. A brief summary of work developed in the realm of computer vision on pedestrian detection and tracking is based on a recent review by Enzweiler \& Gavrila (2009) and is presented in the following sections.

## Pedestrian Hypothesis Generation

In the first step, an initial position of the pedestrian hypothesis is determined. Common methods for hypothesis generation are sliding window techniques, which involve the shifting of windows of various sizes over the image to localize a pedestrian hypothesis. Despite the significant computational cost, several studies employed the sliding window technique, e.g., (Dalal \& Triggs
2005) (Mohan, Papageorgiou \& Poggio 2001) (Sabzmeydani \& Mori 2007) (Szarvas et al. 2005). Variations on this approach were developed for reducing computational cost by employing a series of classifiers along with the instantiation of different hypotheses (Wu \& Nevatia 2007) or by adopting specific assumptions regarding the road and object geometry (Leibe et al. 2007) (Gavrila \& Munder 2007). Approaches for hypothesis generation have been also developed based on stereo vision. For example Alonso et al. (2007) used a localization mechanism for pedestrian hypotheses based on stereo vision in order to overcome the lost depth cues during monocular vision. Motion features have been employed for hypothesis generation, especially with approaches based on background segmentation (Zhao \& Nevatia 2004). Other approaches identify pedestrian hypotheses based on distinctive local features, such as outline detection (Agarwal, Awan \& Roth 2004).

## Pedestrian Classification

The second step after the instantiation of a pedestrian hypothesis is to verify that this hypothesis (object) is indeed a pedestrian using various shape, motion, appearance, outline, and temporal cues. An extensive body of work has been developed for this purpose. Enzweiler \& Gavrila (2009) proposed a broad categorization scheme for pedestrian classification into generative and discriminative approaches along with further delineation in each broad category. The generative approach to pedestrian classification assumes an underlying density model for the appearance of pedestrians. The generative approach can be implemented in a Bayesian framework by defining a class prior and updating the class density distribution using observations from the
concerned video sequence. The generative approach can be further divided into shape models and shape/texture models.

Shape models possess the appealing characteristic of being independent of clothing or illumination. The majority of discrete shape models adopt various 2D geometric models encapsulating the pedestrian. Exemplar-based shape models use specific pedestrian shapes as matching templates for pedestrians, e.g., (Zhou, Gao \& Zhang 2007) (Gavrila 2007). The advantage of the explicit and specific shape examples is however compromised by the practical requirement for a large number of shape examples to adequately cover the space of all observable shapes. Efficient matching methods have been developed to reduce the storage requirement and enable real-time applications, e.g., (Stenger et al. 2006). In contrast to discrete shape models, continuous shape models use a parametric representation of the pedestrian shape. The distribution of pedestrian class shapes is learnt from examples of manually annotated (Heap \& Hogg 1998) or automatically detected pedestrian shapes (Enzweiler \& Gavrila 2008). Various approaches have been developed for learning linear, non-linear, and piecewise linear models of pedestrian shapes (Enzweiler \& Gavrila 2008) (Munder \& Gavrila 2006). Continuous shape models possess the advantage of being more flexible in representing pedestrian shapes than discrete shape models. However, this advantage is realized at the expense of increased computational cost involved in the learning of the distribution parameters of continuous shape models.

As evidenced by several road user detection tasks, the combination of several cues holds consistent advantage over approaches that depend on fewer or
unique cues. Another discriminative feature of pedestrians is their texture, which is more diverse than the texture patterns of motorized road users. Approaches have been developed for the combined parameterization of pedestrian appearance in terms of shape and texture (Fan, Sung \& Ng 2003) (Enzweiler \& Gavrila 2008).

The discriminative approach for pedestrian classification involves the Bayesian learning of the parameters of a decision function that distinguishes between pedestrian and non-pedestrian objects. The main inputs to the decision function are discriminative features. Examples of these discriminative features are local intensity differences at different parts of the image (Papageorgiou \& Poggio 2000). A dictionary of these features is predefined for various orientations and scales of pedestrian examples. Techniques for automated selection of the most discriminative feature have been proposed based on the popular AdaBoost algorithm (Viola, Jones \& Snow 2003) (Cao, Qiao \& Keane 2008) or defining features that adapt to the underlying dataset (Munder \& Gavrila 2006) (Gavrila \& Munder 2007). The use of orientation histograms of image gradients computed at local image subregions as discriminative features has been popularized by the seminal works of Dalal \& Triggs (2005) and extended for various subregions by Zhang, Wu \& Nevatia (2007). Other approaches for feature identification include interest-points popularized by the important work of Lowe (2004), manually annotated collection of edges (Wu \& Nevatia 2007), spatiotemporal features that capture human gait (Sidenbladh \& Black 2003), and crossspectral pedestrian detection using stereo and infrared vision (Krotosky \& Trivedi 2007).

Several models for the decision function have been proposed. Multilayer neural networks have been successfully used in particular with adaptive features (Szarvas et al. 2005). Support Vector Machines have emerged recently as a more powerful tool for addressing classification problems. As opposed to traditional techniques that minimize some arbitrary error measure, Support Vector Machines maximize the margin that separates some hyperplane and the elements that belong to different classes. A growing number of studies demonstrated successful adoption of Support Vector Machines ranging from simple and cost-efficient linear classifiers (Zhu et al. 2006) (Zhang, Wu \& Nevatia 2007) to more sophisticated non-linear Support Vector Machines (Mohan, Papageorgiou \& Poggio 2001) (Munder \& Gavrila 2006). Recent approaches for pedestrian classification rely on part-specific (componentbased) features instead of seeking full-body descriptive features. Examples of this approach are (Mohan, Papageorgiou \& Poggio 2001) (Sidenbladh \& Black 2003) (Alonso et al. 2007). Component-based features have the advantage of requiring fewer training examples and having the potential to overcome partial occlusion. This advantage is however compromised by model complexity, computational cost, and degraded performance in low-quality images.

## Pedestrian Tracking

The third and final step is pedestrian tracking. Tracking involves the assignation of various pedestrian objects in successive images to a common sequence, commonly called a track or a trajectory ${ }^{4}$. One elementary approach

[^7]for pedestrian tracking relies on the geometry and dynamics of pedestrian objects detected in a sequence of frames (Gavrila \& Munder 2007). There is however more tracking-relevant information in a sequence of pedestrian images. Following reasoning similar to pedestrian classification, a mixture of features can be used for the assignment of pedestrian objects to a common track. Appearance models have been used in conjunction with geometry and dynamics in a significant number of applications, for example (Sidenbladh \& Black 2003) (Zhao \& Nevatia 2004) (Ramanan, Forsyth \& Zisserman 2005) (Munder \& Gavrila 2006) (Wu \& Yu 2006) (Wu \& Nevatia 2007) (Zhang, Wu \& Nevatia 2007).

In the preceding categorization of pedestrian detection and tracking, the two steps were presented in a sequential and apparently separate form. It is arguable that discriminative features of pedestrian objects can be extracted by analyzing the sequence of positions. For example, simple maximum speed was used as a discriminative feature in a road user classification developed in this thesis (Ismail, Sayed \& Saunier 2009). A peculiar motion rhythm exhibited by pedestrians due to their movement by ambulation, as opposed to rolling for vehicular objects, is a potential discriminative feature (Ran, Chellappa \& Zheng 2006). Therefore, pedestrian classification can be enhanced by recovering information from pedestrian tracks, thus giving rise to an iterative mechanism for pedestrian classification and tracking. Another approach to link the two steps is the integration of both steps under a Bayesian framework that combines appearance features with pedestrian dynamics using single cues ( $\mathrm{Wu} \&$ Nevatia 2007) or multiple cues (Sidenbladh \& Black 2003) (Ramanan, Forsyth \& Zisserman 2005) (Munder \&

Gavrila 2006). Particle filtering is a popular approach for integrating multiple cues by approximating a joint (combining multiple cues) posterior density by a mixture of weighted random samples (Khan, Balch \& Dellaert 2005).

## Night-time Pedestrian Detection

A particular concern for pedestrian safety arises in night-time conditions. Poor pedestrian visibility leads to an abnormally high rate of pedestrianinvolved collisions during night-time conditions. The literature contains several studies on night-time pedestrian detection and tracking. A method was proposed by Xu , Liu \& Fujimura (2005) for night-time pedestrian detection and tracking from monocular infrared vehicle-mounted cameras. In this study, a model based on Support Vector Machines was used for detection along with a combination of Kalman Filter prediction and mean shift for tracking. A probabilistic model for night-time pedestrian detection was developed by Bi, Tsimhoni \& Liu (2009) using distance and image metrics (clutter metrics, contrast, and blob size).

## Benchmark Performance Evaluation

Despite the wide availability of public video and image libraries for testing, reported performance tests of various methods and algorithms are at variance. For example, Enzweiler \& Gavrila (2009) note the difference in performance reporting within the same study (Viola, Jones \& Snow 2003) as well as in contrast between the previous study and (Leibe et al. 2007). To address the evident discrepancies of individual evaluations, Enzweiler \& Gavrila 2009 conducted an independent evaluation of a number of selected pedestrian detection and tracking approaches. The dataset, called the Daimler benchmark library, contains several thousands of pedestrian images manually
extracted from video sequences. The number of positive pedestrian examples provided for training was 15,660 and the number of negative examples (images empty of pedestrians) was 6,744. The testing dataset contained 250 pedestrian tracks recorded from a monocular vehicle-mounted camera. A dataset comprising 21,790 pedestrian images was used for testing. The following detection and tracking approaches were selected to represent, to a reasonable extent, the range of all potential approaches:

1. Haar wavelet-based cascade (Viola, Jones \& Snow 2003)
2. Artificial neural networks (Woehler \& Anlauf 1999)
3. Histograms of oriented gradients (Dalal \& Triggs 2005)
4. Multi-resolution Texture-based classification (Gavrila \& Munder 2007)

The evaluation methodology was based on comparing the number of positive detections with the ground truth, whether for fact the concerned image contains a pedestrian. In this thesis ground truth is defined as a reference set of data that is considered free of detection, classification, and tracking errors. No 3D scene information was provided for pedestrian detection. The evaluation was conducted under varying settings in order to obtain a nuanced picture of the performance of each approach. Following the enumeration of the detection approaches in the previous list, the following conclusions were reported (Enzweiler \& Gavrila 2009):

- Approach 1 was the most reliable in low-resolution images (approximately 650 pixels/pedestrian).
- Approach 3 provided the best performance in medium-resolution images (approximately 4500 pixels/pedestrian).
- Temporal integration which considers track-relevant information during detection phase improved the performance of all approaches for pedestrian detection.
- Majority of false detections occurred in images containing nonpedestrian vertical features such as poles or traffic signs.
- For real-time ${ }^{5}$ trajectory analysis, approach 1 produced the highest detection rate ${ }^{6}$.


### 2.3.2 Computer Vision Developments for Vehicle Detection

The advantages of video sensors over other vehicle detectors have been recognized for decades. The rich and detailed information about road users and the traffic scene that video sensors provide exceeds traditional sensors such as radar, ultrasonic, and closed-loop detectors (Wang, Xiao \& Gu 2008). Developing a functional real-time video-based traffic surveillance system has been an elusive target for a number of decades starting from later 1970s (Wang, Xiao \& Gu 2008). In general, road user detection and tracking involve the same theoretical problem, irrespective to the type of the concerned road user. However, vehicle detection poses less challenges than pedestrian detection. Vehicles movements are relatively more regular and stereotypical than pedestrians. Vehicles are locally rigid with regular and often monotonous texture. Despite these differences, a categorization scheme similar to pedestrians can be adopted to organize various approaches to vehicle detection. Typical sequence of analysis steps starts with 1) hypothesis

[^8]generation, 2) classification, and 3) tracking. A brief summary presented hereafter of work developed in the realm of computer vision on vehicle detection and tracking from still cameras is based on an earlier review (Saunier \& Sayed 2006) and a recent review (Wang, Xiao \& Gu 2008).

## Vehicle Hypothesis Generation

The purpose of identifying the location of vehicle hypothesis is to focus on subregions of an image, or a frame in a video sequence, which potentially contains vehicles. This is similar to coarse-to-fine reasoning for general pattern recognition, in which more accurate and computationally intensive operations are applied on a subset of the entire search space. Three main categories for vehicle hypothesis generation were identified (Wang, Xiao \& Gu 2008).

The first category is frame subtraction which identifies subregions of the image at which there is an inter-frame change in the image value. These differences are typically evaluated at a block level instead of pixel level in order to improve robustness to image noise (Paragios \& Deriche 2000). The main advantage of this approach is that it is not necessary to pre-define or learn a model for the scene or typical vehicle appearance. The main drawback of this approach is its critical sensitivity to the particular selection of a time interval to calculate the inter-frame difference. While shorter time intervals enable more precise vehicle localization, it may exclude stationary or slowmoving vehicles. The adverse effect of long time intervals is the reduction of the detection and tracking precision.

An important approach in the first category is based on statistical testing for inter-frame differences. This approach assumes that pixel-wise changes are caused by a moving object. To obtain accurate detection using this approach, the moving objects should be sufficiently textured and traverse large displacements. Dedicated spatial-temporal detectors have been adopted to compensate for this drawback (Paragios \& Deriche 2000). Another important variant of this category is Markov Random Field methods. This approach attempts to limit the requirement for prior knowledge of the vehicle size by including the latter in a statistical estimation process (Odobez \& Bouthemy 1995).

The second category for vehicle hypothesis generation contains the popular approaches for motion detection using background-foreground segmentation - a shorthand name for this category is background segmentation. The idea of background segmentation is to subtract the current image of the monitored scene from what it would look like had no road users been present. If a suitable background learning approach is adopted, background segmentation can provide remarkable performance. The main challenges for learning background models include variability of scene illumination, shadows, stationary road users that could blend to the background, and adverse weather conditions. Various approaches for background learning have been proposed and successfully adopted for vehicle detection including: frame average (Odobez \& Bouthemy 1995) and minimum and maximum intensity background model (Paragios \& Tziritas 1999). These two approaches are critically prone to embed stationary vehicles to the background model. An approach for vehicle hypothesis generation that benefits from the
stereotypical vehicle movement, especially the delineation of traffic lanes, is virtual loop detectors (Pang, Lam \& Yung 2004). Any abrupt change in image value within the virtual loop detectors triggers a vehicle hypothesis. If the shortcoming of requiring human localization of virtual loop detectors is ignored, this approach becomes one of the most efficient and accurate in this category. A well practiced commercial application of his method is the Autoscope system (Michalopoulos 1991).

One of the most popular approaches for background segmentation is modelling pixel values of the background model as mixture of Gaussians (Stauffer \& Grimson 1999). This approach assumes that the time-series of pixel values is drawn from an independent mixture of Gaussian (Normal) distributions. The pixel-level model is updated online, thus enabling effective real-time applications. Each pixel is evaluated to determine whether this Gaussian, which most properly describes its current value, belongs to the background model. The incidence or relative frequency of a particular Gaussian model at a specific pixel determines whether it belongs to the background model.

## Vehicle Classification

After identifying a vehicle hypothesis, it is classified into a vehicle or a nonvehicle object. Three main approaches for vehicle classification can be identified (Wang, Xiao \& Gu 2008). Notable of which are knowledge-based methods which classify a vehicle hypothesis based on a priori knowledge of characteristic vehicle features. Notable of distinctive vehicle features are shadows, edge orientation (Moon, Chellappa \& Rosenfeld 2002) (Tsai, Hsieh
\& Fan 2007), texture (Kalinke, Tzomakas \& Seelen 1998), image gradient (Tan \& Baker 2000), 3D pose model that encapsulates the vehicle hypothesis (Costa \& Shapiro 2000) (Lou et al. 2005), wheels (Iwasaki \& Kurogi 2007), and motion patterns (Ismail, Sayed \& Saunier 2010). The last approach was extensively used in this thesis and is described in more detail later in Chapter 6.

## Vehicle Tracking

There is no methodological difference between approaches for vehicle tracking and general methods of object tracking. The latter is documented in several reviews in the realm of computer vision. Various object tracking approaches can be classified into four major groups (Liu, Wu \& Zhang 2008) (Saunier \& Sayed 2006): 1) region-based Tracking, 2) feature-based tracking, 3) contour-based tracking, and 4) model-based tracking. Note that there are common elements in tracking approaches and earlier steps (hypothesis generation and vehicle classification). This is explainable by the fact that there is no clear separation of the previous steps and various approaches lie on their boundaries.

Region-based tracking depends on the extraction of connected subregions (blobs) of the image as holistic objects. Most methods of region identification rely on background segmentation. The assignment of blobs identified in image frames to a single track is most commonly performed using Kalman Filters (Badenas, Sanchiz \& Pla 2000) (Veeraraghavan, Masoud \& Papanikolopoulos 2003) (Wang \& Lien 2008).

Feature-based tracking, the method of choice in this thesis, is concerned with the tracking of local features, such as salient points, corners, and edges. The
main advantage of this approach, similar to component-based pedestrian tracking, is the robustness to partial occlusion. A successful application was demonstrated by Hsieh et al. (2006) for vehicle tracking using line features coupled with Kalman Filter for position prediction. In another application, the popular Kanade-Lucas-Tomasi Feature Tracker (Lucas \& Kanade 1981) was used by Saunier \& Sayed (2006) for detecting and tracking features tracked on moving vehicles.

Active contour approaches (snake models) rely on detecting and tracking a model of the vehicle outline or contour. The vehicle contour is dynamically updated in order to fit the observed vehicle outline. Contour tracking is computationally more efficient than previous vehicle tracking approaches by virtue of the simplicity of describing contour models. Successful applications of contour-based tracking have been demonstrated (Peterfreund 1999) (Paragios \& Deriche 2000) (Zhou, Gao \& Zhang 2007).

The fourth approach to object tracking is the model-based. First, an accurate 3D geometric model is established for the detected vehicle. Second, the 3D models are projected on the image plane by knowledge of its dimensions and orientation of movement along with camera parameters. Projected models are then tracked in subsequent frames. Successful application of this approach was previously demonstrated (Sminchisescu \& Triggs 2001). The major drawback of this approach is the requirement for detailed information on vehicle geometry.

### 2.4 Selected Applications in Transportation Engineering

The following sections review a number of selected studies found in the transportation engineering literature that include applications of computer vision techniques. Subject areas such as pavement data collection, autonomous vehicle control, and license plate recognition were deemed outside the scope of this thesis and therefore were excluded from this review.

### 2.4.1 Road Safety Analysis

The importance of video observation in road safety analysis cannot be underestimated. Video sensors enable the recovery of more information about traffic conflicts and collisions than police and insurance records. Several studies noted the benefit of video monitoring of collision events in providing insight into the contributory factors to collisions, e.g., (Conche \& Tight 2006) (Elmitiny et al. 2010). It is also helpful to resolve conflicting evidence often reported from witnesses or collected from site. Despite the obvious benefit of video data demonstrated in this study, review of video data can become a burdensome task if conducted manually. Addressing this well-entrenched challenge is the main drive for this thesis.

A novel methodology for traffic conflict analysis was proposed based on microscopic road user tracks (Saunier \& Sayed 2007). The tracking accuracy required for this application exceeds the requirement for applications on road user detection and counting. To meet this accuracy requirement, featurebased tracking was selected as the method of choice (Saunier \& Sayed 2006). In subsequent work, Saunier \& Sayed (2007) demonstrated the feasibility of
using a clustering approach for traffic conflict detection using a model that combines K-means clustering with Hidden Markov Chain trajectory modelling. The correct detection accuracy for traffic conflicts was $100 \%$. However traffic events that the authors were uncertain about being traffic conflicts were detected by the system at the expense of increased rate of false alarms. Despite the novelty of this work, it does not lead to a sequential development to the severity measurement of traffic events. The clustering techniques demonstrated in this work cannot be directly adapted for calculating objective measures of traffic conflict severity.

The statistical significance of Post Encroachment Time (PET) as an explanatory variable in right-angle collision prediction models was successfully demonstrated (Songchitruksa \& Tarko 2006). The authors also demonstrated the importance of PET count in discriminating various traffic safety levels within a particular location. A total of 16 signalized intersections were monitored, each intersection for 8 hours. PET measurements were automatically collected by aid of Autoscope system (Michalopoulos 1991). A selection was made for the counting of traffic events based on a PET threshold of 6.5 sec . It was not clear in the study what severity regimes this threshold separates. It was found that the combination of traffic volume and PET counts produced significant model coefficients. This proves the presence of safety information conveyed by PET counts in addition to traditional measures of exposure such as traffic volume.

In a companion study Songchitruksa \& Tarko (2006) proposed a novel approach for modelling collision frequency based on an extreme value
formulation. The continuum of crash severity measure was mapped using PET with negative values being interpreted as collisions. The evaluation of the calibrated extreme value model provided an evidence of the relationship between PET and collision frequency. Despite the originality of this study and the promising extensions thereof, it is challenged on theoretical grounds. PET is a temporal proximity measure that is incapable of comprehending the variable severity of collision events. In that, all collision events would possess negative PET, but the magnitude of PET in these cases is irrelevant to collision severity. In that, PET maps traffic events into two distinct and separate severity regimes that may not belong to a unique severity distribution. The use of extreme value theory could possess far more validity if the employed conflict indicator could map all traffic events, including collisions, into a single severity continuum.

One of the key shortcomings of traffic conflict analysis based on simulated vehicle tracks, e.g., (Cunto \& Saccomanno 2008) (Mehmood, Saccomanno \& Hellinga 2001), is the questionable validity of normative driver behaviour models used in microsimulation for simulating driving mistakes that lead to traffic conflicts. A novel driver behaviour model was proposed by Xin et al. (2008) for representing the imperfection of driving tasks including driver inattention and a detailed perception-response process. A total of 54 vehicle tracks that were involved in a total of 10 collisions were successfully replicated by the driver behaviour model. Vehicle tracks were extracted using NG-VIDEO (Kovvali, Alexiadis \& Zhang 2007). This study represents an important development of traffic conflict techniques based on traffic microsimulation. Several issues were not treated in this study such as
sensitivity of calibrated parameters, there transferability, and extended validation for larger datasets that include other collision types, such as the datasets analyzed in previous work (Saunier, Sayed \& Ismail 2010).

### 2.4.2 Behavioural Analysis

An example of behavioural analysis of pedestrian movement using computer vision techniques was developed by Chae \& Rouphail (2008). The main aspect of pedestrian-vehicle interactions was pedestrian gap acceptance behaviour during crossing maneuvers at roundabouts. The data collected in this study were microscopic pedestrian and vehicle tracks. The video processing system was reported to achieve tracking accuracies for vehicles and pedestrians of $92 \%$ and $90 \%$ respectively. The intersection space was divided into a set of subregions for more robust localization of the interacting road users. The maximum pedestrian volume using the monitored crosswalk was 16 pedestrians/hour interacting with a maximum of approximately 1330 vehicles/hour. This is a relatively small traffic volume and probably entails low-density crowd movement. Various measurements of pedestrian walking speed, critical gap, and driver yielding behaviour were provided. Despite the novelty of using computer vision techniques, the results reported in the study are of limited scope due to the low pedestrian and vehicle volumes.

Malinovskiy, Zheng \& Wang (2008) developed a methodology for road user tracking based on background segmentation for gray scale videos. The challenge of occlusion, in case of proximate movement, was addressed by keeping track of blob merging and splitting events. Composite objects composed of merged blobs were tracked as a coherent blob until splitting of
this blob was observed. After splitting, composite objects were dissected into constituent blobs which are matched to blobs before merging using size and colour distribution features. The validation results of pedestrian count and speed measurement were satisfactory with detection accuracy ranging from $83 \%$ to $100 \%$. The validation was conducted on pedestrian facilities with relatively low-volume (maximum 294 pedestrians/ hour). Other unavoidable shortcoming of approaches based on background segmentation is the inability to resolve occlusion or over-grouping if the merged objects do not split within the camera field of view. Another inevitable shortcoming is the inability to resolve over-grouping if merging of blobs lasted for long time. Lastly, background modelling algorithms are imperfect and prove unreliable in crowded scenes or when road users remain stationary for prolonged period of time.

Malinovskiy, Wu \& Wang (2008) demonstrated an application of the previously described pedestrian tracking system in measuring pedestrian walking speed. The homography matrix was calculated using a user-defined square that is drawn as it would appear in the video while being projected from the road surface. There is an unavoidable subjectivity in this definition since there is no guarantee that the user-defined square takes precisely this shape in reality. The pragmatic solution sufficed for this application which involved a traffic scene familiar to the authors. The approach used for classifying road users into pedestrians and non-pedestrians was based on the jaggedness or the rhythm of road user speed profiles. A satisfactory classification performance was observed with $2.6 \%$ false classification rate.

This classification approach is of limited general validity since tracking noise can cause what is an apparently rhythmic movement. Road user periodic movement has been investigated in this thesis and no satisfactory classification performance, into pedestrians and non-pedestrians, was obtained. This could be explained by the open and mixed-use traffic scenes that were monitored in this thesis. However, the study by Malinovskiy, Wu \& Wang (2008) provides another evidence of the practical validity of approaches based on background segmentation for scenes featuring low traffic volume and adequate pixel representation of road users.

### 2.4.3 Performance Evaluation

Prevedouros et al. (2006) conducted one of the key studies on performance evaluation of video detection systems for the purpose of incident detection. Video detection systems were run for a period of 96 days and monitoring the tunnels of the Attica tolled roadway facility in Athens. The findings of this study were multitude and mixed, most notable of them are sensitivity to traffic volume, camera set-up position, and illumination. There was varied performance of different systems and a general conclusion was rightfully drawn as to the immaturity of commercial applications of computer vision technologies.

Medina, Benekohal \& Chitturi (2009) evaluated the performance of automated vehicle detection technology for an extended observation period of 10 hours distributed over different times of the same day. The most favourable illumination conditions, and also used as base condition, was a cloudy noon. The main source of errors in dawn time was the false detection
caused by the reflection of light from adjacent lanes. More pronounced false detection errors were observed (5-53\%) in sunny morning conditions due to shadows extending over adjacent lanes. An error pattern similar to dawn conditions was observed for dusk conditions. At night, headlight reflection caused false detection problem due to light reflection. The results of this study were corroborated by work by Middleton et al. (2008). In this study, the authors noted the critical sensitivity of video detection systems to camera setup position and variation in illumination during the day.

Chitturi, Medina \& Benekohal (2010) evaluated the variability of the performance of three video detection systems. The automated vehicle detection systems were tested for a total duration of 40 hours. All vehicle detection systems committed false detections at rates from $0.2 \%$ to $36 \%$. The significant impact of shadows on system performance was evidenced by the high false detection rate in sunny conditions. The missed detection rate of all systems was remarkably low.

### 2.4.4 Traffic Performance Monitoring

Cheek, Hawkins \& Bonneson (2008) reported an evaluation of a computer vision technology for the automated measurement of queue lengths upstream a signalized intersection. The study emphasized the advantage of video sensors over other intrusive sensors for the purpose of automated measurement of queue lengths in terms of maintenance and installation costs. Video data was recorded by an array of video sensors that covered the concerned intersection approach in conjunction with the study region upstream its road segment.

One of the important computer vision developments in the transportation literature is a wide-area video detection system called Autoscope. The system is based on the concept of virtual loop detectors. This system was initiated at the University of Minnesota in 1984 (Michalopoulos 1991) (Dickmanns 2002) and is used in a large number of transportation applications.

Using the Autoscope technology, Cheek, Hawkins \& Bonneson (2008) carefully defined virtual loop detectors within the study region in order to capture queue fluctuation. A total of 24 hours of data was collected and analyzed. Virtual detectors were situated every $15.24 \mathrm{~m}(50 \mathrm{ft})$ for a distance of 122 m upstream the stop line. The detectors phase reading would feed a queue estimation algorithm based on Kalman Filters. A total of 500 measurement points were observed and a moderate agreement between predicted and observed queue lengths was found. No explanation was provided for the cases of disagreement. It is likely that prediction errors were caused by issues in the detection accuracy.

The accuracy of automated pedestrian counting techniques in outdoor environment was critically investigated (Greene-Roesel et al. 2008). Computer vision was one of the technologies included in the evaluation. The main advantages of computer vision technologies reported in this study were: wide area coverage, potential of reliable performance in crowded conditions, amenability for manual review for additional information, inexpensiveness of hardware, and the ability to keep a permanent record of data. The main shortcomings of computer vision technologies were: the focus of most commercial products on indoor settings, the challenge of counting in
crowded conditions, and the performance vulnerability environmental factors.

Hu et al. (2008) developed a computer vision application for the detection and classification of road users. The technology is based on background segmentation. The features used in road user classification were the minimum distance from the blob centroid to points on its perimeter. Road user classification was further refined using Kalman Filters in order to enforce the predominant classification. The study reported $98 \%$ accuracy of vehicle detection and classification and $70 \%$ accuracy for pedestrians and bikes.

Hubbard, Bullock \& Day (2008) raised the important limitation in current models for the prediction of pedestrian level of service at signalized intersection, namely the model in the Highway Capacity Manual (HCM 2000). These models do not consider the microscopic interaction between pedestrians and motorized traffic. For example, pedestrian facilities that suffer significant interruption by right-turn motorized traffic may be rated as possessing a level of service of $A$. The authors cited empirical attempts to incorporate pedestrian-vehicle conflicts into level of service models, e.g., (Zhang \& Prevedouros 2003) (Akin \& Sisiopiku 2007). However, this attempt faces the same challenges of aggregate and empirical measures of pedestrianvehicle conflicts in lacking the required level of details to adequately describe road user behaviour. For example, volume-based measures of pedestrianvehicle conflicts fail to capture the genuine hazard of a pedestrian crossing when the latter volume is low. These models also fail to comprehend the
difference between the interaction of multiple pedestrians with a single vehicle and a single pedestrian with multiple vehicles.

A novel computer vision system was developed for automated measurement of vehicle speed and traffic volume based on a hybrid background segmentation and Kanade-Lucas-Tomasi feature tracker (Kanhere et al. 2007) (Lucas \& Kanade 1981). The background model was learnt using the median value of image pixels. The developed system was capable of functioning at ground-level camera setting and of providing a detection accuracy of $98 \%$. The authors focused on vehicle front base as a representative geometry for the vehicle foreground component. Furthermore, the dimensions of foreground blobs were used to classify motorized vehicles into passenger cars and trucks. While testing of this approach was conducted on a limited video sequences with total duration of 45 min , it is not expected that reliance on foreground components only will provide reliable classification results.

Qi, Tang \& Smith (2006) demonstrated a successful application of computer vision techniques for the automated detection of roadway shoulder activities. For a camera set on a light pole, a reliable coverage length of $30-60 \mathrm{~m}$ was obtained. The correct detection rate ranged from $80-100 \%$ for extended test duration of 103.2 hours.

### 2.4.5 The Next Generation Simulation (NGSIM) Program

The need for microscopic data for the development, calibration, and validation of traffic simulation models has been well recognized in the literature. With the advent of more capable computer vision techniques, it has become the method of choice for the Federal Highway Administration of the

United States for satisfying this long-awaited data need. The NGSIM program was instituted for the purpose of collecting microscopic vehicle tracks in order to develop and calibrate microscopic traffic simulation models. Video data was collected at three locations in California for duration of 1-3 days for 8 hours per day. The publicly available dataset contains vehicle tracks extracted from a 45 min subsequence. Vehicle tracks were extracted using an automated video analysis tool ( $\mathrm{NG}-\mathrm{VIDEO}^{7}$ ) which also allows human observer to correct for tracking errors.

The NGSIM public dataset of vehicle tracks spawned a series of important studies on traffic simulation, for example (Hamdar, Treiber \& Mahmassani 2009) (Izadpanah, Hellinga \& Fu 2009) (Kyte et al. 2009) (Thiemann, Treiber \& Kesting 2008). The demand for these studies has been latent for decades. It is foreseeable that NGSIM data will continue to support similar developments in the future. Despite the wealth of data afforded by the NGSIM program, the temporal and geographic scope of data collection is hardly representative of the tremendous variety of traffic conditions and driver behaviour patterns that are typically modeled using microscopic traffic simulation. The transferability of the NGSIM data to other locations and the representativeness of its sample size have not been thoroughly discussed in the literature. The development of more accurate and more efficient computer vision technologies will likely enable more extensive data collection and ultimately the development of enhanced traffic simulation models.

[^9]
### 2.4.6 Traffic Simulation

In one of the earliest applications of computer vision techniques for the development of car-following models, Ahmed (1999) developed an improved model for lane change behaviour at traffic intersections. In order to estimate various model parameters, real-world lane changing behaviour was recovered from video observations. The data extracted from video observations and used in model calibration and validation was mainly composed of vehicle tracks. The extraction of vehicle tracks was conducted using dedicated image processing software in automated as well as semiautomated fashion. The automated and semi-automated extraction of road user tracks was conducted within a pre-specified region of interest that covered $150-200 \mathrm{~m}$ of a four-lane carriageway for a total observation duration of 2 hours. Automated data collection was activated in uncongested conditions. The computer vision technology used in the concerned image processing software was mainly based on region-based vehicle tracking using background segmentation. Semi-automated data collection was conducted by aid of human observers in congested conditions. The pace of semi-automated data collection was 30 person-hours per one minute of video data. The developed and calibrated model exhibited improved performance over legacy models in predicting traffic volume passing through a 1.83 km highway segment in contrast with field observations.

While methodologically innovative, the generality of the developed model is challenged by the limited observational period of 2 hours. The likely explanation of this limitation is inherent to vehicle tracking approaches based
on background segmentation. These approaches typically provide poor tracking quality in congested traffic conditions.

Choudhury \& Ben-Akiva (2008) proposed and calibrated an intersection lane choice model. Various path-planning and maneuverability choices were used to constitute the model. Model development and validation were conducted using microscopic vehicle tracks from a 488m highway segment of US highway 101 as part of the FHWA's NGSIM project. Model calibration was conducted using aggregate vehicle trajectories recorded for 22 min . Model validation was conducted using aggregate vehicle trajectories recorded for subsequent 10 min . Promising results were obtained by comparing predicted traffic volume with observations. The observations period is however limited in time and the level of data aggregation underutilized the details abound in microscopic vehicle tracks.

One of the original studies on the calibration of pedestrian simulation models using microscopic pedestrian tracks was conducted using computer vision techniques (Hoogendoorn \& Daamen 2006). The authors argued that the reliance on macroscopic data, such as traffic flow, speed, and density, might not lead to the optimal model. For example, the heterogeneity of pedestrian characteristics cannot be represented by macroscopic data. The authors posed a more generalized argument that model calibration and validation should be conducted at the same level of aggregation that the simulation model uses to simulate pedestrian movement. Microscopic pedestrian tracks were extracted from an indoor video observation of pedestrian movement within a $10 \mathrm{~m} \times 4 \mathrm{~m}$ area and navigating a 1m wide bottleneck (Hoogendoorn \& Daamen 2003). A
particularity of the monitored pedestrians was their wearing colour helmets in order to aid tracking. The computer vision technology developed to extract microscopic tracks was a combination of background segmentation, colour detection, and clustering for pedestrian detection and Kalman Filtering to refine pedestrian tracks. Subsequent surveys of participating subjects showed that the experimental environment had not affected their movement that was arguably naturalistic. The generality of this study is limited due to the considered sample size and the particular navigational tasks presented to the participants. Furthermore, automated pedestrian tracking was conducted in an especially controlled setting, such as the use of coloured helmets as positional markers. Applying the same computer vision technique with the same level of success in open and mixed-use traffic settings is probably challenging.

Hoogendoorn et al. (2002) presented an original study on the use of computer vision techniques to extract microscopic vehicle tracks for in-depth behavioural analysis and improvement to car-following models. They argued that other data collection systems such as inductive loops, pneumatic tubes, and differential GPS analysis are incapable of extracting microscopic vehicle tracks with adequate accuracy. A high-definition digital camera was attached to a helicopter that hovered above the study area. Data collection lasted 2 hours and covered a 200 m long highway segment. After conducting image rectification to account for perspective, the background model was estimated as the median of each frame pixel value. After detecting isolated foreground components, vehicle blobs, a geometric model was estimated to encapsulate every blob. Validation was conducted on 45 s and 52 s video sequences with a
correct detection rate of $98 \%$ and $90 \%$ respectively. Using the same data, Ossen \& Hoogendoorn (2005) discovered significant heterogeneity in driver characteristics. The heterogeneity of driver characteristics was evident in the simulation model parameters and predicted driver behaviour. An important message taken from the aforementioned work was that heterogeneity of driver behaviour could only be explored in depth using microscopic vehicle tracks.

Previous studies performed at TU Delft ${ }^{8}$ constitute an important development of traffic simulation models by utilizing computer vision technologies. The length of the video sequences, the elementary use of background segmentation techniques, the special camera setting on a helicopter, and the strictly indoor monitoring of pedestrian movement are the main limitations in this work. While it is understandable that these special precautions were instituted to improve the tracking quality, it came at the expense of the transferability and cost of the data collection procedure. Attempts to address these challenges were made in this thesis.

Hoseini \& Vaziri (2006) proposed a driver behaviour model that combines lane-specific car-following models with lane change models. The model is arguably useful for driving cultures with weak or non-existent lane discipline. Using a cellular automate formulation, the proposed model places all movement directions along the same continuum that spans lateral and longitudinal movement. In order to calibrate and validate the proposed model, microscopic vehicle tracks were extracted from video observations.

[^10]The algorithm used is based on background segmentation with the background modeled as the mean pixel value. Different registration lines were defined to detect the presence of foreground components within the monitored highway segment. The observation period extended for a total of 1 hour. The calibrated model was validated against average speed and density observations and satisfactory predictive power was reported.

### 2.4.7 Other Applications

There is a growing body of work on the application of computer vision techniques in traffic control. Shelby et al. (2008) demonstrated the feasibility of using automated video detection technologies for the deployment of adaptive traffic signal system. The proposed system was reported to successfully monitor traffic conditions and automatically update signal timing. Based on field tests conducted in four different sites, there was a substantial reduction in vehicle delay compared to legacy systems. Sun \& Rescot (2008) studied the use of a novel video sensor, omni-directional camera, for the purpose of traffic monitoring at roundabouts. Single camera equipped with a dome mirror is capable of observing all turning movements within the roundabout. The accuracy of vehicle detection was approximately $90 \%$. Boillot, Midenet \& Pierrelée (2006) presented a novel traffic control algorithm for urban intersections based on accurate and real-time traffic data monitoring using computer vision techniques. Different traffic performance measures of the proposed traffic control algorithm were compared to reference traffic control systems proving a clear advantage.

Another important venue for data collection is the automated recovery of road features. This is particularly important for large-scale road network evaluation and for mapping applications. A novel image processing application was developed for the automated recovery of road surface information from video logs (Wu \& Tsai 2006). Pavement data included the automated detection of the location of shoulder marking, shoulder width, and travel lane width. In a sequel study, the authors demonstrated a successful application of the developed technology in the recovery of horizontal curve radii (Tsai, Wu \& Wang 2010). Another related application is the automated recovery of road inventory data. For example, several studies have been conducted on the subject of automated recognition of road signs, e.g., (Fang et al. 2004) (Hu \& Tsai 2009) (Wang, Hou \& Gong 2010) (Wu \& Tsai 2006) (Baro et al. 2009).

Other selected applications of computer vision techniques include vehicle reidentification for travel time estimation and origin-destination surveys (Sun et al. 1999) (O'Kelly et al. 2005), validation of traffic noise models based on providing more accurate traffic volume and speed measurements (Herman \& Nadella 2005), and anomaly detection (Velastin, Boghossian \& Vicencio-Silva 2006). The last application is the subject of Chapter 8 in this thesis.

### 2.5 Privacy Issues

Douma, Frooman \& Deckenbach (2008) argued that the emerging use of advanced computer vision technologies brings about novel legal challenges to the discipline of transportation engineering. The challenges arise from
privacy implications of misuse of or the right to retain gathered data. Privacy is a fundamental element of human rights. The common-law jurisdiction of the US and Canada potentially has no clear regulatory structure to assist concerned parties in utilizing the gathered data without infringing on the rights of privacy afforded to monitored road users (Douma, Frooman \& Deckenbach 2008). Data that would be gathered only through intrusion by legacy data collection methods can now be collected using advanced video surveillance techniques (McClurg 1995). The legal mechanism that protects against intrusion into a private sphere may not be precisely applicable to computer vision technologies even if the outcome is identical. The legal challenges that could reasonably arise are: vicarious criminal liability imputable to the institution overseeing the technology testing or deployment and tort liability due to privacy infringement.

Five aspects of privacy were reported by Douma, Frooman \& Deckenbach (2008) which were adopted from earlier work (Solove 2006). The aspects of life protected by privacy regulations are: spatial, behavioural, decisional, bodily, and informational. Spatial privacy refers to the delineation of private and public places. Behavioural and decisional privacy refers to the rights to protect the disclosure of certain actions or decisions respectively. Bodily privacy refers to a person's body. Informational privacy refers to the protection of personal information during data collection as well as the protection of gathered personal information. The aspects of privacy endangered by computer vision technologies, and arguably by all Intelligent Transportation Systems, are behavioural and informational.

The legal reasoning developed by Douma, Frooman \& Deckenbach (2008) in the case of Intelligent Transportation Systems in the United States will be adapted in the following sections specifically to computer vision techniques. It is plausible that the legal reasoning can be transferred to other legal systems and more proximately to other common-law jurisdiction such as Canada. The legal discussion of the police use of computer vision technologies for law enforcement is considered outside the scope of this review. Following are important doctrines that can be inferred from a number of US Supreme Court decisions (Douma, Frooman \& Deckenbach 2008):
a. The right to collect and use public data should be accompanied by a statutory or regulatory duty to prevent unwarranted disclosure.
b. The protection of personal privacy is comparable to the protection of life and property.
c. The concept of expected privacy limits the inclusion of the inner vehicle space under the privacy framework that covers homes and private areas. Expectation of privacy does not exist for anything put in "plain view". That plain view exclusion obviously covers vehicle and pedestrian movement from a place to another.
d. When data is in "general public use", the use of advanced computer vision technology is legitimate if the observed subjects are visible for the general public eye.
e. It is not clear whether the identity of the person driving a vehicle is a legitimate type of data.
f. When video data is collected by private entities, issues of privacy are covered by tort law. The relevant doctrine is the protection against
"false light", or unexpected publicity. In keeping with this doctrine, privately held data may not be shared with public entities or other private entities.
g. The Video Voyeurism Prevention Act enacted in 2004 in the United States ${ }^{9}$ criminalizes the taking or distributing of some types of photograph without the subject's consent. It appears that the main concern of this act however lies outside the realm of traffic monitoring or computer vision applications.

Extensive treatments of the subject of privacy in face of accelerated development in video surveillance technology have been developed in academia. The concept of privacy in public has been developed to grant some degree of privacy to public activities that are expected to be viewed by the public eye. The treatments also advocate a fuzzy definition of privacy or alternatively the degree of privacy. Many treatments reject the legal equivalence of unaided visual observation in the public and video surveillance on grounds of the level of details and permanent storage of video data.

In general, there is a reasonable agreement and sound judgement behind egregious privacy infringement using video cameras. The tort doctrine of "false light" can reasonably cover illegitimate sharing of privately held video data with the public. However, public inquiry into privately held video data as well as the right of private entities to retain video data collected from public venues is examples of issues that are not clearly covered by a legal

[^11]doctrine. Douma, Frooman \& Deckenbach (2008) note the similarity between the required legal developments to regulate the new technological developments in video surveillance and the development in copyright law to meet the then new technology of chip design.

In so far as development in computer vision technologies, traffic monitoring, and ITS technologies are anticipated, it is important to be aware of legal developments that determines what information is private and to what degree.


## Recovering Real-world Road User Positions

### 3.1 Background

The research work presented in this thesis relies mainly on video sensors as the main source of data acquisition. The use of video sensors to collect traffic data, primarily by tracking road users, has several advantages:

1. Video recording hardware is relatively inexpensive and technically less challenging to use than other positional sensors.
2. A permanent record of the traffic observations can be kept for archiving, future analysis, and human review.
3. Video cameras are often already installed and actively monitoring traffic intersections.
4. Video sensors cover a wide field of view. In many instances, one camera is sufficient to monitor an entire intersection, especially if the video sensor is placed at a vantage point.
5. Video sensors offer rich and detailed data of road user movements.
6. Techniques developed in the realm of computer vision renders automated analysis of video data feasible. Process automation has the advantage of reducing the labour cost and time required for data extraction from videos.

In a typical video sensor, observable parts of real-world objects are projected on the surface of an image sensor, in most cases a plane. An unavoidable reduction in dimensionality accompanies the projection of geometric elements (points, lines, etc.) that belong to a 3-dimensional Euclidian space (world space) onto a 2-dimensional image space. In order to recover the positions of various features that appear in the video, geometric elements must be mapped back from image space to the world space. What makes this step necessary is that metric measurements are only possible in terms of world space coordinates. The process by which this mapping is established is called hereafter camera calibration.

The recovery of real-world tracks of road users supports all forthcoming applications presented in this thesis. More precisely, the measurement of pedestrian walking speed requires accurate estimation of camera parameters given the relatively slow speed at which pedestrians move. Furthermore,
road user positions must be estimated at high accuracy in order to enable reliable measurement of their spatial and temporal proximity. As mentioned earlier, the practical significance of the camera calibration approach presented in this chapter reaches beyond the scope of applications in this thesis. In particular, conducting road user tracking in real-world coordinates can improve the accuracy of the tracking performance by correcting for perspective effect and other distortions due to projection on the image plane ${ }^{1}$.

Camera calibration concerns the estimation of camera parameters sufficient to back-project objects from the image space to a pre-defined surface in the realworld space. In general, the camera model can be parameterized by a set of extrinsic and intrinsic parameters. Extrinsic camera parameters describe camera position and orientation. Intrinsic camera parameters are necessary to convert observations to pixel coordinates. Typically, both extrinsic and intrinsic parameters are estimated in the calibration process.

Three major classes of camera calibration methods can be identified. First are traditional methods, based on geometric constraints either found in a scene or synthesized from a calibration pattern. The second class contains selfcalibration methods that utilize epipolar constraints on the appearance of features in different image sequences taken from a fixed camera location. Camera self-calibration is sensitive to initialization and can become unstable in case of a special motion sequence (Sturm 1997) and in the case where intrinsic parameters are unknown (Bougnoux 1998). Active vision calibration

[^12]methods constitute the third class for camera calibration. They involve calibration under controlled and measurable camera movements.

Only the first class of methods lends itself to traffic monitoring in which cameras have been fixed with little knowledge of their intrinsic parameters and control over their orientation. This is typically the case of already installed traffic cameras. The second class concerns self-calibrating cameras with prior knowledge of camera intrinsic parameters.

Examples of other classification methods include linear and non-linear, explicit and implicit (Wei \& Ma 1994). Non-linear methods enable a full recovery of intrinsic parameters, as opposed to linear methods. Both methods may be combined, e.g., in (Tsai 1987), by obtaining approximate estimates using linear methods with further refinements using non-linear methods. Inferring camera parameters from implicit transformation matrices obtained using implicit methods is susceptible to noise (Phong et al. 2005). Limiting calibration to extrinsic parameters gives rise to the topics of pose estimation (Zhang 1994).

Despite numerous studies on the topic of camera calibration, the following challenges can arise due to particularities of urban traffic scenes:

1. Many of the photogrammetry and Computer Vision techniques available in the literature do not apply due to differences in context, hardware, and target accuracy. Powerful and mature tools such as the self-calibrating bundle in the existing literature are not always possible to apply for relatively close-range measurements in urban traffic
scenes. this is especially the case for images taken by consumer-grade cameras containing noisy or incomplete calibration data (Remondino \& Fraser 2006). In addition, other methods in photogrammetry and Computer Vision depend on observing regularization geometry or a calibration pattern. In the typical cases where video cameras are already installed to monitor a traffic scene, or when only video records are available, this procedure cannot be directly applied.
2. Many of existing techniques rely on parallel vehicle tracks, in lieu of painted lines, for vanishing point estimation (Schoepflin \& Dailey 2003) (Kanhere, Birchfield \& Sarasua 2008). Vehicle tracks can be extracted automatically using computer vision techniques. These methods are particularly useful for self-calibration of pan-tilt-zoom cameras used for speed monitoring on rural highways. However, the vehicle motion patterns in urban intersections are not prevalently parallel. An example is shown in Figure 3.1a and 3.1b.
3. Much of the regularizing geometry in traffic scenes include elements such as road markings that may be altered in many ways. Regularities of traffic scenes provide a wealth of cues to inform the camera calibration process. In this study, one of the monitored traffic sites, BR, exhibited in Figure 3.2a was repainted after the orthographic image was taken, making point localization difficult. Using only point correspondences in this case is unreliable.
4. A significant number of camera calibration methods rely on the observation of one or more sets of parallel co-planar lines. By estimating the points of intersection of these sets of lines, i.e., vanishing
points located at the horizon line of the plane that contains these lines, camera parameters can be estimated. In urban traffic environments, the field of view of the camera can be too limited to allow the depth of view necessary for the accurate localization of the vanishing points, as shown in Figure 3.2a. To achieve desirable accuracy, camera calibration must include additional geometric information.
5. In many cases, cameras monitoring urban traffic intersections are already installed. Many of these cameras function as traffic surveillance devices, a function that does not necessarily require accurate estimation of road user positions. Given the installation cost and intended functionality, in-lab calibration of intrinsic parameters, e.g., using geometric patterns, can be difficult.

b) Pedestrian motion patterns


Figure 3.1 The difficulty of relying on the automated extraction of road user tracks. Figure a) shows the motion patterns of vehicles at a busy intersection in Chinatown, Oakland-California (sequence OK in Table 3.1). Figure b) shows pedestrian motion patterns.

In general, the proposed camera calibration approach was mainly motivated by issues encountered in case studies of video sequences presented in Chapters 4 to 8 as well as other research work on automated road safety analysis outside the scope of this thesis. The particular issues are: the repainting of traffic pavement marking, and the inability to estimate accurately vanishing point(s) because the field of view is too limited or nonlinear distortion is too pronounced. Table 3.1 provides a summary of the camera calibration case studies successfully carried out in the course of this thesis. More detailed description of the practical problems encountered in this study is provided in subsequent sections. As shown in Figures 3.1a and 3.1b, the difficulty of relying on road user trajectories is represented by the lack parallelism of pedestrian as well as vehicle motion prototypes. Many patterns represent turning movements and lane changing maneuvers that do not exhibit parallelism. Parallel vehicle tracks have to be hand-picked which is tantamount to manually annotating lane marking. Figure 3.1 b shows pedestrian motion patterns. It is evident that pedestrian tracks do not exhibit prevalent parallelism within crosswalks.

Table 3.1 Summary of case studies of camera calibration


1 The number of point correspondences available for calibration.
2 The number of line segments annotated in the image space with known real-world length.
3 The number of annotated pairs of lines in the image space the angle between which is known in world space.
4 The number of line segments annotated for equi-distance constraints. The endpoints of each line segment are annotated at two locations in the camera field of view.

As shown in Figure 3.2a, the estimation of the vanishing point location based on lane marking was unreliable. The obtained camera parameters were initially not sufficient to measure pedestrian walking speed in adequate accuracy. The integration of additional geometric constraints enhanced the estimates of the camera parameters and met the objectives of this application. Figure 3.2 b shows a sample frame from video sequence K 1 of traffic conflicts shot in Kentucky. Significant radial lens distortion is observed at the peripheries of the camera field of view. A reliable estimation of the vanishing point location requires the consideration of line segments that extend to the peripheries of the camera field of view. The significant curvature of parallel lines in these locations made the estimation of the vanishing point challenging.
a) Limited field of view.

b) Pronounced linear distortion.


Figure 3.2 An illustration of camera calibration issues that arise in urban traffic scenes. Figure a) shows a frame taken from video sequence BR-1 shot at Vancouver-British Columbia. Figure b) shows a sample frame from video sequence K1 of traffic conflicts shot in Kentucky.

Another challenge faced in this thesis is that some of the analyzed videos sequences were collected by other parties. The camera calibration methodology was motivated by positive particularities of traffic scenes. The geometric regularities abundant in traffic scenes offer geometric information besides the appearance of parallel lines that can increase the accuracy of camera calibration. The majority of the applications supported by this study involved the recovery of real-world coordinates of pedestrian tracks. Pedestrians move significantly slower than the motorized traffic, a characteristic that evidently required higher accuracy for camera parameters. Relying only on geometric information provided by parallel lines yielded camera parameters that provided unsatisfactory pedestrian speed estimates (Kanhere et al. 2007).

The work presented in this chapter concerns a robust camera calibration approach for traffic scenes in cases of incomplete and noisy calibration data. The cameras used in this study were commercial-grade cameras; most were held temporarily on tripods during the video survey time, others were already installed traffic cameras. A strong focus of this study is on the positional accuracy of road users, especially pedestrians. This was possible by relying on manually annotated calibration data, not automatically extracted vehicle tracks as is the case in automatic camera calibration, e.g., (Kanhere, Birchfield \& Sarasua 2008).

The uniqueness of this work lies in the composition of the cost function used for the estimation of camera parameters. The cost function contains information on various corresponding features that lie in both world and
image spaces. The diversity of geometric conditions constituted by each feature correspondence enables an accurate estimation of camera parameters. Features are not restricted to point correspondence or parallel lines, but extend to distances, angles between lines, and relative appearance of locally rigid objects. After annotating (manually defining) calibration data, a simultaneous calibration of extrinsic and intrinsic camera parameters is performed, mainly to reduce error propagation (Yu et al. 2009).

The remaining sections of this chapter describe, in order: a focused review of relevant previous work, the methodology of camera calibration, and a discussion of a number of case studies. Video sequences in these case studies were collected from various locations in the Downtown area of Vancouver, British Columbia, Oakland, California, and a signalized intersection in Kentucky.

### 3.2 Previous Work

There is an emerging interest in the calibration of cameras monitoring traffic scenes, e.g., (Worrall, Sullivan \& Baker 1994) (Pengfei 2004) (Li et al. 2007) (Masoud \& Papanikolopoulos 2007) (Kanhere, Birchfield \& Sarasua 2008) (Yu et al. 2009). An important advantage of traffic scenes for this purpose is that they typically contain geometric elements such as poles, lane marking, and curb lines. The appearance of these elements is partially controlled by their geometry, therefore providing conditions for estimating camera parameters. Common camera calibration approaches define different calibration conditions from a set of corresponding points, e.g., (Tsai 1987) (Zhang 2000),
from the appearance of geometric invariants such as parallel lines (Caprile \& Torre 1990), or from line correspondences (Dubrofsky \& Woodham 2008).

These approaches however overlook other geometric regularities such as road markings, curb lines, and segments with known lengths. The use of geometric primitives is becoming more popular, e.g., in recent work (Masoud \& Papanikolopoulos 2007) and citations therein. However, two main issues can arise in calibrating traffic scenes that cannot be addressed using existing techniques. First, most of the existing techniques construct the calibration error in terms of the discrepancy between observed and projected vanishing points. However, camera locations may be at significantly high altitude or its field of view too limited to reliably observe the convergence of parallel lines to a vanishing point. Finding initial guesses can be also challenging in such settings. Second, a detailed map or up-to-date orthographic image of the traffic scene may be unavailable. In this case, reliance on point correspondences is not possible. The proposed calibration approach draws the calibration information from the real-world lengths of observed line segments, angular constraints, and the dimension invariance of vehicles traversing the camera field of view.

### 3.3 Methodology

### 3.3.1 Camera Model

In the described camera calibration methodology, the canonical pinhole camera model is adopted to represent the perspective projection of real-world points onto the image plane (will also be called image space). A projective
transform that maps from a point $\mathbf{X} \in R^{\mathrm{n}}$ to a point $\mathbf{Y} \in R^{\mathrm{k}}$ can be defined by a $(k+1) \times(n+1)$ full-rank matrix. In the case of mapping from 3-D Euclidean space to the image plane, $k=2$ and $n=3$. In homogeneous coordinates, the projective transform can be represented by a matrix $\mathbf{T}_{3 \times 4}$ and a normalization term $\omega$ as follows:

$$
\omega\left[\begin{array}{l}
\mathbf{Y}  \tag{3.1}\\
1
\end{array}\right]=\mathbf{T}\left[\begin{array}{l}
\mathbf{X} \\
1
\end{array}\right]
$$

Similar to the column vectors in Equation 3.1, $\mathbf{T}_{3 \times 4}$ is defined up to a scaling factor while containing 11 degrees of freedom. In theory, a total of 11 camera parameters can be recovered: 6 extrinsic and 5 intrinsic. The matrix $T$ can be decomposed into two matrices such that: $\mathbf{T}=\mathbf{M} \times \mathbf{N}$, where matrix $\mathbf{N}_{4 \times 4}$ maps from world coordinates to camera coordinates (composed of intrinsic parameters), and matrix $\mathbf{M}_{3 \times 4}$ maps from camera coordinates to pixel coordinates (composed of extrinsic parameters). Two linear intrinsic parameters besides a non-linear parameter are primarily considered in the proposed approach. An additional non-linear parameter, radial lens distortion, is calibrated for the purpose of being used as an initial estimate of the set of calibrated linear camera parameters. Knowledge of extrinsic camera parameters, comprising 3 rotation angles and a translation vector, is sufficient for generating $\mathbf{N}$. Matrices $\mathbf{M}$ and $\mathbf{N}$ are calculated as follows:

$$
\mathbf{N}=\left[\begin{array}{cc}
R_{3 \times 3} & t_{3 \times 1}  \tag{3.2}\\
0_{1 \times 3} & 1
\end{array}\right] \quad \mathbf{M}=\left[\begin{array}{cccc}
f_{y} & -f_{x} \cot \theta & u_{o} & 0 \\
0 & \frac{f_{y}}{\sin \theta} & v_{o} & 0 \\
0 & 0 & 1 & 0
\end{array}\right]
$$

where $f_{x}$ and $f_{y}$ are respectively referred to as the horizontal and vertical focal lengths in pixels, $\theta$ is the angle between the horizontal and vertical axes of the
image plane, $R$ is three-dimensional rotation matrix, $t$ is translation vector, and $\left(u_{o}, v_{o}\right)$ are the coordinates of the principal point. The principal point is assumed to be at the centre of the image in the video sequence.

The non-linear camera parameter considered in this methodology is the radial lens distortion parameterized by the distortion coefficient $k$. The selection of this non-linear parameter was motivated by the pattern of visual distortion of linear road marking visible in cases $\mathrm{K} 1 \& 2$. The projection of points taking into account radial lens distortion is represented by the second-degree polynomial form shown in the following set of equations:

$$
\begin{align*}
& \dot{x}=x *\left(1+k r^{2}\right) \\
& \dot{y}=y *\left(1+k r^{2}\right) \tag{3.3}
\end{align*}
$$

where $(x, y)$ are image space coordinates measured in pixels, $(\dot{x}, \hat{y})$ are the image space coordinate corrected for radial lens distortion and $r$ is the uncorrected distance in pixels from the principal point to a point on the image space.

### 3.3.2 Cost Function

There is no universally recognized cost function for errors in camera models (Masoud \& Papanikolopoulos 2007). Yet, there are stable formulations developed in the literature, e.g., in (Weng, Cohen \& Herniou 1992), for calibration data consisting of point correspondences. It is however more complicated to construct a proper cost function if the calibration error is based on different types of geometric primitives. A proposed cost function is argued to satisfy the following conditions:

1. Uniformly represent error terms from different geometric primitives, i.e., consistent weights and units. This is possible if the cost function is constructed in real-world coordinates.
2. Be perspective invariant, i.e., not sensitive to image resolution or camera-object distance.

It is also desirable that a cost function be meaningful in further image analysis steps so that keeping account of error propagation is possible. For example, it may be desirable to compare the estimated positional error due to videobased tracking to the positional error due to camera calibration. Satisfying the first condition in linear algebra, and without special mapping, entails some assumption and/or approximation. Following are the set of conditions proposed in this approach to represent a calibrated camera model:

1. Point correspondences $\left(\mathrm{CD}_{\mathrm{p}}\right)$. Matching features are points annotated in the image and world spaces. This condition matches the backprojection of points from one space to their positions in a current space. For unit consistency, point positions in world space are compared to the back-projection of points from the image space to the world space.
2. Distance constraints $\left(\mathrm{CD}_{\mathrm{d}}\right)$. This condition compares the distance between the back-projection of two points to the world space and their true distance measured from an orthographic map or in-field.
3. Angular constraints $\left(C D_{a}\right)$. This condition compares the true angle between the two annotated lines to that calculated from their backprojection to world space. Special cases are angles of $0^{\circ}$ in case of
parallel lines, e.g., lane markings or vertical objects, and $90^{\circ}$ in case of perpendicular lines, e.g., lane marking and stop lines.
4. Equi-distance constraints $\left(\mathbf{C D}_{\text {ed }}\right)$. This condition compares the realworld length of line segments observed at different depths of view. This condition preserves the back-projected length of a line segment even if it varies in the image due to perspective effect.

The following cost function is composed of four components, each representing a condition which is an implicit function of the vector of camera parameters $\boldsymbol{\Pi}$ :

$$
\begin{equation*}
f(\boldsymbol{\Pi})=\sum_{\left.i \in \mathbf{C D}_{p}, j \in \mathbf{C D}_{d}, k \in \mathbf{C D}_{a}, m \in \mathbf{C D}_{e d}\left\|\Delta \mathbf{P}_{i}\right\|_{2}{ }^{2}+\left(\Delta s_{j}\right)^{2}+\left(\bar{l} \tan \Delta a_{k}\right)^{2}+\left(\Delta l_{m}\right)^{2},{ }^{2}\right)} \tag{3.4}
\end{equation*}
$$

where,

1. $\mathbf{C D}_{p}, \mathbf{C D}_{d}, \mathbf{C D}_{a}$, and $\mathbf{C D}_{e d}$ are respectively the sets of calibration pointdifference, distances, angular constraints, and equi-distance constraints.
2. $\left\|\Delta \mathbf{P}_{i}\right\|_{2}$ is the real-world distance between observed and back-projected calibration points in the ith set of point correspondences,
3. $\Delta s_{j}$ is the difference between observed and projected distances in the $j^{\text {th }}$ set of distance correspondence,
4. $\bar{l}$ is the average length of the back-projected line segments on the pair of lines that defines the angular constraint,
5. $\Delta a_{\mathrm{k}}$ is the difference between annotated and calculated acute angle between the $\mathrm{k}^{\text {th }}$ back-projected pair of line segments that defines the angular constraint, and
6. $\Delta l_{m}$ is the difference between the real-world length of a line segment calculated at two locations with different depth of view. This can be typically obtained by measuring the distance between two points on a vehicle traversing a traffic intersection.

The back-projection of points in the image space, i.e., mapping from image space to world space, is performed efficiently using the homography matrix $\mathbf{H}$ that corresponds to a set of camera parameters $\Pi$. A least square estimation of the homography matrix is conducted using four points selected from $C$, using $\mathbf{X}$. If the non-linear camera distortion parameter is estimated, backprojection using the homography matrix is not accurate. In this case, backprojection is cast as a minimization problem, such that the reprojection of the estimated world-space position, from world space to image space, achieves a minimum difference from the annotated image position. The initial estimate of this minimization problem is the world-space position of a point using homography. A basic Quasi-Newton non-linear optimization is sufficient for accurate estimation of the world-space position.

The cost function component that represents angular constraints has the useful property of being proportional to the length of the annotated line segments that define the angular constraint. This assigns larger weight to angles more precisely defined using long edges.

The cost function presented in Equation 3.4 represents linear discrepancies between observed and back-projected geometric primitives, all expressed in real-world unit distance. This construction of the cost function clearly meets the previously proposed conditions. It is noteworthy that the construction of
the cost function in pixel coordinates, commonly adopted in the literature, is significantly cheaper to compute than the proposed cost function. In the latter case, point projection to image space is a closed-form operation. The proposed camera calibration approach is designed as an accurate one-time operation to support data extraction from video surveys in which computational efficiency is of lesser importance. In addition, the expression of the projection error in pixel coordinates is implicitly biased toward features closer to the camera (represented by more pixels). This may not be desirable in all applications. For example, the case study based on the video sequence K1, shown in Figure 3.2 b, focuses on events that take place in the furthest intersection approach.

### 3.3.3 Implementation Details

The three intrinsic camera parameters that are estimated through calibration are focal length, skew angle, and radial lens distortion. The extrinsic parameters are the translation and rotation (six parameters) of the camera coordinate system from the world coordinate system. The selection of these camera parameters yields more accurate results than if optimization is conducted for each element of the transformation matrices $\mathbf{M}$ and $\mathbf{N}$ (Equation 3.2).

The minimization of the cost function in Equation 3.4 over the camera parameters is performed using the Nelder-Mead (NM) simplex algorithm (Nelder \& Mead 1965). This algorithm was selected over the commonly used Levenberg-Marquardt (LM) (Marquardt 1963) which failed in some cases to converge when the initial estimate of the camera parameters was not in close
proximity to the globally minimizing set of parameters. When both converged, NM was consistently more computationally expensive. Nevertheless, computational cost is of lesser importance for the one-time high-precision applications targeted by this approach.

The initial estimates for the case studies shown previously in Table 3.1 were obtained using an estimate of the camera position in an orthographic map of the monitored traffic intersections. Estimates were also provided for the camera height and of the location of the back-projection of the principal point on the road surface. The estimate for the focal length was found using previous information and assuming away perspective. Obtaining an accurate initial estimate of the focal length and camera height proved difficult and was in most cases far from the calibrated value. A similar issue was encountered for estimating the camera height of video sequences which were not collected by the authors (sequences K1, K2, and OK). The calibrated camera height for K1 and K2 were 11.5 m and 10.9 m respectively, while their initial estimate was 5.5 m . The implementation of this method was conducted in MATLAB (Mathworks 2010). A toolbox was developed to annotate the calibration data, find initial estimates, conduct the camera calibration and visualize the calibration results. The following section provides a review of four case studies in which the proposed camera calibration approach provided adequate estimates of camera parameter. The intended applications were carried out successfully as described in Chapters 4 to 8 .

### 3.4 Case Studies

The four case studies analyzed using the proposed camera calibration approach are summarized in Table 3.1. Camera calibration was conducted for video sequences collected from the downtown area of Vancouver, British Columbia (video sequences 1-4 from site BR and sequence PG), Chinatown in Oakland, California (OK), and an unidentified intersection in Kentucky (K1 and K2). When possible, real-world data was extracted from an orthographic image from Google Maps and in-field distance measurements.

### 3.4.1 Annotation of Calibration Data

Corresponding points were annotated in image and world spaces. The realworld coordinates of points in the image space can be calculated from their positions on the world map. The true lengths of line segments which constitute distance and equi-distance conditions were calculated from the orthographic image. In case of sequences BR-1:4, true lengths of line segments were collected by in-field measurements (total of 21 measurements). This was necessary to obtain camera calibration with accuracy that supports the measurement of pedestrian walking speed (refer to Table 3.1). Pairs of lines which constitute the angular constraints were annotated in the image space. These lines are parallel lane markings, parallel light poles and road-side signs, and perpendicular road markings. Figure 3.3 shows the calibration data for sequence $B R-2$.

### 3.4.2 Validation

The developed algorithm was compared against the well-known Tsai algorithm for camera calibration. The cost function components in Equation 3.4 were incrementally introduced. Also, an implementation of Tsai's method (Tsai 1987) was used to estimate the camera parameters based on the set of point correspondences obtained for each scene. A supplementary in-site distance measurement was performed at scenes BR-1-4. The sizes of the different calibration datasets for each scene are shown in Table 3.2. Root Mean Square Error (RMSE) was calculated by leaving out one feature observation, from sets $\mathbf{C D}_{p}, \mathbf{C D}_{d}$, and $\mathbf{C D} \mathbf{D}_{a}$ at a time and adding up the error from each feature observation. The total number of iterations required for each scene is the maximum of the number of data points in sets $\mathbf{C D}_{p}, \mathbf{C D}_{d}$, and $\mathbf{C D}_{a}$. For example, the number of iterations is 13 for BR-1 and 12 for BR-2.

Table 3.2 RMSE calibration error using Tsai Algorithm (Tsai 1987) and different cost function compositions. The numbers of point correspondence, distance, and angular constraints are in columns $\mathbf{C D}_{\boldsymbol{p}}, \mathbf{C D}_{\boldsymbol{d}}$, and $\mathbf{C D}_{a}$ respectively.

| Dataset | Tsai | Cost function component |  |  | \# Data Points |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Point <br> correspondences | Distance <br> Constraints | Angular <br> Constraints | $\mathbf{C D}_{p}$ | $\mathbf{C D}_{d}$ | $\mathbf{C D}_{a}$ |
| BR-1 | 0.482 | 0.689 | 0.606 | 0.583 | 13 | 6 | 4 |
| BR-2 | 0.463 | 1.099 | 0.662 | 0.557 | 11 | 12 | 6 |
| BR-3 | 2.040 | 2.329 | 0.458 | 0.528 | 5 | 10 | 5 |
| BR-4 | 0.597 | 2.204 | 0.597 | 0.322 | 9 | 10 | 3 |
| PG-1 | 0.132 | 0.099 | 0.0929 | 0.094 | 22 | 2 | 2 |

Calibration results are presented in Table 3.2. When relying only on point correspondence, Tsai's algorithm outperforms our algorithm except for PG-1. However, the accuracy of the camera calibration improves significantly after the integration of distance and angular constraints. The average reduction in RMSE based on point correspondence training data to all geometric primitives is $42 \%$. In three out of five scenes, the accuracy of our estimates was better than those obtained using Tsai's algorithm.


Figure 3.3 Calibration data for video sequence BR-2. Point correspondences are annotated with their serial numbers. Points marked with red are calculated and points in blue are annotated. The segments in red define the distance conditions. The segments in blue define pairs of lines for angular conditions. Figure a) shows the calibration data (points, and lines) in the image space. Figure b) shows the back-projection of the calibration data to world-space.

The performance at scenes $B R-3$ and $B R-4$ is noteworthy since a limited number of calibration points were available at these scenes. The addition of the angular constraints in most cases reduces RMSE, in exception of scene BR3. An idiosyncrasy of this scene is the definition of its angular constraints using long edges. It is possible that this exception is due to the appreciable contribution of the angular constraints to the cost function. Figures 3.3 shows observed and re-projected calibration data in scene BR-3.

### 3.4.3 Effect of Different Cost Function Components

In order to investigate the effect of using a mix of geometric primitives, the cost function components in Equation 3.4 were incrementally introduced. The sizes of the different calibration datasets for each scene are shown in Table 3.1. Figure 3.4a shows the reduction in back-projection error for sequences BR-1:4 and PG with the introduction of additional cost function components. In order to investigate the effect of the equi-distance constraint, the video sequence OK was selected. This sequence has the largest number of calibration data points. In addition, a special challenge faced in this case study was that the video sequence was observed from an unknown camera setting location. It was also attempted to investigate the improvement in estimation accuracy over features obtained only from the image space, as is the case with the mainstream vanishing point methods for estimation. Figure 3.4 b shows the back-projection error using different compositions of the cost function. The error was calculated in terms of the difference between the calculated and true lengths of a validation set of 12 line segments. These line segments were not included in the calibration data set.

There is a clear advantage of using calibration data in addition to estimates of point correspondences (four corner points which coordinates estimated based on an assumed lane width of 3.5 m ) referred to as case 1 in Figure 3.4b. There is also an advantage over the use of angular constraints only (case 2) which is analogous to camera calibration based on vanishing point estimation. The addition of all cost function components (case 4) provides however only marginal improvement compared to using point correspondences only (case 3). This likely occurs because of the abundance of accurately localized point correspondences in this video sequence. The effect of the addition of cost function components was more evident in sequences K 1 and K 2 . The camera calibration for these sequences was the most challenging. The video sequence, collected from an unidentified site in Kentucky, contains a valuably large number of vehicle-vehicle traffic conflicts that were analyzed in a different study. The effect of non-linear lens distortion was visible for almost all observed line segments.

As shown in Figure 3.5a, there is a clear advantage of adding all cost function components. The back-projection error was calculated based on the difference in the calculated real-world length of line segments observed from two different cameras for the same site, corresponding to datasets K 1 and K 2 . Figure 3.5 b shows the validation results of camera calibration conducted using the complete set of cost function components (case 5).
a) RMSE for BR-1:4 and PG for different cost function compositions.

b) Linear back-projection errors for scene OK


Figure 3.4 Examples of reduced camera calibration error due to the inclusion of various cost function components. Figure a) shows the RMSE error of test sets BR1:4 and PG. Figure b) shows the back-projection error in terms of the difference between the true and calculated lengths of 12 line segments in sequence OK. The 12 segments were not used in the calibration. The length difference is normalized by the segments length: Back projection error $=\left|\mathbf{L}_{\text {true }}-\mathbf{L}_{\text {calculated }}\right| / \mathbf{L}_{\text {true }}$.

a) Back-projection errors for K1 and K2.
b) Discrepancy in linear measurements from the two cameras K1 and K2.


Figure 3.5 Evidence of improvement in calibration accuracy by including different cost function components for video sequences K1 and K2. Figure a) shows the back-projection error measured as the difference between the realworld lengths of a total of 20 line segments calculated from two camera settings at K1 and K2. The discrepancy in the lengths of the validation line segments were normalized by each line segment length (average 12.57 m ). Figure b) shows the lengths of the validation line segments for case 5. Refer to Figure 4 for the indication of cases 1:5.

### 3.4.4 Visualization of Results

In order to visualize the accuracy of the estimated camera calibration parameters, a reference grid is depicted in Figure 3.6 for sequences BR-2, PG, and OK. The reference grids for sequences K 1 and K 2 are shown in Figure 3.7. For sequences K1 and K2, the calibrated radial lens distortion parameter could explain the apparent distortion of the boundaries of the closer sidewalk. The distortion at the further sidewalks could not be completely captured. This demonstrates that additional non-linear parameters are required to capture other types of image distortion evident in this video sequence.

Sample results of applications supported by the estimated camera parameters for these case studies are shown in Figure 3.8. In this figure, sample pedestrian and vehicle tracks displayed in both world and image spaces are exhibited. The selected road users are involved in traffic conflicts. The positional accuracy of tracking using the estimated camera parameters enabled successful detection and severity measurements of these events in an automated fashion.
a) BR-2 (image grid)

c) PG (image grid)

e) OK (image grid)

b) BR-2 (world grid)

d) PG (world grid)

f) OK (world grid)


Figure 3.6 Reference grid for video sequences BR-2, PG, and OK, overlaid on frames of the video sequence and orthographic images. The grid spacing is 1 m and the height of the vertical reference lines (depicted in blue) is 4.0 m . Sequences BR-1 and BR-3:4 are recorded at the same site (BR) with different fields of view.
a) The intersection in Kentucky as it appears from the first camera (K1).

b) The intersection in Kentucky as it appears from the second camera (K2).


Figure 3.7 Reference grids for video sequences K1 K2. The non-linear calibration parameters could capture the distortions at the closer sidewalk of sequences K1 and K 2 . The grid spacing is 2.0 m and the height of the displayed vertical line segment (depicted in blue) is 4.0 m .


Figure 3.8 In this traffic safety application, accurate road user tracks are required to measure their temporal and spatial proximity. Left are the back-projected pedestrian and motorist tracks. Right are the CV-based tracks of the interacting road users. Figures a) and b) show the world and image space of video sequence PG. Figures c) and d) show the world and image space of video sequence OK. Road user depiction in Figures

### 3.5 Conclusions

Camera calibration is necessary for recovering metric information from video sequences. Despite the development of successful methods, current approaches do not address the critical issues that arise when monitoring traffic scenes, especially when high camera calibration accuracy is required. The methodology presented in this chapter was fundamental to video analysis conducted in subsequent parts of this thesis.

In this chapter, a robust methodology for camera calibration was developed and tested. The proposed methodology successfully tackled all practical challenges faced in video analysis conducted in subsequent parts of the thesis. As supported by the reported results, the composition of the cost function representing calibration error proved to enhance the accuracy of the calibration process.

One of the peculiarities of camera calibration noticed in this work is the nonmonotonous effect of introducing different cost function components as is shown in Table 3.2. This peculiarity was also noticed while generating the error reduction for other case studies. Intuitively, the introduction of new cost function components should reduce the estimation error as is shown in Figures 3.4 and 3.5. However, in some cases the introduction of a particular cost function components causes the accuracy of estimation to deteriorate. This is an issue that is worth further investigation. This investigation however lies outside the scope of the research problem that defines this chapter and was therefore delegated for future research.

The formulation of this cost function in a linear algebra entails assumptions regarding the angular constraints. An important extension of this work is the reformulation of the cost function using geometric algebra in which different geometric elements can be uniformly represented. Further improvements to the proposed methodology should consider the inclusion of additional nonlinear parameters such as tangential distortion.

## Automated Pedestrian Data Collection using Computer Vision Techniques

### 4.1 Background

This chapter presents the details of a study on the application of computer vision techniques for the automated measurement of pedestrian microscopic data. The main context in which microscopic pedestrian data was used is the measurement of pedestrian walking speed. Subsequence sections present more details about the motivation of this research work and challenges facing current methods used for measuring pedestrian walking speed.

Walking is the most basic means of travel and is one of the key activities in a sustainable, healthy, resource-efficient and liveable urban environment. New
urban planning concepts have redefined the function and mode-assignment of streets by emphasizing walkability and recognizing the pedestrian as a key road user (Greenberg 2005). AASHTO describes walking activities as the lifeblood of urban streets (AASHTO 2001). The new functional definition of streets entails changing industry standards and professional practice in order to accommodate pedestrian needs for safety and mobility. The reviving emphasis on walking and other non-motorized means of travel is part of a larger theme that advocates the creation of a more sustainable transportation system. The emergence of this theme is likely a public response to global changes in energy resources, a desire for improving the quality of life in urban areas, and a growing environmental awareness. These drives of public support have not shown signs of fading and will likely continue in the future. In addition, the emerging research focus on pedestrians comes as a response to demographic changes.

Along with other developed countries, the population of Canada is aging. Percentage of seniors (65+) in Canada increased from $13 \%$ in 2001 to $13.7 \%$ in 2006 and is projected to reach $23-35 \%$ in 2031 and $25-30 \%$ by 2056 (Martel \& Malenfant 2006). Seniors in British Columbia represent $14.6 \%$ of the total population in 2006, one of the highest in Canada (Martel \& Malenfant 2006). Similar national trends can be observed in the United States although the population is slightly younger (Shrestha 2006).

The effect of demographic changes on the design of pedestrian facilities can be understood by studying the particularities of the older age groups. Older pedestrians have longer information processing and perception times (Fugger
et al. 2000), need generally more illumination (Fozard 1981), are more prone to overestimating the dimensions of crossing facilities as a result of misreading visual cues (Guerrier \& Sylvan 1998). Moreover, older pedestrians are more likely to be involved in accidents (Harkey 1995). Aging brings about general change in physical attributes (Pauls 2008) and in particular walking speed (Fitzpatrick et al. 2006).

The measurement of walking speed of older pedestrians has been an important topic in the literature of pedestrian studies. Newer releases of standard design guides, e.g., MUTCD, are in the process of adopting design parameters that consider more aspects of the elderly pedestrian. Studies in the literature suggest adopting a continuum of design parameters, e.g., walking speed, based on the expected age distribution among pedestrians (Fitzpatrick et al. 2006)(Highway Capacity Manual 2000). Further studies are required to capture the differences among senior subgroups as some studies suggest that this age group is not homogenous as assumed by past studies of walking speed (Stollof, McGee \& Eccles 2007).

### 4.2 Issues with Pedestrian Data

Despite the growing importance of non-motorized traffic and in particular pedestrians, these modes of travel are in general overlooked, and understudied relative to vehicular traffic. For example, current trip counts capture $16-33 \%$ of actual non-motorized trips (Litman 2003), while collecting reliable nonmotorized traffic information is especially challenging (Weinstein \& Schimek 2005). Planning for pedestrian facilities and modelling of pedestrian demand
are areas of research that are yet to be developed to a level that matches vehicular traffic (Pulugurtha \& Repaka 2008). In general, there is a poor integration of pedestrians to current transportation networks and a challenged interlinking with activity areas (James \& Walton 2000). For example, vehicular traffic is traditionally the main focus of level of service improvements, with little attention to negative impact on modes that share the same transportation facility (Milam \& Mitchell 2008). The trade-off between improving the level of service for motorized traffic and the related impact on non-motorized transport is often ignored or cursorily studied in the current state of practice. The effect of permitting longer pedestrian crossing interval times on motorized traffic delay was analyzed in hypothetical case studies, e.g., (Kim et al. 2005). However, little is known about the measures to alleviate motorized delay resulting from the adoption of slower normative walking speed, especially in cases of high motorized traffic volume or short signal cycles.

The limitation in collecting pedestrian data inhibits a better understanding of many pedestrian research issues. For instance, data is required to capture pedestrian response to longer pedestrian crossing intervals, in particular whether this traffic control measure will result in slower crossing speed. Microscopic observational data is required to investigate the ability of individual pedestrians to adapt their walking speed in response to change in signal indication, in anticipation of potential conflict with motorized traffic (Gates et al. 2006) (Stollof, McGee \& Eccles 2007), or in response to external stimuli (Kim et al. 2007). In addition, microscopic pedestrian observations can provide valuable insight for pedestrian modelling, e.g., inter-person spacing and pedestrian maneuvering (Kerridge \& Chamberlain 2005) and obstacle
navigation (Willis et al. 2004). Although at a relatively advanced stage in theory and analysis, pedestrian simulation models are generally based on limited understanding of microscopic pedestrian behaviour (Willis et al. 2004) and limited validity because of a lack of real data (Kerridge \& Chamberlain 2005) (Antonini et al. 2006).

Collecting positional data for pedestrians is particularly challenging due to the less organized nature of pedestrian traffic compared to vehicular traffic (Hoogendorn, Daamen \& Bovy 2003). The main methods for collecting this data can be classified into: manual field observations, manual observations from videos, semi-automated video analysis, and automated video analysis.

Manual field observation, which is the most common method of pedestrian data collection, is in general more expensive, error-prone, and time consuming compared to video analysis (Kerridge et al. 2004). The use of video sensors for measuring pedestrian walking speed has several advantages. First, it captures naturalistic pedestrian movement with limited risk of stirring the attention of observed subjects, who may behave unnaturally if felt being watched. Other advantages include the relative ease of installation, the richness of the data that can be extracted (i.e., complete trajectories), the large area that can be covered and their low cost. However, manual video observations are also time consuming and error-prone.

Semi-automated analysis, or time-lapse analysis, of pedestrian movement involve the use of image processing tools to manually mark or track pedestrians in a sequence of video images, e.g., (Lam \& Cheung 2000) (AlGhadi, Mahmassani \& Herman 2002). Manual operations in semi-automated
video analysis can be laborious and limited in terms of data volume to be analyzed compared to automated methods. Automated video analysis involves the use of computer vision techniques and can overcome many shortcomings associated with manual field observations and manual video analysis.

The current practice of observing pedestrian walking speed generally depends on manual observation of pedestrian crossing time. As discussed earlier, manual techniques face several accuracy and efficiency challenges. In order to cope with the increasing demand for studying pedestrian movement, to accommodate changes in pedestrian characteristics, and to improve signal design, automated techniques need to be further developed. The primary objective of this chapter is to document the development and testing of a prototype system that is capable of extracting real-world pedestrian tracks from a video taken at traffic facilities.

The main purpose of the video analysis system is to enable large volume and accurate recording of pedestrian walking speed. The study is unique in regard to the developed video analysis technique as well as in testing the developed system under different conditions of lighting, crowdedness, and traffic mix in an open and uncontrolled environment. This chapter discusses the technical issues that arose during the system development along with techniques for resolving these difficulties. Walking speed measurements automatically calculated by the system were validated in contrast with walking speeds extracted by human observers. The system accuracy in automatically measuring pedestrian speed was satisfactory and provided support and reliability for analysis results. A case study is introduced for pedestrian
movement in a main commercial corridor in the Downtown area of Vancouver, British Columbia. The case study was validated and demonstrated satisfactory accuracy of the system. A statistical analysis of the case study results is presented at the end of this chapter along with reports on the findings.

### 4.3 Previous Work

### 4.3.1 Studies on Walking Speed

Walking speed is a fundamental property of pedestrian flow that is important in a wide range of applications. The ability to predict pedestrian movement under different external circumstances and individual attributes of pedestrians is an important underpinning for the design of pedestrian facilities (Al-Azzawi \& Raeside 2007). Many types of transportation studies require prior knowledge of walking speed, such as: planning and management of crowd movement, developing pedestrian simulation models, estimating facility level of service, and designing pedestrian signals. There are various contextual and individual variables which influence walking speed. Numerous studies in the literature dealt with the determinants of walking speed based on quantitative and/or theoretical treatments. Examples of studies that involved substantial walking speed observations are presented in Table 4.1. Other studies discussed the effect of the following factors on walking speed: carried object (Morrall, Ratnayake \& Seneviratne 1991), area type (Al-Masaeid, Al-Suleiman \& Nelson 1993), crowd density(Fruin 1970)(Virkler \& Elayadaph 1994) (Goh \& Lam. 2004), temperature (Walmsley \& Lewis 1989), noise (Boles \& Hayward 1978), city size temperature (Walmsley \& Lewis 1989), feeling of insecurity or being monitored (Smith \&

Knowles 1979), crossed lane use (Bowerman 1973), platoon movement (Golani \& Damti 2007), and whether walking is indoors or outdoors (Lam \& Cheung 2000). For a comprehensive review of the evolution of walking speed refer to cited works in (Knoblauch, Pietrucha \& Nitzburg 1996)(LaPlante \& Kaeser 2004) (Fitzpatrick et al. 2006)(Hoogendoorn \& Daamen 2006) (Stollof, McGee \& Eccles 2007). However, as suggested from Table 4.1, none of the key studies in the literature has made use of automated pedestrian speed collection. Methods used in practice are unable to capture microscopic changes in speed and position (Shi et al. 2007). This remark highlights shortcomings in the current techniques used in the practice of pedestrian data collection and signifies the practical need for this research.

Table 4.1 Sample of previous studies on pedestrian walking speed

| Study | Reported 15th Percentile ( $\mathrm{m} / \mathrm{s}$ ) | Reported 50th Percentile ( $\mathrm{m} / \mathrm{s}$ ) | \% difference from standards ${ }^{1}$ | Sample | Method ${ }^{3}$ | Significant <br> Factors | Insignificant Factors |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| (Bowman \& Vecellio 1994) | - | 1.04 | -20\% | 360 | 1 | 1,7 | 7 |
| (Dahlstedt 1978) | 0.67 | - | -26\% | N/A | 1 | 1 | - |
| (Fitzpatrick et al. 2006) | 0.9 | - | 0\% | 2552 | 2 | 1 | 5,8,6,2 |
| (Guerrier \& Sylvan 1998) | 0.66 | - | -27\% | 263 | 2 | 1 | - |
| (Gates et al. 2006) | 0.92 | - | 2\% | 1947 | 1,2 | 1,5,6 | 2 |
| (Hoxie \& Rubenstein 1994) | 0.86 | - | -4\% | 1229 | 1 | 1 | - |
| (Hui et al. 2007) | - | 1.22 | -6\% | 1882 | 2 | 1,2 | - |
| (Knoblauch, Pietrucha \& Nitzburg 1996) | 0.97 | - | 8\% | 7123 | 1 | 1,3 | 2,4-8 |
| (Lam \& Cheung 2000) | Model | - | N/A | 16453 | 3 | 4,6,9,10,11 | - |
| (Lam, Morrall \& Ho 1995) | Model | - | N/A | N/A | 2 | 4,6,9,11 | - |
| (Lee and Lam 2006) | Model | - | N/A | 14886 | 3 | 4,11 | - |
| (Montufar, Michelle and Nakagawa 2007) | 0.88 | - | -2\% | 1792 | 1 | 1,3,4 | - |
| (Stollof, McGee \& Eccles 2007) | 1.03-1.16 | - | -64\% | 2603 | 1,2 | 1 | - |
| (Tarawneh 2001) | 0.97 | - | 8\% | 3500 | 1 | 1 | 2,4,5 |
| (Ye et al. 2008) | Model | - | N/A | 2089 | 2 | 11 | - |

1 We refer to the most recent recommended updates for MUTCD as standards
2 Significance is statistical and/or practical. The assessment of the practical significance of walking speed factors was either directly reported in the studies or performed by the author. Insignificant factors were treated in a similar manner.
3 Number indications: 1) Field observations, 2) Manual video analysis, 3) Semi-automated video analysis, 4) Automated analysis (None found)
4 Number indications: 1) Age and/or walking problems, 2) Gender, 3) Season/weather (precipitation, snow, temperature), 4) Pedestrian facility type (Crosswalk, sidewalk, stairway, midblock crossing, experiment setting), 5) Group size, 6) Traffic control (Pedestrian signal type, unsignalized, speed limit), 7) site specifications (Marking, geometry, road classification, median, lane usage), 8) Vehicular traffic, 9) Indoor/outdoor, 10) Activity area (Shopping, commercial, recreational, etc.), 11) Pedestrian traffic characteristics (flow, density, directional split).

### 4.3.2 Techniques of Measuring Walking Speed

Automated pedestrian data collection relies mostly on video sensors, including infrared and thermal imaging cameras (Kerridge et al. 2004), as well as sometimes on Light Detection and Ranging (LIDAR) sensors (Cui, Zhao \& Shibasaki 2006). This study advocates the use of video cameras (in the visible spectrum) because alternative sensors are still more expensive and more widely available, or their resolution in space and time may be limited (for example 16 by 16 "pixels" at 3Hz in the device presented in (Kerridge et al. 2004)). Using multiple cameras can help address occlusion issues, but requires their registration to take advantage of the setup, and this work focuses on a simpler single-camera system.

### 4.3.3 Challenges of Pedestrian Tracking in Computer Vision

In order to automatically extract pedestrian data from video data, road users must be detected, tracked from one frame to the next and classified by type; at least as pedestrians and non-pedestrians. Automated pedestrian monitoring using video data is a complex task, especially in the type of "open" and "busy" urban environment on which this research is focused. Open environment refers to the mixed traffic, including motorized vehicles and pedestrians, the variable structure, and the multiple flows of moving objects that may enter, leave the scene in various regions, and stop for varying amounts of time in the field of view (e.g., at traffic lights or stop lines, or park on the side of the road). Busy environment refers to the concentrated presence of road users, especially in what relates to high level of crowdedness. This is a much more challenging
type of environment than more controlled ones such as rural highways which have received more attention up to now.

Although great progress has been made in recent years, tracking performance is difficult to report and compare because implementations are not publicly available and common benchmarks are limited. Tracking pedestrian and mixed traffic in crowded scenes is still an open problem. Most vision-based pedestrian data collection took place in idealized conditions, e.g., heads and feet present all the time (Hoogendorn, Daamen \& Bovy 2003), low pedestrian volume (Malinovskiy, Wu \& Wang 2008) (Chae \& Rouphail 2008), or heavily controlled indoor experiments including markers on pedestrians (Hoogendorn, Daamen \& Bovy 2003) (Kerridge et al. 2004). The collected datasets are typically small and, in some cases, require significant manual input to correct the automated results and to supplement with additional data (Chae \& Rouphail 2008).

### 4.4 Methodology

The main objective of this section is to describe the developed system components and to document various algorithmic modifications required to meet the requirements of pedestrian detection and tracking. Figure 4.1 shows different components of the prototype system. The following sections describe various system components.


Figure 4.1 Layout of the pedestrian detection and tracking prototype system. The figure shows the five main layers of the system. Depicted also is the data flow among system modules from low-level video data to a database of detected, tracked, and classified road user.

### 4.4.1 Camera Parameters

The main objective of camera calibration is to find a set of parameters to establish a mapping from world coordinates to image plane coordinates. Once this mapping is created, real-world coordinates of points that appear in the video can be recovered. The extrinsic parameters specify the translation and orientation of the camera coordinates relative to the world coordinates. Intrinsic parameters describe the perspective projection of the road scene onto the image plane. Both sets of parameters can be obtained by minimizing the difference between the projection of geometric entities, e.g., points and lines, onto world or image plane spaces and the real-world measurements of these entities.

The reliance on point correspondences at the site monitored in the case study presented in this chapter (section 4.5) was hampered by a recent surface painting of the intersection that left only a handful of common features on both the orthographic satellite image and the video images. This difficulty was overcome, to some extent, by relying on distance constraints to inform the calibration process. Linear field observations were performed to obtain the true lengths of entities that appeared in the video images. Another practical difficulty arose because it was not possible to conduct a lab-based camera calibration in order to find all the intrinsic camera parameters aside from the focal length. All camera parameters had to be estimated based on information collected from the traffic scene. This increased the processing time required for the convergence criterion to be met, that is for the gradient of the objective function to be less than 1e-05. Accurate camera parameters were required in this study since the error magnitude in speed estimation that results from
position estimation is significant at low speeds. Camera calibration was conducted following the methodology presented in Chapter 3. The different camera settings used for video data collection in this chapter were referred to as case studies BR-1 to BR-4.

The obtained camera calibration results were very satisfactory (average percentage error in linear measurements was 4\%). Figure 4.2 shows sample road user tracks projected on an orthographic satellite image of the scene. Similar studies in the literature used artificial construction of an orthographic image using video image rectification, e.g., (Laureshyn \& Ardö 2006). The approach followed in this study by projecting the video data on an independent site map proved helpful in visually verifying the accuracy of projection - especially with the difficulties faced in obtaining calibration data. In addition, it was possible to collate pedestrian tracks obtained from different camera settings into a single site map, whereas video image rectification produces a setting-dependent site map (Masoud \& Papanikolopoulos 2007).
a) Tracks in image space
b) Tracks in world space


Figure 4.2 Pedestrian tracks at site BR-2. Figure a) shows road user tracks in the image space. Figure b) shows the same tracks projected on an orthographic image. The trajectories are classified by object type (vehicles or pedestrians) and direction. Trajectory clusters 1,2, and 3 are for pedestrians moving SoutheastNorthwest, Northwest- Southeast and Crossing respectively, while cluster 4 is for vehicles.

### 4.4.2 Feature Tracking and Grouping

A feature-based tracking system was initially developed for vehicle detection and tracking as part of a larger system for automated road safety analysis in (Saunier \& Sayed 2006). Tracking features is done through the well known Kanade-Lucas-Tomasi Feature Tracker. Additional "filters" are added to keep only relevant features. First, stationary features are not tracked and are discarded. This and the movement of objects imply that new features must be regularly generated to keep tracking the whole field of view. Second, feature tracker errors are dealt with by enforcing regularity motion checks, i.e., bounds on acceptable feature acceleration and change in direction.

Since a vehicle can have multiple features, the next step is to group the features, i.e., decide what set of features belongs to the same object, using cues like spatial proximity and common motion. The grouping method described in (Beymer et al. 1998) was extended to handle intersections in (Saunier \& Sayed 2006). A graph is constructed over time: the vertices are feature tracks, edges are grouping relationships between tracks and connected components (groups of features) correspond to vehicle hypotheses. Two parameters are crucial for the success of the method: the connection distance $\mathrm{D}_{\text {connection, }}$ i.e., the distance between two features for their connection, and the segmentation distance $\mathrm{D}_{\text {segmentation, }}$ i.e., the threshold on the difference between the minimum and maximum distances between two features above which these features are disconnected. Features must also be tracked simultaneously for a minimal period of time to make sure that the common motion condition is enforced. The tracking accuracy for motor vehicles has been measured between $84.7 \%$ and $94.4 \%$ on three different sets of sequences (Saunier \& Sayed 2006). This means that most trajectories are detected by the system, although over-grouping and over-segmentation still happens and may create other problems.

### 4.4.3 High-level Object Processing

Different than previous work by (Saunier \& Sayed 2006), the traffic scenes analyzed in this thesis are mixed featuring road users with very different sizes, e.g., passenger cars and pedestrians. In their work, he connection and segmentation distances could only be adjusted for one type of road user. To address this issue, the original system has been extended by obtaining the type of the road users. The parameters are initially set for pedestrians, with the
undesirable effect of producing over-segmented vehicle objects. To address this shortcoming, once the groups of features belonging to cars are identified, their constituting features are reprocessed by the grouping algorithm using larger connection and segmentation distances. Effectively, feature grouping is conducted at two levels, one for pedestrians and the second is for vehicles. This modification extended the video processing time, however results were encouraging.

In the system settings used in this Chapter, a simple test on the maximum speed reached by road users was sufficient to discriminate between pedestrians and motorized road users in most cases.

### 4.4.4 Manual Input to the Video Analysis System

The point of an automated system is to minimize user input, especially to eliminate the need for continuous supervising. The main input provided to the video analysis system was a reliable set of tracking parameters. Various adjustments made to the tracking parameters were conducted following a trial-and-error fashion along with visual inspection of tracking results. Since the world coordinates are recovered, the parameters could be reused unchanged in various scenes. This approach for parameters selection proves satisfactory in this chapter and Chapter 5. The challenges faced in Chapter 6 required the use of a more sophisticated approach for parameter selection.

### 4.5 Case Study

This section describes the analysis of video sequences collected at an open and busy intersection, in the Vancouver Downtown area. The objective of this analysis is to test the ability of the system to measure the walking speed of pedestrians in a variety of settings. Validation was conducted as follows:

1. Select an intersection on a main commercial corridor in Vancouver, British Columbia with a nearby camera setting location. The intersection should contain a variety of pedestrian facilities. Also, the location should be on the main course of crowd movement outbound of a concurrent event in order to test the system.
2. Record high-definition video data for the intersection in day- and nighttime conditions.
3. Select a random sample that represents $10 \%$ of the detected and tracked pedestrians (individuals or groups).
4. Calculate the average walking speed by measuring the time the elapses during observing the crossing between two check lines.
5. Compare the system-based and observer-based walking speeds.

### 4.5.1 Data Collection

Videos were collected for pedestrian movement at a traffic intersection on Robson Street which is a major commercial and business corridor in the Vancouver Downtown area with an active walking environment. A total of seven footages were recorded from 8:00 PM till 12:00 PM in order to capture normal night-time pedestrian movement as well as crowd movement to and from a fireworks event that took place in the same time. The timing of the video
survey was intended to be concurrent with a fireworks event in order to capture higher pedestrian volumes and to provide walking speed information for local transportation authorities in order to assist in predicting outbound crowd movement in future events.

The camera was set on the 29th floor of a high-rise building that overlooks that intersection. Figure 4.2 shows a video image and an orthographic satellite image of the intersection along with real-world tracks of pedestrian movement as obtained using the video analysis system. The recorded video sequences covered a wide variety of observation conditions that often exist in pedestrian facilities. Various pedestrian density conditions were monitored, ranging from crosswalks with low pedestrian volumes to concentrated crowd movement. Pedestrian movements were monitored at sidewalks, crosswalks, and along, Robson Street, a thoroughfare that was closed for motorized traffic.

### 4.5.2 Data Analysis

Tracks shown in Figure 4.3 depict the movements of individual pedestrians as well as groups of pedestrians. Tracked objects, i.e., individuals and groups which reached a speed higher than a specific threshold, $3.5 \mathrm{~m} / \mathrm{s}$, were classified as motorized traffic and filtered out. Figure 4.4 shows a compilation of sample of pedestrian and vehicle speed profiles in time. Pedestrians exhibit a characteristic rhythmic movement. Attempts to use this idiosyncrasy as a classification feature were not successful, mainly due to incomplete pedestrian tracks. Therefore, only a maximum speed threshold was used for road user classification. Pedestrian tracks were clustered using the k-means algorithm. Each track was represented by a four-dimensional vector, each element being
the average movement orientation over a section of the track. The first and last sections cover $20 \%$ of the entire duration during which the pedestrian object existed, starting from both ends. The two intermediate sections were selected at one third of each pedestrian track with a length of $10 \%$ of the track duration. Selection of several clustering variables was necessary to capture turning pedestrian movements. The number of clusters was selected based on visual observation of the prevalent streams of pedestrian movement in each video record. The four trajectory clusters that appear in Figure 4.3 are: SoutheastNorthwest movement (cyan), Northwest-Southeast movement (blue), crossing movement (green and black), and vehicles (red).


Figure 4.3 Road user trajectories (tracks) transformed to world coordinates. Tracks of motorized road users are depicted in red. Remaining tracks are colorcoded based on a k-means clustering of pedestrian tracks.

Night-time footage was the most challenging to analyze due to the poor visibility of pedestrians in dim corners of the intersection. A specific set of feature tracking parameters were used to detect and track more features. As shown in Figure 4.5, the results obtained were generally satisfactory. Data however could not be recovered from low-light areas. Dark-clothed pedestrians were difficult to detect without rendering the integration of large volume of uninformative and low-quality features. Table 4.2 shows a summary of the dayand nigh-time tracking parameters.

Table 4.2 Summary of tracking parameters

| Tracking Parameter | Day-time Condition sequences BR4,5,6 | Day-time Condition sequences BR1,2,3 | Night-time Condition |
| :---: | :---: | :---: | :---: |
| feature-quality ${ }^{\text {a }}$ | 0.06 | 0.06 | 0.01 |
| min-feature-distance-klt ${ }^{\text {a }}$ | 3 | 4 | 4 |
| Minimum displacement to track feature (m) ${ }^{\text {b }}$ | 0.01 | 0.015 | 0.015 |
| Maximum acceleration to keep features $\left(\mathbf{m} / \mathbf{s}^{2}\right)^{b}$ | 2.5 | 3 | 3 |
| Connection distance (m) ${ }^{\text {b }}$ | 1 | 0.5 | 0.5 |
| Connection cosine ${ }^{\text {b }}$ | 0.5 | 0.8 | 0.8 |
| Segmentation distance (m) ${ }^{\text {b }}$ | 0.4 | 0.2 | 0.2 |
| Minimum number of features per group ${ }^{\text {b }}$ | 4 | 5 | 5 |

a Refer to OpenCV documentation for a complete description of these parameters at the following links:
http://lopencr.willowgarage.com/documentation///feature detection.html\#goodfeaturestotrack
http://opencv.willowgarage.com/documentation/c/motion_analysis_and _object_tracking.html\#calcopticalflowpyrlk
b Refer to the feature grouping algorithm for more details (Beymer et al. 1998).


Figure 4.4 The horizontal axis shows the frame number (surrogate for time) and the vertical axis shows the observed speed in $\mathrm{m} / \mathrm{s}$. Pedestrian profiles (in green) exhibit a characteristic rhythm.

Walking speed data was collected at user-defined registration areas for each tracked object that falls in a specific movement cluster. The definition of a registration area was necessary for gathering walking speed data in desirable specific spatial context. Since walking speed varied during the time a tracked object was present within the registration area, the average walking speed within this duration was recorded. Figure 4.6 shows the registration area defined for the indicated crosswalk. Registration areas were defined for other pedestrian facilities (two sidewalks, two unmarked crosswalks, and another marked crosswalk) in order to gather walking speed data.

### 4.5.3 Validation

The validation process in this study was concerned with walking speed measurements. The average walking speeds for a $10 \%$ random sample drawn from tracked pedestrian objects was compared to manual video observations of the walking speed. Walking speeds were manually calculated based on the time required by moving objects to traverse the shortest distance between two check lines. The check lines were selected to be the road markings of the crosswalk across Robson Street as shown in Figure 4.7. Figures 4.8 (a) and (b) show a comparison between measured and automatically calculated walking speeds. There is a very good agreement between manual and automated walking speed values $($ RMSE $=0.0725 \mathrm{~m} / \mathrm{s}$ and $0.0548 \mathrm{~m} / \mathrm{s}$, respectively). The residual errors can be attributed to the inaccuracy of manual speed calculation in which the pedestrians are unrealistically assumed to follow the shortest path between two check lines, inaccuracy in camera calibration, and irregularities in pedestrian tracks due to noise in feature detection.


Figure 4.5 A sample frame from night-time video analysis. Displayed are red bounding boxes around pedestrian objects and their walking speed.


Figure 4.6 The figure shows pedestrian trajectories that crossed through the marked data collection area. Trajectories are collated and projected to the world image from different videos with different fields of view and hence may be truncated in different regions.


Figure 4.7 Two sets of check-lines for collecting manual observations of average walking speeds. The spacing between the upper set of check-lines (crosswalk) is 3.61 m and the spacing between the bottom set of check-lines (long segment) is 11.58 m .

### 4.5.4 Discussion

The case study was intended to monitor pedestrian movement under several conditions. The monitored pedestrian facilities were a crosswalk, two sidewalks, and two unmarked crosswalks. A summary of walking speed statistics is presented in Table 4.3. Figures $4.9(\mathrm{a})$ and $4.9(\mathrm{~b})$ show sample distributions of pedestrian walking speed for crossing and sidewalk movements respectively. Pedestrians moving from Northwest to Southeast had to walk up a $5 \%$ longitudinal grade. The average walking speed for all pedestrian objects was $1.22 \mathrm{~m} / \mathrm{s}$ and the average and 15th percentile crossing speed was 1.31 and $0.93 \mathrm{~m} / \mathrm{s}$ respectively. This value is consistent with studies in the literature as shown in Table 4.1.

Table 4.3 Summary of walking speed statistics

| Movement | No. Pedestrian objects | Average$(\mathrm{m} / \mathbf{s})$ | Stan. Dev. ( $\mathrm{m} / \mathrm{s}$ ) | $P$-value (difference in means between column and row movement types) |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | Southeast- <br> Northwest <br> UCW | Southeast - <br> Northwest SW | Northwest Southeast $U C W$ \& $S W$ |
| SoutheastNorthwest UCW ${ }^{1}$ | 907 | 1.41 | 0.26 | - | - | <0.0001 |
| SoutheastNorthwest SW ${ }^{2}$ | 1148 | 1.04 | 0.28 | - | - |  |
| Northwest - Southeast UCW | 289 | 1.26 | 0.30 | <0.0001 | - | - |
| Northwest - Southeast SW | 44 | 0.97 | 0.24 | - | 0.0333 |  |
| MCW ${ }^{3}$ | 162 | 1.31 | 0.37 | 0.0002 | - | 0.0069 |
| Night-time | 656 | 1.13 | 0.21 | - | - | <0.0001 |
| 1 UCW: unmarked crosswalk 2 SW: sidewalk 3 MCW: marked crosswalk |  |  |  |  |  |  |



Figure 4.8 Figure a) shows validation of walking speed measurements at daytime. Horizontal axis depicts walking speed based on the time interval required to walk between two check lines. Vertical axis depicts automatically measured average walking speed based. Figure b) shows validation of walking speed measurements at night-time conditions.

b) Southeast-Northwest crossing



Figure 4.9 Figure a) Walking speed frequency distribution for pedestrians moving through the data collection area shown in Figure 4.4 across Robson St. Figure b) walking speed frequency distribution for pedestrians moving from Southeast to Northwest through corresponding data collection areas on both sidewalks of Robson Street.

There is a statistically significant ( $\mathrm{p}<0.05$ ) difference between walking speed at crosswalks and at sidewalks, walking uphill (from Northwest to Southeast) and opposite direction. There is no statistically significant $(p=0.0616)$ difference between walking speed along marked and unmarked crosswalks. However this result is deemed as inconclusive since it was measurably close to statistical significance. There is a statistically significant difference between NorthwestSoutheast walking speed at night during a road closure and at day time along the sidewalks. This was likely due to the larger space afforded for pedestrians during a road closure as well as the leisurely nature of walking back from a night event.

As discussed before, one of the major advantages of video-based data collection is to capture walking speed variability, quantified by the standard deviation of
speed measurements over the time interval within a registration area. It was observed that pedestrians walked faster along unmarked crosswalks in case of approaching vehicles. The variability in crossing speed was recorded for movements along marked and unmarked crosswalks. There is a statistically significant ( $\mathrm{p}<0.0001$ ) higher variability of walking speed at unmarked crosswalks compared to marked crosswalks.

### 4.6 Conclusions

Pedestrian walking speed has been the subject of continuous research. There is a recent revival in pedestrian studies that is motivated in part by demographic changes. It is believed that future data collection is necessary to develop a better understanding of pedestrian movement and the factors that influence walking speed.

The majority of commercial techniques developed for automatically collecting traffic data focus on vehicular traffic. The technological aspects of automated pedestrian data collection are generally more involved than vehicular traffic. The majority of walking speed studies in the literature does not make use of automated video analysis for collecting pedestrian data. In this study, an automated system for collecting pedestrian walking speed using video analysis was developed and tested. A system previously developed for vehicle detection and tracking was significantly modified to adapt for particularities of pedestrian movement and to discriminate pedestrian and motorized traffic. The system was tested on real video data collected at Downtown area of Vancouver, British Columbia, during day- and night-time conditions. It was found that
pedestrians walk faster at marked crosswalks than sidewalks. Walking speed was more variable at unmarked crosswalks compared to marked crosswalks. Gradient and lighting conditions were identified as statistically significant variables that influence walking speed.

Several conclusions can be drawn from this research work. First, the accuracy of walking speed calculations was sensitive to the camera calibration parameters. Several challenges were faced during the recovery of the camera parameters due to site-specific conditions. The robust camera calibration methodology presented in Chapter 3 was successfully used. Second, night-time conditions proved to be the most difficult as expected because of the obscurity of pedestrian outlines and video recording noise. A special set of detection parameters was used for night videos and results obtained were satisfactory. Finally, the literature of pedestrian observational studies is yet to benefit from automated video analysis techniques. It is expected that the system presented in this study will be further improved by adding other appearance-based techniques.

## Automated Detection of Pedestrian-Vehicle Conflicts

### 5.1 Background

There is a growing emphasis on the sustainability of transportation systems. This emphasis is often manifested by promoting public transit and improving the traffic conditions for non-motorized modes of travel. Walking is a key non-motorized mode of transport that connects different components of a multimodal transport network and interfaces with external activity areas (land use). Building safe and walking-friendly pedestrian facilities is fundamental to encouraging and accommodating walking activities. For example most modern municipalities are required to have in place official
community plans to manage growth and many, if not most, of them contain policies that promote pedestrian activities. There also an increasing public fund allocation for safety programs that focus on problem areas such as pedestrian injuries.

The study of pedestrian safety focuses mainly on the interaction with other motorized and non-motorized traffic and the conformity to traffic control regulations. The main focus of this chapter is on analyzing traffic events that involve conflicting movements between pedestrians and vehicles. As discussed in Chapter 1, road safety analysis has traditionally relied on historical collision data. However, there are some shortcomings to this approach: rarity and randomness of road collisions, extended observational periods, and concerns with the quantity and quality of collision data. Collision data reporting is often incomplete and biased toward highly damaging collisions. Collision auditing is conducted after collision occurrence, at which time the causes, specific location, and behavioural aspects of the event are subject to judgement - if ever reported.

The shortcomings of relying on collision data for pedestrian safety analysis are even more acute. For example, collisions involving pedestrians are less frequent than other collision types. Pedestrian-involved collisions accounted from 1992 to 2001 for $3.6 \%$ of the total number of collisions in British Columbia (British Columbia Traffic Collision Statistics 2005). In addition, pedestrian traffic volumes are less readily available than motorized traffic volumes due to the difficulties of collecting pedestrian data. The identification of pedestrian exposure to the risk of collision is therefore
challenging without tracking pedestrian and vehicles. Pedestrians, being vulnerable road users, when involved in collisions, have considerably higher chances of being severely injured, with little chance of the collision being classified as property-damage-only. From 1992 to 2001, pedestrians accounted for $14.8 \%$ of traffic collision victims (i.e., injured or killed) in British Columbia and $15.2 \%$ in Canada.

The use of surrogate safety measures has been advocated as a complementary approach to address these issues and to offer more in-depth analysis than relying on accidents statistics alone. One of the most developed methods rely on traffic conflict analysis. Traffic Conflict Techniques (TCTs) involve observing and evaluating the frequency and severity of traffic conflicts at an intersection by a team of trained observers. The concept was first proposed by Perkins and Harris in 1967 (Perkins \& Harris 1968). A traffic conflict takes place when "two or more road users approach each other in space and time to such an extent that a collision is imminent if their movements remain unchanged" (Amundsen \& Hydén 1977). A common theoretical framework ranks all traffic interactions by their severity in a hierarchy, with collisions at the top and undisturbed passages at the bottom (Svensson \& Hydén 2006).

TCTs hold several advantages over collision-based safety measures. Traffic conflicts are more frequent than traffic collisions. TCTs were shown in some studies to produce estimates of average accident frequency that are comparable to accident-based analysis (Migletz, Glauz \& Bauer 1985). Traffic conflicts are manually collected by a team of trained observers, either on site or offline through recorded videos. Despite the considerable effort that is put
into the development of training methods and the validation of the observers' judgement, such data collection is subject to intra- and inter-observer variability. This can compromise the reliability and repeatability of traffic conflict data collection. In addition, the training and employment of human observers makes traffic conflict studies costly. In a recent study (Fitzpatrick et al. 2006), the effort for extracting pedestrian and motorist data from videos was deemed "immense". This type of data is not only difficult to collect, but also its usefulness is sensitive to the accuracy of the collection process.

Due to limitations of manual data collection, a growing trend of using automated data collection systems has caught on in the field of transportation engineering. In particular, automated video analysis has attracted considerable interest. Video sensors are now widely available (traffic cameras are already installed on many roadways) and are relatively inexpensive.

Previous work on the automated analysis of video data in transportation has mainly involved the detection and tracking of vehicular traffic, e.g., (Saunier \& Sayed 2007). Particularities of pedestrians make their detection and tracking in video sequences challenging. Problems arise from their intertwined tracks, groupings, varied appearance, non-rigid nature, and the generally less organized nature of pedestrian traffic as compared to vehicular traffic; which are subject to standard "rules of the road" and lane discipline.

This work strives to address some of the previous shortcomings and research recommendations. This chapter discusses the development and testing of an automated video-analysis system that meets the following objectives:

1. Detect and track road users in a traffic scene, and classify them into pedestrian and motorized traffic.
2. Identify important events in a video sequence. The definition of an important event in this study is "any event that involves a crossing pedestrian and a conflicting vehicle in which there exists a conceivable chain of events that could lead to a collision between these road users". To be conceivable, a reasonable chain of events leading to a collision should be considered. The actual quantitative interpretation of this general definition is given in the experimental study. The pre-condition for an important event to occur in this study is that a left-turning vehicle enters the monitored crosswalk in the presence of a pedestrian or a group of pedestrians already in the crosswalk. Excluded were the events that involved the following unlikely chain of events: a vehicle reverting its travel direction, a pedestrian changing movement from walking to running ( $>3.5 \mathrm{~m} / \mathrm{s}$ ), and a collision involving pedestrians standing beyond the curb line.
3. Report objective measures of severity indicators for all events.

The system can either work autonomously, or be used to assist human experts by sifting through large amounts of video data and identifying important events that is worthy of further investigation. The system was tested on video data recorded for two days at a location in the Downtown area of Vancouver, British Columbia. The task of calculating traffic conflict indicators for each event that involved a pedestrian-vehicle interaction was performed in a fully automated way.

### 5.2 Previous Work

### 5.2.1 Pedestrian-vehicle Conflicts

Cynecki (1980) described a conflict analysis technique for pedestrian crossings, citing fundamental differences between vehicle-vehicle and pedestrian-vehicle conflicts, and indicating desirable characteristics for conducting a conflict study. Two of these characteristics, repeatability and practicability of traffic conflict studies, can greatly benefit from automated video analysis, which offers a cost-efficient and objective means for traffic conflict analysis. In subsequent bodies of work, several studies adopted traffic conflict analysis to study the level of safety of pedestrian crossings, e.g., (Lord 1996) (Van Houten et al. 1997) (Tiwari, Mohan \& Fazio 1998) (Tourinho \& Pietrantonio 2003) (Malkhamah, Miles \& Montgomery 2005) (RodriguezSeda, Benekohal \& Morocoima-Black 2008). While the majority of past work was based on observer-based traffic conflict analysis, few studies, e.g., (Malkhamah, Miles \& Montgomery 2005), developed a relationship between conflict indicators and automatically measured parameters, such as motorist deceleration rate. In a recent study (Chae \& Rouphail 2008), an automated analysis of video data was performed to investigate the interactions between pedestrians and vehicles at roundabout approaches.

### 5.2.2 Severity Conflict Indicators

Various conflict indicators have been developed to measure the severity of an interaction by quantifying the spatial and temporal proximity of two or more road users. The main advantage of conflict indicators is their ability to capture the severity of an interaction in an objective and quantitative way.

Concerns however remain regarding the lack of a consistent and accurate definition of conflict indicators (Chin \& Quek 1997). Conflict indicators developed in the literature are capable of capturing and connoting different proximal, situational, and behavioural aspects of traffic conflicts. Each indicator however possesses drawbacks that limit its ability to measure the severity of recognized traffic events. For a review of conflict indicators and their relative advantages and limitations, the readers are referred to (Archer 2004).

### 5.2.3 Pedestrian Detection and Tracking

To study pedestrian-vehicle conflicts, all road users must be detected, tracked from one video frame to the next, and classified by type; at least as pedestrians and motorized road users. This is a challenging task as described in Chapter 4. In addition to specific problems when tracking pedestrians, common problems are global illumination variations, multiple object tracking, and shadow handling. Various approaches for pedestrian tracking have been described in Chapters 2 and 4. Although great progress has been made in recent years, the tracking performance of different pedestrian tracking systems is difficult to report and compare, especially when many of these systems are not publicly available or their details disclosed, and when benchmarks are rare and not systematically used. Tracking pedestrians and mixed traffic in crowded scenes is still an open problem. To the author's knowledge, no attempt has yet been made to develop a fully functional videobased pedestrian conflict analysis system. However, only for the purpose of
detecting and analyzing pedestrian-vehicle conflicts, feature-based tracking provided reliable performance as demonstrated in Chapter 4.

### 5.3 Methodology

This section describes the development of a video-based system for the automated analysis of pedestrian conflicts. The system has 5 basic components: 1) video pre-processing; 2) feature processing; 3) grouping; 4) high-level object processing; and 5) information extraction. A depiction of the video analysis system was provided in Figure 4.1 in Chapter 4. The following sections describe the work performed in context of the work presented in this chapter.

### 5.3.1 Camera Calibration

The main purpose of camera calibration is to estimate a set of camera parameters to project objects onto the video sensor (image plane). The inverse transformation that recovers world coordinates of objects in the video images can also be obtained from the camera parameters. Camera parameters are classified into extrinsic and intrinsic parameters. Extrinsic camera parameters specify the displacement of the camera's coordinates relative to world coordinates. Intrinsic parameters are required to establish a perspective projection of objects defined in the camera coordinates onto the image plane. Both sets of parameters can be obtained by minimizing the difference between the projection of geometric entities, e.g., points and lines, onto world or image plane spaces, and the actual measurements of these entities in projection space as described in Chapter 3.

The calibration data used in this study, scene PG described in Table 3.1, was composed of a set of 22 points selected from salient features in the monitored traffic scene that appear in the video image, as shown in Figure 5.1. The world coordinates of the calibration points were collected from an orthographic image of the location obtained from Google Maps. The only intrinsic parameter considered in this study was the camera focal length. In this chapter, image plane coordinates were back-projected onto the road surface, i.e., assuming the positions of all road users are projected on the plane $\mathrm{Z}=0$.

The calibration accuracy obtained by applying the methodology presented in Chapter 3 was satisfactory. The average percentage error in coordinate estimates was less than $1 \%$. The camera calibration problem faced in this study was relatively simple due to the abundance of lane marking features which appear in the orthographic image of the traffic scene.

Figure 5.2 shows the projection of a sample of pedestrian tracks on an orthographic satellite image of the traffic scene. Similar studies in the literature used artificial construction of an orthographic image using video image rectification e.g., (Laureshyn \& Ardö 2006). The approach followed in this study by projecting the video data on an independent site map proved helpful in visually verifying the accuracy of the resulting projection; especially given the difficulties faced in obtaining calibration data. In addition, it was possible to collate pedestrian tracks obtained from different camera settings into a single site map, whereas video image rectification produces a setting-dependent site map.

### 5.3.2 Video Formatting

Depending on the video source, it may be necessary to encode the video in a suitable format for later processing, as well as correct recording artefacts such as interlacing. For this study, a digital video recorder was used that encoded video to a suitable format (AVI container and MPEG encoder).
a) Image space

b) World space


Figure 5.1 The 22 points used to estimate the camera calibration are displayed on a video frame in Figure a) and on an orthographic satellite image of the traffic scene in Figure b). Bulleted points ( $\bullet$ ) are manually annotated and xshaped points ( $x$ ) are projections of annotated points using the estimated camera parameters.


Figure 5.2 A sample of pedestrian tracks is projected on an orthographic satellite image of the traffic scene. Vehicle tracks are depicted in red and pedestrian tracks are in black.

### 5.3.3 Object Tracking

Detection and tracking and grouping of road user features were conducted using the same methodology previously presented in section 4.3.2. Difficulties occur in scenes where there is mixed traffic use and road users vary in sizes, e.g., vehicles and pedestrians, and the connection and segmentation distances can only be adjusted for one type of road user. To address this issue, the original system has been extended by identifying the types of the road users. The parameters were adjusted for pedestrians, and consequently the motorized vehicles are over-segmented. Once the groups of features belonging to non-pedestrian objects had been identified, the feature
were processed a second time by the grouping algorithm using larger connection and segmentation distances.

In the current system, a simple condition on the maximum speed of each road user was sufficient to discriminate between pedestrians and motorized road users in most cases. This test will typically classify bicyclists as motorized road users, which may lead to detect what are truly pedestrian-bicyclist conflicts as pedestrian-vehicle conflicts (or collisions due to erroneous size estimation).

### 5.4 Case Study

The system was tested on traffic video recorded for two days during daytime at a crosswalk in Downtown Vancouver (case study PG in Chapter 3). The objective of the case study is to assess the capability to identify instances of important events, and to calculate severity conflict indicators for each event.

### 5.4.1 Site Description and Data Collection

The study area is the intersection between W Pender St. and W. Georgia St. in the Downtown area of Vancouver, British Columbia, Canada. The main interacting movements are pedestrian crossing and left-turn vehicles. Leftturn traffic at signalized intersections poses a particularly increased risk of collision for pedestrians (Lord 1996). Furthermore, this intersection is unique in that it is a skewed intersection within a corridor grid of streets all containing right-angle intersections. Hence, there is a high possibility of observing an adequate number of pedestrian-vehicle conflicts. In this study,
important events occurred when a pedestrian and a vehicle co-existed inside the monitored crosswalk. The potential for observing pedestrian-vehicle conflicts was confirmed by testimonials of persons familiar with traffic movement at this site.

A video camera was set on the $6^{\text {th }}$ floor of a building that overlooks the intersection and aimed towards the west. Video recording was conducted for a total of 20 hours over two business days. Approximately, a total of 7000 leftturning vehicles and 2100 pedestrians were observed. These volume estimates are derived from the automated video analysis.

### 5.4.2 Calculation of Conflict Indicators

The system detects all events constituted by the pairs of pedestrians and vehicles that are in the traffic scene simultaneously. Among these events, this study is interested in important events, as defined in the introduction, and traffic conflicts, which are a subset of important events. The complement of important events over the space of all traffic events are defined as undisturbed or uninterrupted passages.

In order to compensate for the limitations of individual conflict indicators, four conflict indicators were calculated in this study. The first is Time-toCollision (TTC) defined as "...the time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained." (Hayward 1968). An accurate estimation of TTC however requires considerable field measurement of road user positions, speed and direction of movement. This work relies on the traditional operational definition of a collision course, extrapolating the road users' movements with constant
velocity (used in (Svensson \& Hydén 2006) for example). This hypothesis is however simple and may lead to unrealistic collision-course estimates.

Other conflict indicators are used to capture different proximity aspects. PostEncroachment Time (PET) (Cooper 1984) is the time difference between the moment an offending road user leaves an area of potential collision and the moment of arrival of a conflicted road user possessing the right of way. Gap time (GT) is a variation on PET that is calculated at each instant by projecting the movement of the interacting road users in space and time (Archer 2004). Deceleration to Safety Time (DST) is defined as the necessary deceleration to reach a non-negative PET value if the movements of the conflicting road users remain unchanged (Hupfer 1997). Allen et al. (Allen, Shin \& Cooper 1978) ranked GT, PET and Deceleration Rate as the primary measures for left-turn conflicts. DST was selected since it captures greater details of the traffic event. TTC was selected since it is the primary traffic conflict indicator in the literature. The values of conflict indicators used in event detection are the minimum TTC, the minimum GT, the maximum DST and PET. Figure 5.3 shows sequences of severity conflict indicators calculated for a traffic conflict event. Algorithm 5.1 presents a description of the method used in this study to calculate the severity indicators.

## Algorithm 5.1: Algorithm for calculating conflict indicators for a pedestrian-vehicle event

Definitions: 1) A generic position function $F: \mathbb{N}^{+} \rightarrow \mathbb{R}^{2}$ returns the world-space position of a road user $(x, y)$ at time instant $t$ such that $F(t)=(x, y)$.
2) A generic velocity function $\dot{F}: \mathbb{N}^{+} \rightarrow \mathbb{R}^{2}$ returns the velocity components of a road user $(\dot{x}, \dot{y})$ at time instant $t$ such that $\dot{F}(t)=(\dot{x}, \dot{y})$.
3) A generic position extrapolation function $E: \mathbb{N}^{+} \times \mathbb{R}^{2} \times \mathbb{R}^{2} \rightarrow \mathbb{R}^{2}$ returns the position at time $t+\Delta t$ of a road user with current position $(x, y)$ and velocity $(\dot{x}, \dot{y})$ at time $t, E(\Delta t, x, y)=(x, y)+(\dot{x}, \dot{y}) * \Delta t$.
Input: Let $P: \mathbb{N}^{+} \rightarrow \mathbb{R}^{2}$ be the pedestrian position function, defined for $t \in\left[t_{p 1}, t_{p 2}\right]$. Let $H_{f 1}, H_{f 2}, H_{r 1}$, and $H_{r 2}$ be the position functions of the vehicle front and rear corners respectively, that are all defined for $t \in\left[t_{v 1}, t_{v 2}\right]$.
Let $\dot{P}$ and $\dot{H}$ be the pedestrian and vehicle velocity functions, respectively
1- $\quad$ Let $W$ be the segment demarcating the crosswalk
Let $c_{1}=0.25 \mathrm{~m} / \mathrm{s}$ be a speed threshold and $\Delta t_{e}=10 \mathrm{~s}$ be maximum extrapolation time.

Output: Time series of TTC, DST, and GT, and the PET

## begin

for each pair consisting of a pedestrian and a vehicle whose observed trajectories intersect at point set $P_{o}$

Let $T_{o}$ be the times at which each road user occupies the points in $P_{o}$
Find the times $T_{1}$ at which the observed vehicle rear corner positions $H_{r 1}$ and $H_{r 2}$ intersect $W$
$\mathrm{PET}=\max \left(\max \left\{T_{1}\right\}-\min \left\{T_{o}\right\}, \max \left\{T_{o}\right\}-\min \left\{T_{1}\right\}\right)$
for each $t \in\left[\max \left\{t_{p 1}, t_{v 1}\right\}, \min \left\{t_{p 2}, t_{v 2}\right\}\right]$ such that $\|\dot{P}(\mathrm{t})\| \geq \mathrm{c}_{1}$ and $\|\dot{H}(\mathrm{t})\| \geq \mathrm{c}_{1}$

Find the intersection points $P_{2}$ between the extrapolated positions of the pedestrian $E(\Delta t, x, y) \mid \Delta t_{e} \geq \Delta t \geq 0$ and of the vehicle front corners $E\left(\Delta t_{e}, H_{f}(t)\right) \mid \Delta t_{e} \geq \Delta t \geq 0$ for $f \in\left\{f_{1}, f_{2}\right\}$

Find the intersection points $P_{3}$ between the extrapolated positions of the vehicle rear corners $E\left(\Delta t_{e}, H_{r}(t)\right) \mid \Delta t_{e} \geq$ $\Delta t \geq 0$ for $r \in\left\{r_{1}, r_{2}\right\}$ and $W$

2-
Find the times $T_{2}$ and $T_{3}$ at which each road user occupies the intersection points in $P_{2}$ and $P_{3}$

Calculate $\Delta t=\Pi \mathbf{C}(t)$ such that $E(\Delta t, P(t))$ lies inside the extrapolated positions of the vehicle outline

Calculate $\mathbf{G T}(t)=\min \left\{T_{2}\right\}-\max \left\{T_{3}\right\}$
and $\boldsymbol{D S T}(t)=\frac{\left(\max \left\{T_{2}\right\}-\mathrm{t}\right) *\|\dot{H}(t)\|-\max _{\mathrm{p} \in \mathrm{P}_{2}}\|P(t)-p\|}{1 / 2 *\left(\max \left\{T_{2}\right\}-\mathrm{t}\right)^{2}}$
if the pedestrian leaves the conflict area before the vehicle then

Recalculate $\mathbf{G T}(t)$ and PET such that it is the time between the instant a pedestrian clears the conflict area and the instant of arrival of the front of the conflicting vehicle arrival.

Notes:
1- This definition of a "conflict area" is adopted from (Lord 1996).
2- Several algorithmic details were implemented to deal with tracking errors, e.g., tracked objects that are detected or lost during the traffic event. Details are omitted for brevity.

### 5.5 Validation

Various manually designed detection conditions defined over the composite values of the severity conflict indicators are used to identify important events. These results are compared on a sample of events manually classified by a human observer, using the definition of important events given in this chapter and the conflict definition in the US FHWA observer's guide (Parker \& Zegeer 1989). The pre-condition for an important event to occur in this
study is that a left-turning vehicle enters the monitored crosswalk in the presence of a pedestrian or a group of pedestrians already in the crosswalk. Excluded were events that involved the following unlikely chain of events: a vehicle reverting its travel direction, a pedestrian changing movement from walking to running ( $>3.5 \mathrm{~m} / \mathrm{s}$ ), and a collision involving pedestrians standing beyond the curb line. Sources of misclassification that may lead to inaccurate indicator are:

1. Errors in pedestrian and vehicle detections. These errors include: noise in tracked object position that could lead to unrealistic extrapolation of a road user's position, over-segmentation, lost detections of a road user, appearing or disappearing during a traffic event.
2. Incapability of the used conflict indicators to measure the level of severity of a traffic event.

While in some cases, it was evident why the erroneous classification of the traffic event took place, it was difficult in other cases to explicate the error source. In this study, the overall performance of the system was considered with respect to detecting and tracking road users, as well as making judicious use of the severity information measured by the conflict indicators. In this study, the detection conditions used for identifying conflicts and important events were defined by scaling serious conflict threshold values which delimit serious conflicts from other traffic events by a severity factor.


Figure 5.3 Conflict indicators for a sample traffic event. The left figure describes the traffic event shown in
figure 5.4a. The right figure describes the traffic event shown in figure 5.4b.

Table 5.1 shows the details of the detection conditions and the summary of detection results for various severity factor values. The thresholds of the identification conditions (shown in column 1) were determined by scaling the serious conflict threshold on each severity indicators by a severity factor ax, where the subscript $x$ refers to the concerned conflict severity indicator. The following typical severity thresholds are taken from the literature: $1.5 \mathrm{~s}, 3$ $\mathrm{m} / \mathrm{s}^{2}, 1 \mathrm{~s}$, and 1 s , for TTC, DST, PET, and GT respectively. For TTC (and similarly for PET and GT), all events that involved TTC $<1.5^{*}$ аттс were detected as important events. For DST, all events that involved DST $<1.5$ / adst were detected as important events. Thus defined, higher severity factors would cover events with lower conflict severity. Increasing the factors lead to a higher chance of detecting conflicts at the expense of misclassifying undisturbed passages as important events. If a severity factor is not mentioned for an indicator, it means that it is not used in the condition.

The total number of conflict events in the analyzed video sequence is 17 . The number of traffic conflicts considers the actual number of pedestrians involved, e.g., a conflict involving a vehicle and two pedestrians is counted as two conflicts. Only PET may allow detecting important events as well as conflicts separately from the other indicators. This is consistent with a study in the literature that used PET for conflict detection (Malkhamah, Miles \& Montgomery 2005). Other conflict indicators however could not solely detect an adequate percentage of important events and traffic conflicts. A combination of the four conflict indicators enabled the system to automatically capture $89.5 \%$ of true conflicts and $71.7 \%$ of important events while however detecting $54.5 \%$ of undisturbed passages as important events.

Table 5.1 Summary of validation results

| Identification Conditions ${ }^{1}$ | Percentage of each event types correctly identified by the system |  |  | Percentage of undisturbed passage falsely identified by the system as important events |
| :---: | :---: | :---: | :---: | :---: |
|  | Traffic <br> Conflict ${ }^{\text {I }}$ | Important Events ${ }^{2}$ | Uninterrupted Passages |  |
| $\mathrm{a}_{\text {TIC }}=1$ | 5.3 | 4.3 | 93.2 | 6.8 |
| $\mathrm{a}_{\text {TTC }}=2$ | 31.6 | 23.9 | 87.2 | 12.8 |
| $\mathrm{a}_{\text {TIC }}=5$ | 36.8 | 39.1 | 66.0 | 34.0 |
| $\mathrm{a}_{\text {DST }}=1$ | 0.0 | 0.0 | 100.0 | 0.0 |
| $\mathrm{a}_{\text {DST }}=2$ | 5.3 | 3.3 | 96.6 | 3.4 |
| $\mathrm{a}_{\text {DST }}=5$ | 47.4 | 51.1 | 63.0 | 37.0 |
| $\mathrm{a}_{\mathrm{GT}}=1$ | 21.1 | 27.2 | 80.4 | 19.6 |
| $\mathrm{a}_{\text {GT }}=2$ | 26.3 | 32.6 | 75.7 | 24.3 |
| $\mathrm{a}_{\text {GT }}=5$ | 42.1 | 41.3 | 66.0 | 34.0 |
| $\mathrm{a}_{\text {PET }}=1$ | 5.3 | 0.0 | 99.6 | 0.4 |
| $\mathrm{a}_{\text {PET }}=2$ | 10.5 | 2.2 | 98.3 | 1.7 |
| $\mathrm{a}_{\text {PET }}=5$ | 89.5 | 42.4 | 88.5 | 11.5 |
| $\begin{gathered} \mathrm{a}_{\mathrm{PET}}=1 \mathrm{OR} \mathrm{a}_{\mathrm{GT}}=1 \mathrm{OR} \\ \mathrm{a}_{\mathrm{DST}}=1 \mathrm{OR} \mathrm{a}_{\mathrm{TIC}}=1 \end{gathered}$ | 21.1 | 28.3 | 74.5 | 25.5 |
| $\begin{gathered} \mathrm{a}_{\mathrm{PET}}=2 \mathrm{OR} \mathrm{a}_{\mathrm{GT}}=2 \mathrm{OR} \\ \mathrm{a}_{\mathrm{DST}}=2 \mathrm{OR} \mathrm{a}_{\mathrm{TIC}}=2 \end{gathered}$ | 36.8 | 43.5 | 67.2 | 32.8 |
| $\begin{gathered} \mathrm{a}_{\mathrm{PET}}=5 \text { OR } \mathrm{a}_{\mathrm{GT}}=5 \text { OR } \\ \mathrm{a}_{\mathrm{DST}}=5 \mathrm{OR}_{\mathrm{TIC}}=5 \end{gathered}$ | 89.5 | 71.7 | 45.5 | 54.5 |
| 1 Observer-based conflict identification was performed according to the US FHWA Observer Manual (Parker \& Zegeer 1989). <br> 2 The definition of an important interaction is an event that involves a crossing pedestrian and a conflicting vehicle in which there exists a conceivable chain of events that could lead to a collision between these road users. |  |  |  |  |



Figure 5.4 Sample of automatically detected important events along with road users' trajectories. The numbers under each image are respectively the min TTC (seconds), PET (seconds), maximum DST ( $\mathrm{m} / \mathrm{s}^{2}$ ), and min GT ( s ). In the images, the road user speed is indicated in $\mathrm{m} / \mathrm{s}$.

### 5.6 Discussion

The main functional purposes of the developed video analysis system described in this chapter is to automatically identify important events that expose pedestrians to a reasonable risk of collision, including conflicts, and relay their record to a human observer for further examination. Combining information from four conflict indicators proved successful in reporting the majority of conflicts identified by a human observer. Figure 5.4 shows sample frames of important events automatically detected by the system. The capability of each conflict indicator to characterize important events was compared to manually annotated events in the dataset. As shown in Table 5.1, none of the conflict indicators was solely capable of capturing important events. The following limitations of the selected conflict indicators were identified in this study:
a. A prerequisite for TTC is the existence of road users on a collisioncourse, that is vehicles will collide if their "movements remain unchanged". The existence of a collision course is not however a necessary condition for capturing "dangerous proximity." Some dangerous interactions could not be captured by TTC because the involved road users were not on a collision course. A typical case occurs when a motorist passes behind a pedestrian at a perilously close distance. A perturbation however of the speed or direction of movement of the motorist, or slight delay on the part of the pedestrian, could potentially create a collision course and hence a calculable TTC. This issue is discussed in detail in Chapter 7.
b. The extrapolation of road users' movements with constant speed and direction could lead to erroneously small values of TTC and DST. It is observable that while TTC can function as a severity measure, it overestimates the actual conflict severity in many events. A typical situation occurs when a pedestrian is considered on a collision course with turning vehicles of which the velocity vector happens to point at the pedestrian. However, this method for extrapolating the road user movement is widely used in the literature.
c. PET was the most reliable parameter for detecting important events. Despite its simple definition, PET has inherent drawbacks in accurately capture conflict severity. Events in the video sequence in which motorists decelerated to near-stop to avoid collision usually have PET values that do not reflect the true severity of the event.

### 5.7 Conclusions

This chapter presented an automated system and methodology that furthers the development of previous work on video analysis to capture the movements of pedestrians at crossing locations. The movement paths of pedestrians and transversal trajectories of vehicles were analyzed and a group of conflict indicators were calculated for each pedestrian-vehicle interaction. The video analysis system developed for this purpose provided the ability to automatically calculate conflict indicators and report important interactions to a human observer for further examination of traffic interactions. The quality of four conflict indicators, Time-to-Collision, Post-

Encroachment Time, Gap Time, and Deceleration-to-Safety Time, were assessed in regard to their ability to comprehend the severity of traffic conflicts. None of the conflict indicators were individually capable of capturing all dangerous interactions between road users. However, a combination of the four indicators proved useful in the identification of important events and traffic conflicts. A possible continuation of this work involves the collection of additional video data at traffic intersections with high pedestrian-involved collision hazard potential. Future work also includes testing, as well as improving, the system's accuracy to detect and track road users in more crowded traffic scenes. As evidenced in this study, there is a need to develop safety measures that address the limitations of current conflict indicators, and draw on the extensive movement data made available by automated methods, such as the automated video analysis system described herein.

## Automated Analysis of Pedestrian-vehicle Conflicts: A Context for Before-And-After Studies

### 6.1 Background

"[Pedestrian exposure to the risk of collision is] very difficult to measure directly, since this would involve tracking the movements of all people at all times" (GreeneRoesel, Diógenes \& Ragland 2007).

The challenge of gaining insight into the mechanism of action that endangers road users transcends the focus on pedestrian exposure to the entire realm of road safety. The accurate estimation of exposure as well as other quantities fundamental to road safety analysis, e.g., severity of a traffic interaction, can
greatly benefit by analyzing road user positions in space and time, i.e., road user tracks (Saunier \& Sayed 2007). Manual annotation of road user positions is time- and resource-expensive, especially when pedestrians are studied, e.g., (Lam \& Cheung 2000) (AlGhadi, Mahmassani \& Herman 2002). As demonstrated in previous chapters, automated extraction of road user positions from video observations has been advocated as a resource-efficient and potentially more accurate alternative.

Video sensors are selected as the primary source of data in this research. Video data is rich in detail, recording devices are becoming less expensive, and video cameras are often already installed for monitoring purpose. Pedestrian tracking in video sequences is traditionally more challenging than other road users (Forsyth et al. 2005). Pedestrians are locally non-rigid, are prone to visual occlusion due to crowdedness, and are more variable in shape and appearance. Despite these challenges, vision-based applications in the field of pedestrian studies have been demonstrated with an increasing level of practical feasibility (see a review of relevant work in Chapter 2). One of the focus areas of pedestrian safety that could greatly benefit from vision-based road user tracking is before-and-after (BA) evaluation of safety treatments. BA studies are a key component of road safety programs that aim at measuring the safety benefits (or absence thereof) derived from a specific engineering treatment.

Catering to the safety of non-motorized modes of travel, in particular for walking, is essential to meeting the ever-growing demand for building a sustainable transportation system. The prevalent collision-based paradigm of

BA studies is based on estimating the reduction in collisions, in terms of frequency and consequence, which can be attributed to the evaluated treatment. In order to draw statistically stable conclusions, e.g., explicating the effect of the treatment away from all other confounding factors, collisions are typically observed for a relatively long period (1-3 years) before as well as after the introduction of the treatment. As presented before, the reliance on collision data for BA analysis invites the following shortcomings:

1. Attribution: police reports and interviews often do not enable the attribution of road collisions to a single cause or a set of causes with satisfactory accuracy,
2. Data Quantity: road collisions are rare events that are subject to randomness inherent to small numbers, and
3. Data Quality: collision records are often incomplete and lack important details, and the quality of road collision reporting has been deteriorating in many jurisdictions.

Shortcomings in collision-based BA studies are even more pronounced in the study of pedestrian safety. Pedestrian-involved collisions are more injurious and less frequent than vehicle collisions (Cynecki 1980). Exposure measures, such as pedestrian volume, are often difficult to obtain and are expensive to collect through in-field surveys (Pulugurtha \& Repaka 2008). Surrogates and/or statistical predictors of these types of data are often used in practice, e.g., (Greene-Roesel, Diógenes \& Ragland 2007). It is often the case that the safety analysis may not afford long-term collision observation after the introduction of a measure (Hua et al. 2009).

Arguments that support the adoption of traffic conflict techniques find more ground in BA studies of pedestrian safety treatments. Traffic Conflict Techniques (TCTs) are based on analyzing the frequency and severity of traffic conflicts at an intersection, typically observed by a team of trained observers. Traffic conflict is defined as: "an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remained unchanged" (Amundsen \& Hydén 1977). Traffic conflicts are more frequent than road collisions and are of marginal social cost. Traffic conflicts provide insight into the failure mechanism that leads to road collisions. BA studies based on traffic conflicts can be conducted over shorter periods. A theoretical framework, advocated in this study, ranks all traffic interactions by their severity in a hierarchy, with collisions at the top, undisturbed passages at the bottom, and traffic conflicts in between (Svensson \& Hydén 2006).

The traditional approach of collecting traffic conflict data is challenged on several accounts. Inter- and intra-observer variability is a common challenge for the repeatability and consistency of results from traffic conflict surveys (Glauz \& Migletz 1984). Field observations are costly to conduct and demand staff training. Despite decades of conceptual developments, there is no universal operational definition of a traffic conflict, e.g., objectively measurable interpretation of words "approach", "risk of" and "unchanged" in aforementioned conceptual definition (Chin \& Quek 1997). Finally, the estimation of objective conflict indicators, such as Time to Collision (Hayward 1968) using field observations can be difficult.

Automating the process of traffic conflict analysis is greatly appealing in the context of BA studies of pedestrian safety treatments. Process automation may enable the analysis of pedestrian-vehicle conflicts in an accurate, objective, and cost-efficient way. The goal of this study is to demonstrate a novel application of automated video analysis for the BA analysis of a scramble phase treatment analyzed manually in previous work (Bechtel, MacLeod \& Ragland 2003). In later stage, the practical use of the developed system as an assisting tool is demonstrated. The length of the video sequence to be reviewed by an observer could be greatly reduced. This study is another step in a research direction that is, to the best of the author's knowledge, unique in the field of road safety and pedestrian studies.

The objectives of this chapter are to first, report several technical improvements to the video analysis system. Second, the work presented in this chapter is also intended to demonstrate the feasibility of conducting BA analysis using video data collected from a commercial-grade camera, from a relatively low altitude, and using a video not collected initially for the purpose of automated video analysis.

### 6.2 Previous Work

### 6.2.1 Conflict-based Before-and-After Studies

There is a significant body of work on the evaluation of pedestrian safety treatments using non-collision data. The literature contain studies that rely on traffic conflicts, e.g., (Van Houten et al. 1997) (Huybers, Van Houten \& Malenfant 2004) (Medina, Benekohal \& Wang 2008) (Gårder 1989) (Kim \&

Teng 2004) (Bechtel, MacLeod \& Ragland 2003) (Acharjee, Kattan \& Tay 2009) and behavioural surrogates such as motorist yielding rate (Turner et al. 2006). The difficulties in relying on collision data in conducing BA studies is acknowledged in the literature, e.g., (Turner et al. 2006) (Hua et al. 2009), in which surrogates safety measures were used. The studies that concerned the evaluation of pedestrian scramble were predominantly conducted using traffic conflicts (Gårder 1989) (Kim \& Teng 2004) (Bechtel, MacLeod \& Ragland 2003) (Acharjee, Kattan \& Tay 2009) (except for (Vaziri 1998)). There is some agreement that scramble phase treatment reduces pedestrian-vehicle conflicts except when pedestrian compliance rate is low (Abrams \& Smith 1977) (Gårder 1989). Among the reviewed studies, the study by (Malkhamah, Miles \& Montgomery 2005) was the only one in which data required for evaluation, motorist deceleration, was automatically collected.

The previously identified issues with the observer-based traffic conflict analysis were echoed by a recent evaluation study of pedestrian treatments in San Francisco (Hua et al. 2009). The authors noted issues with the subjectivity of the definition of traffic conflict, inter-observer agreement, and the labour cost of extracting observations from video data were highlighted. The use of automated video analysis tools is being increasingly advocated to overcome these shortcomings.

### 6.2.2 Video-based Road User Detection and Tracking

The main steps in the procedure of video analysis followed in this chapter are adopted from the methodology presented in Chapter 5 . Some modifications were achieved in order to overcome the challenge of classifying pedestrians
moving during pedestrian scramble. In order to study pedestrian-vehicle conflicts, all road users must be detected, tracked from one video frame to the next, and classified by type, at least as pedestrians and motorized road users. To the author's knowledge, the work presented in Chapters 5 and 6 represent the first attempt to develop a fully functional video-based pedestrian conflict analysis system.

### 6.3 Methodology

Video analysis performed in this Chapter was conducted using the core methodology presented in Chapter 5. Following is a brief description of improvements in the system, mainly to meet video analysis challenges faced in this study.

### 6.3.1 Road User Classification

To analyze pedestrian-vehicle conflicts, it is necessary to identify pedestrians and motorized vehicles. The system described in Chapters 4 and 5 used a speed classifier, a threshold on the maximum speed reached by road users during their existence for classification. This "speed classifier" however proved inadequate for the BA dataset available for this study. This was largely due to the large number of false alarms generated when a pedestrian walks faster than the maximum speed threshold within crowd movement the end of the pedestrian interval or during the pedestrian scramble.

A new method was developed for that purpose, inspired by previous work done by the authors. A small subset of actual road users' trajectories, called
prototype trajectories, is identified using an incremental unsupervised algorithm described in (Saunier, Sayed \& Lim 2007), relying on the Longest Common Subsequence (LCSS) similarity (Vlachos, Kollios \& Gunopulos 2005). The LCSS is a similarity measure that could match tracks of different length.

Let $A$ and $B$ be tracks of two moving objects of lengths $n$ and $m$ respectively, $A=\left[\left(a_{x, 1}, a_{y, 1}\right), \ldots,\left(a_{x, n}, a_{y, n}\right)\right]$ and $B=\left[\left(b_{x, 1}, b_{y, 1}\right), \ldots,\left(b_{x, n}, b_{y, n}\right)\right]$. For a track $A$, let $\operatorname{Head}(A)$ be the sequence of road user positions defined as follows: $\operatorname{Head}(A)=\left[\left(a_{x, 1}, a_{y, 1}\right), \ldots,\left(a_{x, n-1}, a_{y, n-1}\right)\right]$. Given a real number $\varepsilon \geq 0$, the basic similarity measure $\operatorname{LCSS}_{\varepsilon}(A, B)$ is defined as follows (Vlachos, Kollios \& Gunopulos 2005):

- $\quad 0$ if $A$ or $B$ is empty,
- $1+\operatorname{LCSS}_{\varepsilon}(\operatorname{Head}(A), \operatorname{Head}(B))$ if $\left|a_{x, n}-b_{x, n}\right|<\varepsilon$ and $\left|a_{y, n}-b_{y, n}\right|<\varepsilon$, and
- $\quad \max \left(\operatorname{LCSS}_{\varepsilon}(\operatorname{Head}(A), B), \operatorname{LCSS}_{\varepsilon}(A, \operatorname{Head}(B))\right)$ otherwise.

The constant $\varepsilon$ controls the matching threshold for the Chebyshev distance ( $L_{\infty}$-norm) used by default (it is chosen over the Euclidean distance because it is less expensive to compute while yielding good results), but can be replaced by any distance, and more conditions can be added. In this work, a second similarity measure $\operatorname{LCSS}_{\varepsilon, \alpha}(A, B)$, with $0 \leq \alpha \leq 1$, is used by supplementing the trajectories with the velocity at each instant and adding the condition that the cosine of the velocities be below $\alpha$. The associated distances are obtained by scaling the similarities to $[0,1]$ as follows:

$$
\begin{equation*}
D_{\varepsilon}(A, B)=1-\frac{L C S S_{\varepsilon}(A, B)}{\min (n, m)} \tag{6.1}
\end{equation*}
$$

$$
\begin{equation*}
D_{\varepsilon, \alpha}(A, B)=1-\frac{L C S S_{\varepsilon, \alpha}(A, B)}{\min (n, m)} \tag{6.2}
\end{equation*}
$$

where $D$ is the LCSS-based similarity measure or LCSS distance. The prototypes are learnt using $D_{\varepsilon}(A, B)$ to yield a smaller set. The "prototype classifier" uses the 1 nearest-neighbour method with the distance $D_{\varepsilon, \alpha}(A, B)$ and a threshold $\delta(0 \leq \delta \leq 1)$ on the distance to limit the matches to the closest prototypes. The object is assigned the type of the closest prototype. Given that a threshold is used, an object trajectory may have no prototypes with a distance of $\delta$, in which case the default speed classifier is used.

The prototypes need therefore to be labelled. This labelling is a one-time semi-automated operation, where the prototype trajectories are first classified using the speed classifier, then reviewed and corrected if needed by a human annotator. An example of labelled prototypes is given in Figure 6.1. A comprehensive comparison of the classifier on a subset of 1063 manually annotated trajectories was done and the results are presented in Table 6.1 and Figure 6.2. It shows the clear superiority of the prototype classifier over the speed classifier.

Table 6.1 A comparison between the speed and prototype classifiers

| Classifier | Speed <br> Threshold | Max <br> $\boldsymbol{P C C}^{\boldsymbol{I}}$ | Max <br> Kappa | True positive <br> rate $^{\mathbf{2}}$ | False <br> positive rate |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Speed classifier | $2.90 \mathrm{~m} / \mathrm{s}$ | 0.85 | 0.70 | 0.96 | 0.26 |
| Speed classifier with a <br> moving average filter | $2.30 \mathrm{~m} / \mathrm{s}$ | 0.87 | 0.73 | 0.93 | 0.21 |
| Prototype classifier | - | 0.97 | 0.95 | 0.98 | 0.04 |

1 Percentage correct classification (PCC) represents the number of road user trajectories correctly classified (vehicle into vehicle and pedestrian into pedestrian) over the total number of trajectories.
2 A positive is the classification of a road user into a pedestrian and a negative is the classification of a road user into a vehicle. A true positive is a pedestrian classified into a pedestrian (ped-ped). A false positive is vehicle into pedestrian (veh-ped). A true negative is veh-veh and a false negative is ped-veh. The rates are computed by dividing over the number of trajectories in each road user class.


Figure 6.1 Road user prototypes for the before-and-after scramble phase. Figure a) shows the pre-scramble vehicle prototypes(pre-scramble-veh). Figures $\mathrm{b}, \mathrm{c}$, and d show pre-scramble pedestrian prototypes, post-scramble vehicle prototypes, and post-scramble pedestrian prototypes, respectively. The color coding is the result of a k-means clustering in 4 classes based on the prevalent prototype direction.


Figure 6.2 Receiver Operating Characteristic (ROC) Curve for the speed and prototype classifier (for the smoothed max speed classifier, the road user speed is smoothed with a moving average filter). The threshold for the speed classifiers is $3 \mathrm{~m} / \mathrm{s}$.

### 6.3.2 Validation of Tracking Performance

The tracking results of the system need to be evaluated. The safety analysis presented in this Chapter relies on road user tracks. Since most existing research has embraced instantaneous per-frame performance measures, a new algorithm was developed to automatically assign detected objects (the output of the system) to ground truth objects (manually annotated tracks) (Saunier, Sayed \& Ismail 2009).

The results are the unique assignment of these objects: correct assignments (one detected object-to-one labelled object), over-segmentations (one labelled object to many detected objects), over-groupings (one detected object to many
labelled objects), missed detections (unassigned labelled object), and false detections (unassigned detected object). In this Chapter, the results were condensed into correct assignments, missed and false detections, and the performance measure is the following cost function that measures the overall tracking error:

$$
\operatorname{Cost}=\frac{\alpha_{f d}{ }^{*} N_{f d}+\alpha_{m d} * N_{m d}}{N}
$$

where $N$ is the number of annotated objects, $N_{f d}$ and $N_{m d}$ are respectively the number of false and missed detections, $\alpha_{f d}$ and $\alpha_{m d}$ are respectively the weights for false and missed detections, set respectively to 0.25 and 0.75 . The choice of weights is prompted by a target of minimizing missed detections, which might translate into missed pedestrian-vehicle interactions, while still trying to minimize, to a lesser extent, the number of false detections, to reduce the number of falsely detected interactions, called false alarms. This framework was used to optimize the cost function over the space of a few key tracking parameters, namely the connection distance $D_{\text {connection, }}$ the maximum distance between two features for their connection, and the segmentation distance $D_{\text {segmentation, }}$ the maximum difference between the minimum and maximum distance between two features. Data was annotated for 1495 frames, resulting in 41 tracked objects. The space of ( $D_{\text {connection, }} D_{\text {segmentation }}$ ) was searched systematically (refer to Figure 6.3) and yielded the selection of (0.45, 0.12 ). Figure 6.4 presents sample frames with manually annotated data and the result using the automatically tuned parameters.


Figure 6.3 Plot of the cost function with respect to ( $D_{\text {connection }}, D_{\text {segmentation }}$ ).
a) Manual annotation of pedestrian positions.

b) Automatically detected pedestrian objects.


Figure 6.4 Sample frames from validation results. The number of missed detections is $3 / 32$ with 29 false detections mainly due to over-segmentation. Figure a) shows a sample frame from a post-scramble sequence with labelled pedestrians. Figure b) shows the pedestrians tracked in the same frame using the optimized tracking parameters. The bicyclist annotated with a box in Figure b) is correctly identified as a non-pedestrian (given a screen label 'ca').

### 6.3.3 Camera Calibration

The positional analysis of road users requires an accurate estimation of the camera parameters. The camera parameters calibrated in this study are six extrinsic parameters (which describe the location and orientation of the camera) and two intrinsic parameters (which represent the projection on the image space). Once calibrated, it is possible to recover real-world coordinates of points appearing in the video sequence on the pavement surface.

Since videos were collected by a third party, access to the camera was not possible and therefore all camera parameters must to be inferred from video observations and an orthographic image of the intersection. A mixed-feature camera calibration approach was introduced in Chapter 3. Each calibration feature imposes a condition based on its shape, position, and length in both image and world spaces. An additional calibration feature was necessary to enhance the accuracy of the camera calibration based on the parallelism of calculated vertical line (depicted in blue in Figure 6.5) to a manually annotated vertical direction (observed from light poles).

The accuracy of the estimated parameters was tested using a set of 12 line segments, whose true lengths were estimated from the orthographic image. This set of observations was not used in calibration. The calibration error is represented by the discrepancy between calculated and annotated segment lengths normalized by the length of each segment. The accuracy of the final estimates was satisfactory $(0.1 \mathrm{~m} / \mathrm{m})$ and no further error in conflict analysis was attributed to inaccurate estimated camera parameters.

### 6.3.4 Conflict Indicators

Conflict indicators are advocated as an objective and quantitative measure of the severity (proximity to collision) of a traffic event (Svensson \& Hydén 2006). This study concerns traffic events which include a potential conflict between a pedestrian and a non-pedestrian road user. The four conflict indicators calculated in this study are: Time to Collision (TTC), PostEncroachment Time (PET), Deceleration-to-Safety Time (DST), and Gap Time (GT). TTC is defined as "...the time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained" (Hayward 1968).

PET is the time difference between the moment an offending road user leaves an area of potential collision and the moment of arrival of a conflicted road user possessing the right of way (Allen, Shin \& Cooper 1978). GT is a variant of PET calculated at each instant by extrapolating the movements of the interacting road users in space and time (Archer 2004). Deceleration to Safety Time (DST) is defined as the necessary deceleration to reach a non-negative PET value if the movements of the conflicting road users remain unchanged (Hupfer 1997).

An accurate in-field estimation of objective conflict indicators is challenging and inherently subjective. Semi-automated methods have been used in previous studies in which road user positions are manually annotated (Svensson \& Hydén 2006). This process is time-consuming and does not support large-scale data collection.


Figure 6.5 alibration of the video camera. Figures a) and b) show the calibration features. Points are labelled. Segments in red are distance constraints. Segments in blue constitute angular constraints. The inferred camera location is marked. Figures c) and d) show the projection of a reference grid from the world space in c) to image space in $d$ ). World images are taken from Google Maps.

The calculation of conflict indicators in this study follows main lines of Algorithm 5.1. The videos analyzed in this study include significantly large number of road users; especially pedestrian movement during pedestrian scramble. Issues with large data structures arose and the following measures were taken:

1. Road user tracks are extrapolated at their extremities in time by the amount of 3 seconds assuming constant velocity. This extension of the observed road user tracks was conducted to detect conflicts in the further crosswalks of the intersection that occur after vehicle yielding. Vehicles are not tracked when stationary and the image quality at further crosswalks could not enable instant re-tracking when movement is resumed.
2. The list of traffic events to be analyzed is reduced based on the following proximity heuristic:
a. Collect five sample frame numbers selected uniformly from the time span in which the two road users co-exist.
b. Calculate at every point $i$ along a pedestrian trajectory the spacing $S_{i}$ between this point and the potentially conflicting vehicle.
c. Discard this event if $\min \left(S_{i}\right)>10 m$.
3. The remaining list of events is further reduced using the following motion similarity heuristic:
a. For each of the previous sample frame numbers, calculate the smoothed average (window of 10 frames) of the direction of movement.
b. Calculate the angle between the average movement directions of the pedestrian and the vehicle.
c. If the cosine of this angle is greater than 0.9 , discard this event.
4. Road users are assumed to be represented by points, e.g., centroid.
5. The collision area is the point of intersection of pedestrian and vehicle tracks.
6. The objective definition of a collision course is the extrapolation of road user positions that lead to a minimum spacing shorter than the distance traversed by the conflicting vehicle at current speed in 1.5 sec . Extrapolation of road user positions are based on assuming they will maintain a constant velocity.

The tracking parameters used in this study lean toward over-segmentation of road users, i.e., tracking of multiple objects over the same road user. An example is show in Figure 6.6. This increases the chance of tracking of road users, especially pedestrians, at further crosswalks. To reduce this effect, events with calculable conflict indicators that involve road users within a proximity constraint are grouped into one event. This is implemented by creating a graph connecting pedestrian objects and interacting vehicle objects for which there are calculable conflict indicator. All pair-wise spacing between vehicle objects at the moment of their min TTC and min GT are computed. Vehicle objects are further connected if their spacing is below a threshold of 3 m . The subgraph of connected vehicle objects is replaced by a new vehicle object whose resultant conflict indicators are taken as the minima of TTC, PET and GT and the maximum of DST. Details of this grouping are presented in Algorithm 6.1. Figure 6.6 provides additional illustration.

## Algorithm 6.1: Algorithm for grouping pedestrian-vehicle event

Definitions: 1) A pedestrian object $P_{i}$ is $i^{\text {th }}$ in the list of all pedestrian objects that exist in the list of traffic events to be analyzed.
2) A vehicle object $V_{j}$ is $j^{\text {th }}$ in the list of all vehicle objects that exist in the list of traffic events to be analyzed.
Input: Let $V_{j, T T C}$ be the position of the $j^{\text {th }}$ vehicle object at the position that exposed the interacting pedestrian with the shortest Time to Collision (TTC).
Let $V_{j, G T}$ be the position of the $j^{\text {th }}$ vehicle object at the position that exposed the interacting pedestrian with the shortest Gap Time (GT).

Output: An updated list of traffic events in which all successfully grouped events will comprise a unique pair of pedestrian and vehicle objects.

## begin

1- $\quad$ for each pedestrian object $P_{i}$ find within the list of vehicle objects the subset of $n$ vehicle objects $V_{i, j}$ that coexist with $P_{i}$ in the same traffic event.
2- $\quad$ Create an adjacency matrix $A_{n x n}$ that represent the spacing between the positions of every pair of vehicle object $V_{i, j}$ at the time of minimum TTC. Elements in $A$ that correspond to vehicle objects that do not possess a calculable TTC (not on a collision course) are assigned a token value ( 0 ) that is discarded later.

3- $\quad$ Find the connected graphs of all vehicle objects in $V_{i, j}$ in which every pair $l, m$ of connected nodes satisfied the condition $A_{l x m} \leq$ connection_threshold. The threshold is taken 3.0 m in this study.
4- Repeat steps 2 and 3 for vehicle positions at the moment of minimum GT.
5- Combine the list of connected graphs and remove redundancies.
6- Create a new event with TTC at every time step equals the minima at each common time instant of all sequences of TTC observations for all $V_{i, j}$, PET equals the minima of all PET, GT equals the minima of GT observations at every time instant, and DST equals the maxima of all sequence.
7- $\quad$ Remove but one from the list of events all recorders that contains $V_{i, j}$.
8 - $\quad$ Add the new events created in 6 to the list of traffic events to be analyzed.


Figure 6.6 Conflict clustering. Figure a) shows an interaction between a pedestrian and an over-segmented vehicle (tracked twice, object 5638 on the front side and the other 5639 encompasses its horizontal projection). The spacing between these vehicle objects and the pedestrian at minimum TTC and GT are 2.18 m and 1.53 m respectively. Both are below a spacing threshold of 3 m and are therefore grouped. Figure b) shows an illustration of the graph implementation.

### 6.4 Discussion

The analysis of four hours of video was conducted automatically at a pace of approximately one hour of video/day/machine (Intel 1.80 GHz , 2GB Memory, C++ implementation). Sample frames with superimposed road user tracks are shown in Figure 6.7. The spatial distribution of traffic conflict positions is shown in Figure 6.8. A conflict position is taken as the location of the conflicting vehicle at the moment when there was a minimum time separation from the pedestrian. The time separation is measured by TTC as well as GT. There is an evident change in the density of traffic conflicts per unit area and time. The spatial distribution of traffic conflicts migrated away from the crosswalks after the scramble phase. The density of traffic conflicts per unit area was also reduced.

The distributions of the calculated conflict indicators before-and-after scramble are shown in Figure 6.9. To obtain robust measurements of individual conflict indicators, the following modifications were applied in addition to the measures described earlier in the paper:

1. When calculating TTC, the average of the individual TTC calculated for the conflicting pedestrian and vehicle is used to represent the event. The difference in individual TTC measured for the conflicting road users arise because they may satisfy the proximity condition for a collision although their trajectories do not precisely meet at the same point. A modified spatial proximity measure was introduced to represent a collision instance. A collision is defined when the pedestrian and the conflicting vehicle become closer than 2.0 m .
2. The average of the 10 most extreme values is used to calculate min TTC, $\min$ GT, and max DST. The extreme values for GT are calculated based on their absolute value. The representative average is given the sign of the most extreme value.
3. Events with PET value less than a pre-defined noise threshold of 0.25 sec were eliminated. There were no collisions observed in the video sequence and these measurements were mainly due to tracking noise.

There is an evident reduction in the frequency of traffic conflicts. It was not attempted to conduct statistical analysis of this data for three reasons:

1. Validation of the video analysis system on this data sequence was not comprehensively conducted to measure the reliability of the estimates. To some extent meet this demand, a random sample of 366
automatically detected traffic events (266 for the before period and 100 events for the after period ${ }^{1}$ ) was considered for manual review. Among the 266 events in the before period, 13 events were found to involve classification error. In addition, 4 more events involved the misclassification of cyclists as pedestrians although the former appears to move with the mainstream pedestrian movement. Among the 100 events reviewed for the after period, 12 events were found to involve misclassification of pedestrians into vehicles.
2. It is not clear how the severity of traffic events measured by the calculated conflict indicators should be inducted in a statistical analysis.
3. It was not possible to account for confounding factors that may have affected the safety level other than the concerned treatment.

Misclassification of pedestrians into vehicles was still evident, however at a much lower frequency than speed-based classification. Figure 6.7 shows a sample frame in which a pedestrian is misclassified as a vehicle while walking in a scramble phase. However, the issue of road user misclassification was greatly marginalized by the aforementioned prototype classification. The lack of an inference mechanism that is based on the microscopic traffic conflict analysis conducted in this chapter motivated the work presented in Chapter 7. The issue of partial and/or multiple tracks of road users is probably the major challenge for the analyzed data. The degree to which Algorithm 6.1 was effective in addressing this issue was not

[^13]investigated. However, more work is likely required to mitigate the effect of over-segmentation and incomplete tracks.

### 6.5 Conclusions

This study carried out in this chapter demonstrated the feasibility of conducting before-and-after evaluation of pedestrian safety measures using automated analysis of video data. Pedestrian tracking in video data is an open problem for which some improvements have been investigated. The reliance on motion prototypes achieved a clear advantage over classification methods used in previous studies.

The context of this study is the evaluation of the safety benefit of the introduction of the pedestrian scramble phase. A two-hour video sequence was analyzed for pre- and post-scramble. Despite that the video analyzed in this study was not collected initially for the purpose of automated analysis, tracking accuracy was satisfactory. The automated analysis of four conflict indicators shows a reduction in conflict frequency. In addition, there was a general reduction in the spatial density of conflicts after the safety treatment.

It was not attempted in this study to draw a statistical inference regarding the safety benefit of the pedestrian scramble. It represents an important continuation of this work, and potentially a different paradigm of safety diagnosis that considers the frequency as well as severity of traffic events. A framework for safety diagnosis places all traffic events on a continuum of severity from uninterrupted passages to traffic collisions (Svensson \& Hydén 2006). Such framework can clearly benefit from automated video analysis. An
important continuation of this work can also be to conduct a comparison between the severities of traffic interactions measured by the system against expert rating.


Event :2939 objects: 1501 | 5074 | TTC :1.162 PET :2.8 max DST :7.260 min GT :1.105

Event :2609 objects :1317|4966| TTC :3.318 PET :2.266 max DST :3.355 min GT :1.198


Event :2913 objects :1496|4992(unseen)| TTC :1.815 PET :0 Event: 2306 objects: $1313 \mid 4812$ TTC 0 PET 2.07 max DST : 0.437 min GT : 1.113 -0.075 GT 2.83


Event :2372 objects :1167 :4804 | TTC :2.973 PET :0 max An occurrence of misclassification DST :-0.379 min GT :1.473

Figure 6.7 Sample frames with automated road user tracks. The captions display "Event" the event order in the list of potential interactions, "objects" the numbers of the interacting objects, and the indicated conflict indicators.


Intensities are in number of conflict positions per square meter per 2 hours.
Figure 6.8 Before-and-after spatial distribution of traffic conflicts. A conflict positions is selected as the position at which the motorist was separated by either a minimum Gap Time (GT) or minimum Time to Collision (TTC). Figure a) shows the before spatial distribution of conflict locations based on min GT. Figure b) shows the after distribution of conflict positions based on min GT. Figure c) shows the before distribution of motorist position at min TTC. Figure d) shows the after distribution of conflict positions based on min TTC.


Figure 6.9 Distribution of different conflict indicators values for before and after scramble phase. Analyzed video durations are 2 hours before and 2 hours after. $|\mathrm{PET}|$ and $|\mathrm{GT}|$ are the moduli (unsigned) values of the Post Encroachment Time and Gap Time conflict indicator.


# Methodologies for Aggregating Traffic Conflict Indicators 

### 7.1 Background

This section presents a number of arguments that describe the motivation behind methodological developments presented in subsequent sections. A theoretical background is provided at the beginning, followed by a sequence of arguments that end with the central hypothesis of this chapter.

### 7.1.1 Theoretical Preliminaries

In theory, any road user navigating a traffic intersection is expected to become exposed to the risk of collision. It is implausible that every contributing factor to the risk of collision can be marginalized through meticulous road and
vehicle design. There always exists a possibility for the driver to suffer from distraction or lapse in concentration and for the vehicle control process to be compromised. Meanwhile, there is always a chance that the road geometry will not forgive such driving mistakes or mechanical failures. It has been suggested that for every passage of a motorist within the domain of a traffic facility, there is a chance set-up for a collision to take place (Hauer 1982). Furthermore, every road user passage can be seen as a trial that is tested by some underlying chance of failure. The precise mechanism that exposes every road user to the risk of collision is not directly observable and can only be inferred from empirical realizations in the form of road collisions or traffic conflicts. The mainstream methods of road safety analysis rely on the analysis of road collision observations. From a chance set-up perspective, the extent of the chance of a road collision may be inferred from observing the number of actual collisions relative to the total number of trials.

While the intuition behind the chance set-up concept is sound, its implementation is forbiddingly difficult. Due to the rarity and randomness of road collisions, reaching an accurate estimation of the underlying chance of collision requires prolonged observational periods. Moreover, there is no clear definition of what constitutes a trial. It is implausible that for every pair ${ }^{1}$ of road users co-existing within a traffic facility, there exists a conceivable chain of events that leads to a collision. It is arguable that a vehicle stopping at an intersection approach does not pose any reasonable risk of collision to a vehicle navigating the intersection during the latter's permitted phase.

[^14]Similarly, a pedestrian stepping down to a crosswalk may not become exposed to the risk of collision with all vehicles currently present at the intersection. Therefore, it can be argued that a genuine trial must entail a reasonable chain of events that may occur for the chance of collision to materialize.

The previous definition of what constitutes a genuine trial, or real exposure to the risk of collision, is undeniably subjective. Exposure is an abstract concept that commonly defines events at which specific agents come in contact with a source of hazard. This broad definition is applied to disciplines such as industrial safety, epidemiology, and road safety. In the context of pedestrian safety, the source of hazard concerned is the potential for dangerous physical contact with a non-pedestrian road user. Exposure in this context can be defined as the number of traffic events for which there existed a reasonable chain of event that could lead to physical contact between road users. It has been argued that the only direct way to measure pedestrian exposure, and arguably the exposure of other road users, is by tracking pedestrians and conflicting vehicles all the time (Greene-Roesel, Diógenes \& Ragland 2007). Without the use of tracking mechanism, for example computer vision techniques, the latter objective is infeasible. Because the adoption of advanced tracking techniques is very recent to the discipline of road safety, several proxy measures of exposure have been proposed in the literature. Examples are time (Hauer 1982), pedestrian volume (Davis, King \& Robertson 1988), the product of pedestrian volume and vehicle volume (Cameron 1977) and the square root of this product of volumes (Greene-Roesel, Diógenes \& Ragland 2007). However, these surrogates suffer invariably from common conceptual
limitations. For example, the naïve product of the volumes of conflicting streams can lead to over-estimating exposure. Understandably, not every pedestrian using a crosswalk is endangered by every other vehicle in the conflicting vehicular stream. Another limitation to using traffic volume as a measure of exposure is its remoteness from representing the true number of events in which the pair of road users could have potentially collided. One immediate benefit of the extraction of road user tracks from video sequences, as demonstrated later in this chapter, is providing a significantly more accurate measure of exposure.

### 7.1.2 The Severity Dimension

In this section, it is hypothesized that collision risk and severity of a traffic event are conceptually equivalent. In order to proceed with this argument, it is important to adopt explicit definitions. Risk is an abstract concept that can be defined as the product of the probability of an undesirable outcome and its consequences. Similarly, the severity of a traffic event comprises two aspects, the proximity to collision and the physical damage from which road users will suffer if they collide. A frequentist interpretation of the probability of collision defines it as the rate at which collision events take place if the concerned traffic event is repeated for an infinitely large number of times. This interpretation however does not precisely reflect the reality of traffic events. Given the numerous characteristics of the conflicting road users, it is unrealistic that a specific traffic event is in any way repeatable. Bayesian interpretation of the probability of collision possesses more relevance to traffic events. Following a Bayesian interpretation, the probability of collision can be
defined as an abstract chance of this undesirable occurrence. A Bayesian model assumes that a traffic event is one realization of an underlying generative model for which an a priori knowledge is available. The probability of collision can be determined by inferring the matching, up to a likelihood value, between the positions of each road user and some underlying generative distribution, or a movement prototype.

Indeed, road users exhibit stereotypical movements due to regularized geometry, lane marking, crosswalk delineation, and temporal separation by traffic signals. The use of motion patterns to predict road user positions has been a novelty in the work by (Saunier \& Sayed 2008). Conceptual refinement of the previous work was briefly described in section 2.1.1 and is presented in more detail in section 9.3. Extrapolation of road user positions proposed by (Saunier \& Sayed 2008) is one of many forms of uncertainty that encapsulate the probability of collision. In addition to the uncertainty of future road user positions, other uncertainties still exist regarding the type of evasive action to be taken and the consequences of the potential collision. It is possible to categorize these uncertainties into three types:
a. Extrapolation. This type of uncertainty pertains to the knowledge of all possible road user positions given that they remain unaware of the impending collision. Examples of measures of this uncertainty are the work by (Saunier \& Sayed 2008) (Saunier, Sayed \& Ismail 2010) as well as various proximity measures proposed in the literature, see citations in (Archer 2004) (Laureshyn 2010).
b. Evasive action. This uncertainty concerns the particular set of actions a road user may perform in order to avoid collision. All reviewed studies on this subject assume that the only evasive action available to road users is deceleration (Davis 2003) (Cunto \& Saccomanno 2008) (Davis, Hourdos \& Xiong 2008). However, it is recognized that in case of pedestrian-vehicle conflicts, deceleration, swerving, or a combination of both may constitute evasive action taken by motorists or pedestrians (Malkhamah, Miles \& Montgomery 2005). In general, the set of possible evasive actions should conceptually expand beyond deceleration action.
c. Consequences. This uncertainty concerns the estimation of the physical consequences of the potential collision. Various surrogates were developed in the literature to represent the severity of a collision. For example, relative speed between road users (Hydén 1987) or the strength of the required evasive action (Oh et al. 2010).

It has been argued that measuring the probability of collision of individual events is the ideal bottom-up approach for safety modelling (Lord, Washington \& Ivan 2005). The prevalent approach for calculating probability of collision is based on aggregate-level statistical modelling, most prominently the modelling the occurrence of collisions as a Poisson process. This approach however is mainly driven by model-fit improvement without insight into either the mechanism of action between interacting road users or into the heterogeneity of traffic events in terms of the probability of collision entailed in each event. A disaggregate measurement of the probability of collision of all traffic events is the precise solution to the shortcoming with
aggregate-level collision prediction models. As previously argued, severity of a traffic event can be jointly represented by both proximity to and consequences of collision. The universal severity dimension theorized by (Hydén 1987) can be interpreted as the risk of collision measured for each traffic event. Moreover, it is argued that the risk of collision, both proximity and consequences thereof, is the mechanism used intuitively by field observers to subjectively measure severity of traffic events (Laureshyn 2010).

Despite the theoretical appeal for calculating probability of collision, such an undertaking is challenged on practical grounds. The dimensionality of collision events, especially when considering driver behaviour, is indefinite. In fact one can argue that a precise calculation of this probability is not feasible. Two main challenges stand before achieving this objective. The first challenge facing the development of a probabilistic conflict indicator is conceptual; there is no formal model in which the three uncertainties can be combined. The second challenge is the most daunting. The data required to develop and validate such probabilistic models is proportional to its complexity. For example, to model the uncertainty regarding the type and strength of evasive action, a comprehensive library of evasive actions taken by road users involved in a variety of traffic conflicts is required. To the author's knowledge, no data for this purpose has been gathered.

Given the remoteness of a comprehensive measurement of the probability of collision, surrogates are commonly used in the literature. Example of these surrogates are objective severity measures, whether deterministic objective
conflict indicators ${ }^{2}$, such as Time to Collision (TTC), Post Encroachment Time (PET), Gap Time (GT), and Deceleration to Safety Time (DST), or probabilistic indicators such as the severity index proposed by (Saunier, Sayed \& Lim 2007). However, none of these conflict indicators formally combines all three types of uncertainties.

A multitude of other conflict indicators have been developed for the purpose of traffic conflict observation. A recent review of pedestrian-vehicle conflict indicators yielded a set of 54 different indicators (Laureshyn 2010), the majority of which are calculable using positional data. Based on this review, it was noted that there is little work on investigating the different severity aspects that each conflict indicator represents. The proximity of various conflict indicators to the true probability of collision appears to be presumed based on commonsense. It is undeniable that there is some construct validity of these conflict indicators in representing the probability of collision. However, it is not known for fact whether they comprehensively represent severity or whether they reflect partial aspects of severity.

### 7.1.3 Conflict Indicators as Partial Images

A number of conflict indicators have been advocated as the preferable mapping from positional and temporal attributes of conflicting road users to the theorized severity dimension (Hydén 1987)(Svensson \& Hydén 2006). The main advantage of conflict indicators is their objective nature which constitutes a significant advantage over subjective severity measures in terms of consistency of measurements. More precisely, the advantage is the reliability

[^15]of objective conflict indicators over other severity measures. Reliability of measurement refers to the invariance of the conflict indicator to all factors extraneous to positional and temporal attributes of road users. For example, if the tracks of a pedestrian and vehicle are known, their TTC is calculable regardless of the time, the location, and the traffic context of their interaction. However, some of the factors eliminated from consideration in evaluating conflict indicators may in reality be relevant to the true severity of the concerned traffic event.

For example, identical TTC values may be obtained regardless of whether the conflicting road users are aware of each others. Intuitively, there is a higher chance for an event involving road users unaware of each others to develop up to higher severity levels. The coupling of subjective assessment of traffic events and conflict indicator measurements has been reported in the Malmö study (Grayson et al. 1984). In this study, different teams were asked to measure the severity of a common set of traffic events in order to gauge the inter-observer as well as inter-technique variability. It was noted in this study that observers incorporated in their assessments severity aspects other than TTC and Post Encroachment Time (PET). That is, observers found it necessary to complement the measurement of conflict indicator values, putting aside the questionable accuracy of observer-based measurement of positional data, with other contextual variables. Moreover, a weak agreement between conflict indicator values and subjective assessment of the severity of traffic events was found (Shinar 1984). Furthermore, in validating the Swedish Traffic Conflict Technique it was found that serious traffic conflicts rated as such by subjective
human assessment was in stronger correlation with collisions than serious conflicts rated by objective conflict indicators (Svensson 1992).

Based on evidence in the literature, it is plausible that various conflict indicators appear to represent partial images of the true severity of traffic events. Not surprisingly, the trained observer appears to be able to fathom much closer to the true severity of traffic events than conflict indicators based solely on positional data. Unfortunately, the observer provides this measure at much lower reliability than is sufficient to establish a sound practice of traffic conflict analysis.

Another key study on the correlation between collision and traffic conflicts adopted a combination of TTC and a subjective observer-based severity assessment of traffic conflicts (Sayed \& Zein 1999). In this study, field observers were required to record both TTC and their assessment of the risk of collision involved in each traffic conflict. The severity aspects covered by this subjective risk measure were the "seriousness of the observed conflict [,]... the perceived control that the driver has over the conflict situation, the severity of the evasive maneuver and the presence of other road users or constricting factors which limit the driver's response options". The subjective risk measure was introduced to supplement the intrinsic shortcoming of TTC in order to comprehensively representing severity of traffic conflicts.

An example of the inability of conflict indicators to comprehend subtle contextual differences is presented in Figures 7.1a and 7.1b. The two sample frames shown in Figure 7.1 display two apparently similar traffic events with comparable minTTC values. Isolating the measurement of conflict indicators
from the context within which the traffic event occurred could lead to counterintuitive severity measurements. Figure 7.1a shows a pedestrian that arrived early to the crosswalk during the pedestrian scramble phase. The conflicting motorcycle had not arrived during a permitted phase and was recorded while decelerating before the stop line. Moreover, it is unreasonable the motorcyclist would accelerate through the cluster of pedestrians blocking its way to the intersection. Within this context, a human observer may discount the severity of the traffic event shown in Figure 7.1a. Conversely, the pedestrian shown in Figure 7.1b arrived late to the pedestrian scramble and is shown running to clear the intersection. A human observer may rightfully rank the event in Figure 7.1a at much less severity than the event in Figure 7.1b. This severity differential may not be captured by minTTC which is low for both cases. Note that a heuristic commonly entertained such as excluding motorized vehicles behind stop lines from severity evaluation will not work in these cases.
a) $\min \mathrm{TTC}=3.1 \mathrm{~s}$

b) $\min \mathrm{TTC}=2.2 \mathrm{~s}$


Figure 7.1 Two events of only subtle difference in context and comparable values of minimum Time to Collision ( $\min$ TTC). The subtle difference in context however entails significantly different severity.

The previous shortcomings relate to the validity of conflict indicators. Validity concerns the ability of a conflict indicator to comprehend the true severity of a traffic event and truthfully represent the risk of collision to which road users are exposed. While extensive work has been performed on the validity of traffic conflict techniques, surprisingly little work has been done on validating the entire set of conflict indicators proposed in the literature. Previous work has been conducted on the validation of TTC against PET in comprehending the severity of traffic events in which the latter conflict indicator was favourable (Grayson et al. 1984). Little, if any, investigation has been conducted on the validity of other conflict indicators abound in the literature. As they stand, conflict indicators reflect different and sometimes independent severity aspects. It is however possible to group conflict indicators into two classes. The first class requires the presence of a collision course, and the other that measures the mere spatial and temporal proximity of conflicting road users.

The first class of conflict indicators, and potentially the more developed, measures the proximity to a collision point. Examples of the first class are TTC and probabilistic representation thereof (Saunier, Sayed \& Ismail 2010). A variant of TTC, Time to Accident, has been extensively used in the development of the Swedish Conflict Technique and has been validated for this purpose (Svensson 1992). Most notable of the second class is PET, which represents the observed temporal proximity of the conflicting road users. PET has been adopted in another key study in which it was proven to be a reliable predictor of road collisions if observed over an extended period of time (Songchitruksa \& Tarko 2006). The two conflict indicators, TTC and PET,
however do not represent the same collision mechanism. Arguably, they reflect different partially overlapping severity aspects.

TTC represents the proximity of conflicting road users to a potential collision point, while PET represents their proximity to each other. Generally, TTC is more suited to comprehend the severity of traffic events that involve the risk of rear-end collision. PET is of little validity in this case since it is dependent on the speed of the lagging road user as opposed to their relative speed. PET is better suited for representing the severity of crossing events. The two conflict indicators are not necessarily calculable for all events. Moreover, when calculable, they may represent variant severity measurements. For example, a vehicle that avoids collision with a pedestrian by decelerating and coming to a stop may have an unrepresentatively long $\mathrm{PET}^{3}$. On the other hand, road users involved in a dangerous and proximate crossing may have not been on a collision course, and therefore have no calculable TTC. The reasoning applied to contrast TTC and PET can be extended to other conflict indicators to prove the existence of situations in which they yield variant severity measurements.

The extended discussion presented in this section conceivably leads to the main hypothesis of this chapter:

> Hypothesis 7.1: Conflict indicators measure partially overlapping and sometimes independent severity aspects of traffic events.

[^16]
### 7.1.4 Aggregation of Road Safety Cues

The level of road safety is an abstract concept that can only be inferred from measurable indicators. Road safety indicators can be qualitative or quantitative. Due to the abstract nature of road safety and the immense dimensionality of its characteristic space, different indicators may provide various and often independent aspects of road safety. Ultimately, important road safety treatment decisions must be taken based on a singular inference on the underlying level of safety. A number of systematic approaches have been proposed to combine different road safety cues into composite indices, e.g., (Al-Haji 2005) (Hermans, Van den Bossche \& Wets 2009). A theoretical framework was proposed for the general development of composite road safety indicators (Hermans, Van den Bossche \& Wets 2009). A central component of safety index development is the normalization, weighing, and aggregation of different indicator values. Previous developments focused on the integration of different road safety cues into macroscopic safety indices. The same reasoning and theoretical framework can be adopted at the microscopic level of individual traffic events. As opposed to a single conflict indicator, a set of conflict indicators can be used to measure the severity of traffic events. Different conflict indicators can be integrated in order to obtain a more accurate measure of the severity of traffic events. The main objective of this chapter is to present a new methodology for integrating different conflict indicators into a severity index. A case study is presented within a before-and-after context using the traffic conflict data obtained from automated video analysis of the pedestrian scramble safety treatment discussed in Chapter 6.

### 7.2 Methodology

In general, conflict indicators concern the measurement of spatial proximity, temporal proximity, or a combination thereof. The values of conflict indicators can be directly interpreted based on experience or in reference to some severity thresholds. Accordingly, conflict indicators can be viewed as mappings from a range of positional measurements of conflicting road users onto the severity dimension. Techniques in the literature proposed the division of the severity dimension into discrete severity levels. All points with the domain of a severity level are considered of uniform value. Some techniques proposed a division into 30 severity levels, e.g., (Svensson 1998), while other techniques defined only two categories: mild and severe, e.g., citations for PET and Deceleration to Safety Time (DST) in (Malkhamah, Miles \& Montgomery 2005). The following sections present two methodologies for measuring the severity of traffic events using a set of conflict indicators. The last section discusses the problem of aggregating severity measurements conducted at the traffic-event level using three different exposure measures.

### 7.2.1 Integration of Conflict Indicators

This section concerns the problem of integrating different severity cues measured into a single severity index. Individual severity cues are provided by conflict indicator measurements. Two methodologies are presented in this section for integrating different conflict indicators and for mapping their composite values into the severity dimension: single-step integration and multi-step integration.

## Integration Approach A: Single-step Integration.

In this approach an integration function $\Psi($.$) is constructed to map a set of$ conflict indicator values into the severity dimension. Let $x_{1}, x_{2}, \ldots, x_{n}$ be the individual values of $n$ conflict indicators, then the severity value represented by these conflict indicators is constructed as follows:

$$
\begin{equation*}
I=\Psi\left(x_{1}, x_{2}, \ldots x_{n}\right) \tag{7.1}
\end{equation*}
$$

where $I \in[0,1]$ is a dependent variable which domain is the severity dimension. In subsequent sections of this chapter, $I$ is referred to as some severity index. The calibration of this integration approach requires reference severity measurements of a large sample of traffic events. This type of data is currently unavailable. Therefore, this approach has not been implemented in this chapter.

## Integration Approach B: Multi-step Integration

In this approach each conflict indicator value $x_{i}$ is independently mapped to the severity dimension by an individually defined mapping function $\Psi_{i}\left(x_{i}\right)$. The last step is to draw a representative value from the set of individual mappings of different conflict indicators. Following are proposals of representative values:

$$
I= \begin{cases}\frac{1}{n} \sum_{\mathrm{i}} \Psi_{\mathrm{i}}\left(x_{i}\right) & \text { approach } \boldsymbol{B} \mathbf{1}  \tag{7.2}\\ \max _{i}\left(\Psi_{i}\left(x_{i}\right)\right) & \text { approach } \boldsymbol{B} \mathbf{2} \\ \text { quantile }\left(\Psi_{\mathrm{i}}\left(x_{i}\right)\right) & \text { approach } \mathbf{B 3}\end{cases}
$$

The multi-step integration approach can be viewed as a special case of the single-step integration. The interpretation of both is however distinct. The first
integration approach (approach A) considers the interdependence of different conflict indicators in representing severity. The second set of approaches (B1, B2, and B3) assumes that every conflict indicator provides a unique and independent severity measure. In multi-step integration, it is necessary to draw a representative value from the individual mappings of conflict indicators. Equation 7.2 provides sample strategies for drawing representative values from individual mappings of conflict indicators.

Selecting the average of individual mappings (approach B1) of conflict indicators is favourable when: 1) comparable validity in representing the true severity of a traffic event is assumed for every conflict indicator and 2) differences in severity among conflict indicators are attributable to random road user characteristics. For example, TTC and PET satisfy these conditions since each of them measures independent proximity measures. This independence can be further extended to the two other conflict indicators analyzed in Chapters 5 and 6.

By definition, PET, DST and Gap Time (GT) do not require the presence of a collision course. Therefore, there is no plausible reason to believe that the presence of collision course will influence these three conflict indicators. At the event when a collision course is present, TTC measures the proximity of conflicting road users to a collision point, while PET, DST, and GT measure the temporal proximity of road users to each other while traversing the collision point (precisely the conflict zone defined around a collision point). One can obtain variable TTC values for the same GT and DST values, and vice versa. One can also obtain various TTC values for the same PET value, mainly
because the latter depends on actual, as opposed to extrapolated, road user positions. Moreover, PET, DST, and GT are defined in terms of entrance to and exit from a conflict zone. The boundaries of the conflict zone are dependent on the angle between the trajectories of road users. The size of the conflict zone is negatively proportional to the angle between directions of conflicting road users. The time of entry to and exit from the conflict zone is dependent on the boundary definition of the conflict zone. It can be argued that, temporal separation proximity measures such as PET, DST, and GT will tend to underestimate the severity of a traffic event for acute-angle interactions.

Using elementary geometry, it is possible to approximate the relationship between DST and GT in the case of pedestrian vehicle conflicts as follows:

$$
\begin{equation*}
D S T=\frac{2 v_{v}}{t_{p}^{2}} * G T \tag{7.3}
\end{equation*}
$$

where $v_{v}$ is instantaneous vehicle speed and $t_{p}$ is the time expected for the pedestrian to reach the centroid of the conflict zone. The relationship between the two conflict indicators is dependent on the value of the proportionality variable $\frac{2 v_{v}}{t_{p}^{2}}$. The only situation when this proportionality variable is stable around a specific number is when motorists drive at increasingly higher speeds while pedestrians are further back from a potential conflict zone. Conceivably, there are traffic configurations in which road users approach each other at speeds independent from the time proximity of the conflicted road user. For example, in the case of pedestrian scramble presented in Chapter 6, pedestrians are conflicted with through as well as right-turn vehicular traffic. The average approach speed of through movements is likely
to be higher than right-turn movement. There are situations also when this case is reversed. An example is when the intersection exit is blocked for through traffic by a queue build-up.

The adoption of an extreme value of individual conflict indicator mappings, for example the maximum or the minimum value, implies the variability among conflict indicators in comprehending the severity of the concerned traffic event. For example, if it is the case that various conflict indicators, in independence of each other, tend to underestimate severity, then drawing the maximum of individual mappings is more suited than drawing the average value, as is entailed by approach B2. It is straightforward to show that if severity is misrepresented by overestimation, then selecting a minimum value is a more accurate representation of the true severity.

Integration approach B2 may however lead to erroneous severity measurement in the case when extreme values are induced by tracking errors. In order to mitigate this potential for error, an order statistic or quantile value may be used as an approximate to estimated extreme value (approach B3). However, in situations when few conflict indicators are used, the use of an order statistic may be challenging.

### 7.2.2 Mapping to Severity Indices

The previous section discussed the integration of different individual mappings of each conflict indicator $x_{i}$, represented by the corresponding function $\Psi_{i}\left(x_{i}\right)$. The construction of these mappings is closely tied to the validity of the corresponding conflict indicators in representing true severity. If there exists a mapping defined over the range of a conflict indicator that
will correctly represent true severity, then the validation of this mapping is proven. As discussed earlier in section 7.1.3, little work can be found in the literature on validating conflict indicators. In general, the severity interpretation of different conflict indicators is based on experience and judgement. Two main mappings are proposed in this section: function mapping and distribution mapping. The mapping development was restricted to four conflict indicators: TTC, PET, DST, and GT. The mappings are also restricted to measuring the severity of pedestrian-vehicle conflicts, conceivably from the perspective of pedestrian safety. Following subsections provide more in-depth discussion of these mappings.

## Function Mapping

In this mapping approach, closed-form functions are established in order to map individual conflict indicators into severity indices. At first attempt, the development of such functions was based on the author's subjective interpretation of conflict indicator values (Ismail, Sayed \& Saunier 2010). A different set of mapping functions were developed in this chapter based on a calibrated mapping function that yields severity values consistent with benchmarks values in the literature. Following are the functional forms of the mappings:

Functional Mapping Approach 1: subjective function parameters.

$$
\begin{array}{rr}
I(\mathrm{x})=e^{-\frac{x}{p 1}} & \forall \mathrm{x}=T T C>0 \\
I(P E T, G T, \text { orDST })=e^{-p 2 *\left(p 3 * x+e^{-p 3 * x}-1\right)} & \forall \mathrm{x}=P E T, G T, \text { orDST }>0 \tag{7.5}
\end{array}
$$

where $p 1, p 2$, and $p 3$ are specific mapping parameters that define its shape, $I($.$) is the mapping function that takes the value of the conflict indicator as an$ argument and outputs a severity index that ranges from 0 for events with no reasonable exposure to the risk of collision and 1 for all collisions. Note that the proposed mapping is by construction unable to comprehend the variable severity of collision events.

The functional form selected for Equation 7.4 is adopted from a similar formulation by (Hu et al. 2004). The functional form presented in Equation 7.5 is adopted from a generic development of penalty functions with minor modification to yield an indexed value (Ronold \& Christensen 2001).

## Functional Mapping Approach 2: Benchmark-based function parameters.

The functional forms proposed in Equations 7.4 and 7.5 are preserved. The same functional form in Equation 7.5 is used for mapping PET, GT, and DST. The function parameters were calibrated based on the severity benchmarks in the literature shown in Table 7.1.

Table 7.1 Severity benchmark values for constructing mapping functions

| Conflict <br> Indicator | Severity Level | TTC (sec) | PET/GT (sec) | DST (m/s ${ }^{\text {2 }}$ ) | Severity Index |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 00000000000 | 1 (highest) | 1.6 | 3 | 1 | 0.8 |
|  | 2 | 5 | - | 2 | 0.6 |
|  | 3 | 8 | - | 4 | 0.4 |
|  | 4 (lowest) | 11 | 8.5 | 6 | 0.2 |

Thresholds in the literature found for severe conflicts were used to demarcate the highest severity level in Table 7.1 which was selected to be 0.8 . Three more thresholds were selected for lower severity thresholds. Other TTC values in Table 7.1 were selected from the severity measures found in the Swedish Conflict Technique (Svensson 1998) assuming a constant conflicting speed of $20 \mathrm{~km} / \mathrm{h}$ and assuming that the highest severity level of 30 corresponds to a severity index of 1 . The highest severity threshold for GT and PET as well as severity thresholds for DST were reported in (Malkhamah, Miles \& Montgomery 2005). Two assumptions were made in Table 7.1 outside of relevant findings in the literature. The first is the division of severity levels to five subdivisions, constituting five homogenous and successive severity intervals. Second, the least severe temporal proximity for PET/GT is selected to be the time consumed for a pedestrian to walk value corresponding 10.0 m . The spatial proximity threshold is intrinsically defined in the calculation of conflict indicators to demarcate the boundary between exposure events and uninterrupted passages. Exposure events were selected for further proximity analysis while uninterrupted events were discarded for this purpose. Refer to Chapter 6 for further details on the conflict indicator calculations.

## Distribution Mapping

The idea behind this mapping is to represent severity by the relative frequency of a conflict indicator value. Ideally, if a large-scale pool of conflict indicator measurements is available, relative frequency will be closely related to the anomaly in conflict indicator value. According to the severity hierarchy theory as well as empirical evidence in (Archer 2004) (Svensson 1998), severe events are observed with low frequency. The pool of conflict indicator
measurements used to establish the distribution of the four conflict indicators was obtained from the work presented in the Chapter 6. Instead of using empirical cumulative distributions, which could be expensive to calculate, a Gamma distribution was fit to the conflict indicator observations. To deal with negative values for PET and GT, two sets of distribution parameters were estimated from positive and negative conflict indicator values. For negative PET and GT, their absolute values were used to estimate the set of distribution parameters for negative conflict indicator values.

## Asymptotic Argument

Experience and observational judgement were mainly used to develop severity benchmarks for the different approaches to function mapping. However, the definition of these severity benchmarks must to some extent be related to the relative frequency of conflict indicator values. It can be argued that given a larger pool of conflict indicator measurements and a larger pool of expert opinion on the subjective severity of traffic events, function and distribution mappings will converge to a unique mapping. Furthermore, it is likely that such mapping will be dependent on other contextual variables. The sample of conflict indicator measurements used in this chapter was limited to a total of six hours. This can barely provide representative benchmarks for measuring the abnormality of conflict indicators. Figure 7.2 displays function and distribution mappings for different conflict indicators. As is shown, no convergence to a unique mapping was evident. Therefore, further analysis was pursued using independent application of both function and distribution mappings.


Figure 7.2 A depiction of two mappings from conflict indicators to severity index. Shown also are the parameters for function mapping 1. Mapping parameters (p1:p5) are shown in the legend that were collected from benchmarks in the literature. For example p1 = 8.0.

### 7.2.3 Aggregation of Severity Measurements

The main objective of calculating conflict indicators is to obtain severity measures at the traffic-event level, or alternatively at a microscopic level. The bottom-up approach for safety analysis advocated in this thesis must provide at the end some inference on the underlying level of safety. This entails a significant reduction in dimensionality of the data to few and even single quantities. Little statistical work has been conducted on drawing an inference about the level of road safety from the severity distribution of traffic events. The only work found on this subject was in the context of before and after studies (Svensson 1998). The statistical analysis was mainly based on testing the difference in shape between the severity hierarchy before and after the implementation of a safety treatment. However, testing for shape difference is not capable of comprehending the difference in distribution among individual severity levels. Statistical testing for shape difference has to be supplemented with a thorough review of the difference in frequency at each severity level. However, there are no developed models to aid in relating the change in relative frequency at each severity level and the underlying level of safety. In order to circumvent this methodological gap, aggregation of microscopic individual severity measurements should be conducted to produce higherlevel measures.

Aggregation of microscopic events is closely tied to representing all traffic events along a dimension that reveals event-level severity variations. The characteristics of traffic events can be described within a significantly high number of dimensions. Individual attributes of traffic events are multitude,
e.g., location, type, time, relative speed, spacing, and severity. Moreover, many of these attributes change over time. Therefore, aggregation will unavoidably entail reduction in dimensionality accompanied with some loss of information. The advocated aggregation strategies in this chapter are each defined by a specific aggregation attribute. The dimension that represents the variability of this attribute will be adopted to represent also the collective variations among all observed traffic events. Collapsing all attribute dimensions is required except along the dimension that represents the aggregation criterion. Subsequently, representative values can be drawn from the distribution of the aggregation attribute among all traffic events.


Figure 7.3 A schematic for different aggregation approaches. Note that datapoints are deinfed over ordered pair of road users to avoid redundant recording of the same traffic event.

Two main aggregation attributes were adopted in this chapter, time and road user. Aggregation over time describes severity of traffic events along the time dimension. All severity measurements are referenced to the moment of analysis. One of the advantages of this aggregation approach is that important temporal patterns can be recognized using this aggregation approach. However, the key advantage of aggregating over time is the simplicity of extrapolating severity measurements outside the time span of observations. A prime example of aggregation over time was adopted in a key study on extreme value model for road collision (Songchitruksa \& Tarko 2006).

The simplicity of adopting time as an aggregation attribute, or alternatively as a surrogate for exposure, comes at the expense of lacking insight into road user interactions. The most direct shortcoming of aggregating over time is the inability to represent the variation in severity measurements among traffic events that take place at the same moment. Another critical shortcoming of aggregation over time is the tendency to under-represent severity if traffic events exhibit irregularity over time. For example, if the average severity per moment (or frame in a video sequence) is selected as a representative value, aggregate severity will be underestimated if the same number of traffic events takes place within shorter time periods. These shortcomings are intrinsic to aggregation over time and can be overcome by adopting different aggregation attribute.

Another aggregation attribute adopted in this study is road users. This aggregation approach provides more insight into road user interactions. For
example, it is possible to represent the true severity of traffic events irrespective of their temporal regularity. Two main shortcomings face the aggregation over road users. The first shortcoming is the relative difficulty of extrapolating severity measurements outside the observational time span as compared to aggregation over time. Road user counts, especially pedestrian counts, are expensive to obtain for extended time periods. The second shortcoming is the inability to represent the presence of the multiple interactions that a single road user may take.

In order to address this last shortcoming, aggregation should be conducted along the event dimension. The same pattern emerges; aggregating over events instead of road users provides a more accurate representation of road user interactions. However, this enhancement comes with significantly more expensive extrapolation of severity measurements outside the observational time span. In fact, the author is not aware of the presence of any temporal conversion factors for the number of traffic events. Aggregation over events was not directly conducted in the case study presented in this chapter because of the significant computational expense; mainly memory requirements to represent the data structure shown in Figure 7.3. Instead, aggregate measurements were normalized by the number of events.

The methodology proposed in previous sections was used for a case study on the conflict data analysis conducted in the Chapter 6. The following sections provide selected results in addition to relevant discussions.

### 7.3 Case Study

This case study is based on video data collected in 2004 for the evaluation of a pedestrian safety treatment in Oakland, California (Bechtel, MacLeod \& Ragland 2004). In this Chapter, a total of six hours of video data were analyzed for the before as well as after periods - three hours for each period. The distributions of conflict indicators were obtained for a subset of all video sequences, a total of four hours, as was presented in Chapter 6 . An additional hour for each observational period was analyzed in this chapter. The following sections will discuss findings of this analysis in light of Hypothesis 7.1 posed in section 7.1.3. Further analysis was conducted to investigate the sensitivity of the results to the selection of the mapping approach and the selection of the aggregation approach. A particularity of aggregation over time in the following results is that the most severe value of each conflict indicator was registered for the entire life span of the traffic event. This was necessary to overcome the computational cost imposed by the size of the traffic event data structure described in Figure 7.3. This approximation suffices for the demonstration of previous methodological developments intended in the analysis of this case study.

### 7.3.1 Empirical Independence of Conflict Indicators

The discussion presented in section 7.1.3 revolved around the hypothesis that conflict indicators provide different and possibly independent severity measurements. The correlation between various conflict indicator measurements was conducted in order to investigate this hypothesis. Only
four conflict indicators were considered in this analysis, namely TTC, PET, GT, and DST.

First, all pairs of conflict indicators that belong to the same traffic event with at least one calculable value were considered. For example, given that one of the pair of conflict indicators reported a calculable value, the pair can be stacked into the matrix used for calculating correlation measures. This was conducted for the joint test of correlation as well as the common calculability of conflict indicators. Table 7.2 shows both the Pearson linear correlation coefficients and Spearman correlation coefficients for different combinations of conflict indicators. The severity interpretation of signed and unsigned values of PET and GT, corresponding to vehicle passage in front of or behind the pedestrian, is not well known. Therefore, the absolute values of PET and GT were also considered in the analysis. Second, testing was conducted for pairs of jointly calculable conflict indicators. For example, pairs of conflict indicators are considered only if both of them report calculable values. This is to explicate the effect of mutual calculability. Similarly, Table 7.3 shows Pearson and Spearman correlation coefficients for different combinations of conflict indicators. Spearman correlation coefficient is slightly more relevant to this context since a linear relationship between the values of conflict indicators may be impacted by the lack of a uniform range definition for conflict indicators, except for the case of pairs of GT and PET.

In general, there is no strong correlation between TTC and any other conflict indicator, except for a 0.67 Spearman correlation with |PET|, when both indicators are mutually calculable. This means that in this video sequence
absolute temporal proximity reflects to some extent the critical presence on a collision course. In addition, there is a strong correlation between PET and GT when both are mutually calculable (0.70 Pearson and 0.87 Spearman correlation coefficients). This is generally expected since the temporal proximity measured by both indicators is to some extent similar. A mild correlation between DST and GT is found for both cases of pairwise calculability. This provides evidence that in this data, the correlation variable presented in Equation 7.1 did not achieve enough stability to provide strong correlation between the two dependent quantities. While correlation results are subject to several interpretations, the general conclusion that can be drawn is in support of Hypothesis 7.1. Note that the correlation results presented in Tables 7.2 and 7.3 are limited to the video data analyzed in this chapter and may be of limited generality absent further empirical evidence.

### 7.3.2 Results of Different Aggregation Approaches

The average values of different conflict indicators were calculated for various mapping approaches and aggregation approaches. In addition, two bounding percentile values, the $15^{\text {th }}$ and $85^{\text {th }}$, were obtained to gauge the dispersion of every conflict indicator. Average values and estimated bounds are provided for index values calculated for each traffic events and using two mapping approaches. For the analysis presented in this section, all function mappings were conducted using parameters inferred from benchmarks presented in the literature presented in Table 7.1. Mappings were also conducted for average values of the four conflict indicators (integration approach B1, Equation 7.2). Elementary error theory was used to obtain the percentile bounds for
individual aggregations. Finally, the difference between the averages of indices calculated from every possible traffic event and the mapping applied to average conflict indicators was obtained. Sample results are presented in Tables 7.4 and 7.5 aggregating over road users. For Tables 7.4 and 7.5, subsequent pairs of tables present results for before and after conditions in respective order.

Table 7.2 Correlation coefficients for pairs of conflict indicators with at least one calculable value

| Conflict <br> Indicator | TTC | PET | DST | GT | $\mid$ PET $\mid$ | $\mid$ GT $\mid$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| TTC | 1 | $-0.06(0.11)$ | $-0.10(0.05)$ | $-0.07(0.04)$ | $-0.43(0.04)$ | $-0.14(0.05)$ |
| PET | $-0.10(0.21)$ | 1 | $0.02(0.12)$ | $0.02(0.05)$ | $0.25(0.20)$ | $-0.04(0.08)$ |
| DST | $-0.15(0.05)$ | $0.04(0.21)$ | 1 | $0.22(0.09)$ | $-0.29(0.03)$ | $-0.09(0.03)$ |
| GT | $-0.28(0.09)$ | $0.09(0.13)$ | $0.59(0.12)$ | 1 | $-0.12(0.02)$ | $0.58(0.28)$ |
| $\mid$ PET $\mid$ | $-0.57(0.03)$ | $0.35(0.36)$ | $-0.49(0.05)$ | $-0.26(0.04)$ | 1 | $-0.26(0.05)$ |
| $\|G T\|$ | $-0.59(0.03)$ | $-0.08(0.26)$ | $-0.01(0.07)$ | $0.50(0.11)$ | $-0.73(0.03)$ | 1 |
| Upper echelon contains pair-wise Pearson linear correlation coefficients. Lower echelon (shaded) <br> contains Spearman $\rho$ rank correlation coefficient. Values in parentheses are the standard deviation of <br> the correlation coefficients calculated for all pairs within a sample of all $1 / 2$ hours of video data (12 <br> samples). |  |  |  |  |  |  |

Table7.3 Correlation coefficients for only pairs of commonly calculable conflict indicators

| Conflict <br> Indicator | TTC | PET | DST | GT | $\mid$ PET $\mid$ | $\mid$ GT $\mid$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| TTC | 1 | $-0.07(0.37)$ | $-0.30(0.09)$ | $-0.09(0.07)$ | $0.42(0.14)$ | $0.14(0.08)$ |
| PET | $0.28(0.32)$ | 1 | $0.49(0.29)$ | $0.70(0.10)$ | $0.25(0.20)$ | $0.06(0.26)$ |
| DST | $-0.57(0.09)$ | $0.46(0.34)$ | 1 | $0.22(0.09)$ | $-0.04(0.13)$ | $-0.08(0.03)$ |
| GT | $-0.10(0.07)$ | $0.87(0.05)$ | $0.59(0.12)$ | 1 | $0.30(0.20)$ | $0.58(0.28)$ |
| $\mid$ PET $\mid$ | $0.67(0.06)$ | $0.35(0.36)$ | $0.01(0.14)$ | $0.56(0.21)$ | 1 | $0.43(0.15)$ |
| $\|G T\|$ | $0.23(0.05)$ | $0.40(0.30)$ | $-0.01(0.07)$ | $0.50(0.11)$ | $0.70(0.07)$ | 1 |
| Up $\mid$ |  |  |  |  |  |  |

Upper echelon contains pair-wise Pearson linear correlation coefficients. Lower echelon (shaded) contains Spearman $\rho$ rank correlation coefficient. Values in parentheses are the standard deviation of the correlation coefficients calculated for all pairs within a sample of all $1 / 2$ hours of video data ( 12 samples).

Average values of every conflict indicator in their respective units are presented in the second columns, entitled Average Indicator, of Tables 7.4 and 7.5. For example, the average of all calculable TTC values for each road user in the before period is shown to be 4.85 sec . The $15^{\text {th }}$ and $85^{\text {th }}$ percentile bounds are provided for each conflict indicator and index in smaller table cells. For example, the $15^{\text {th }}$ percentile value for the distribution of calculable TTC values for all road users in the before period is shown to be $4.85-2.93=1.92 \mathrm{sec}$. The fourth and fifth columns of Tables 7.4 and 7.5 show the function mapping of each average conflict indicator and the percentile bounds, respectively. For example, the function mapping of the average TTC value shown in Table 7.4 can be calculated as: $I(4.85)=e^{-\frac{4.85}{8}}=0.54$. Similarly, the upper bound for the function mapping of the average TTC can be calculated as follows: $I(4.85-2.93)-I(4.85)=e^{-\frac{1.92}{8}}-0.54=0.24$. Using the same steps of calculation, distribution mappings can be conducted by aid of Figure 7.2. Results of distribution mappings of conflict indicators are shown in columns 6 and 7 of Tables 7.4 and 7.5. The average value of individual index values from different conflict indicators is shown in columns 8 and 9 of Tables 7.4 and 7.5. For example, the average function mapping of all individual averages of conflict indicators, entitles Individual Aggregation, can be calculated as follows: Individual function Aggregation $=\frac{0.54+0.29+0.28+0.02+0.50+0.47}{6}=0.35$. The $15^{\text {th }}$ percentile bound can be calculated using elementary error theory as -0.22 .

It is noteworthy that Tables 7.4 and 7.5 present various aggregations without taking into account the frequency of observations of conflict indicators and
indices per road user. Appendix B provides complete set of results for other combinations of aggregation approaches including aggregation over time and considering frequency of observation. In general, there was no noticeable effect of taking into account frequency of observation on the variance of conflict indicators and indices from before and after periods.

Table 7.6 presents results if only positive values for DST (DST + ) are taken into account. This additional analysis was conducted because it is possible that an average measure for a DST may be misleading given that DST assumes positive and negative values with completely different interpretations. The same issues were addressed for PET and GT by conducting independent analysis for positive and negative values for each indicator. This sign segregation for PET and GT is presented in Tables 7.4 and 7.5 as PET+, PET-, GT+, and GT-. Note that all results in Tables 7.4 to 7.6 are accompanied with $15^{\text {th }}$ percentile and $85^{\text {th }}$ percentile bounds minus the average value.

The following observations are noted from the analysis of results of different aggregation approaches:
a. There is a significant dispersion in all conflict indicator and indices values. It is difficult to provide explanation for this observation except that the severity hierarchy was investigated into adequate depth that a wide variation of severity levels was observed.
b. There was no evidence of a measurable difference in average values between before and after conditions. However, the wide variation in conflict indicator and indices values resulted in variant results for statistical tests for difference. Table 7.7 presents two-sample $t$-test for
the statistical significance, at the 0.05 significance, of the difference in mean values of conflict indicators and severity indices between before and after periods. There was a general consensus of a statistically significant reduction in average conflict indicators from before to after for indices aggregated over time. Indices aggregated over road user yielded mixed results. This could be explained by the fact that traffic interactions lasted on average longer for the before period, a characteristic that could not be captured by aggregation over road user.
c. There was no significant difference in results with and without using frequency for calculating average values. This indicates that there was a general balance for the number of conflict indicator observations per frame and per road user.
d. There was marginal difference in the averages of various indices calculated for every traffic event and individual indices mapped from average conflict indicators. The variation in each value was significant, thus casting doubt over the statistical significance of the previous conclusion.
e. Function mapping tends to consistently yield results lower than the distribution mapping. A direct explanation of this observation, as also exhibited in Figures 7.1 and 7.2, is that if compared with a larger pool of observations, the distribution mapping may yield less abnormality values. In other words, the limited reference observations collected in this study created a bias toward overestimating severity if the distribution mapping was used.

Table 7.4 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user

| Conflict <br> Indicator | Average <br> Indicator |  | Individual Index Value |  |  |  | Individual Aggregation |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Function |  | Distribution |  | Function |  | Distribution |  |
| TTC (sec) | 4.85 | -2.93 | 0.54 | -0.14 | 0.35 | -0.20 |  |  |  |  |
|  |  | 2.54 |  | 0.24 |  | 0.44 |  |  |  |  |
| PET+ (sec) | 7.52 | -4.36 | 0.29 | -0.23 | 0.46 | -0.29 |  |  |  |  |
|  |  | 4.28 |  | 0.49 |  | 0.43 |  |  |  |  |
| PET- (sec) | -6.63 | -3.50 | 0.28 | -0.23 | 0.47 | -0.27 |  |  |  |  |
|  |  | 3.47 |  | 0.48 |  | 0.38 |  |  |  |  |
| DST (m/s ${ }^{2}$ ) | 0.29 | -0.52 | 0.02 | -0.02 | 0.37 | -0.37 |  |  |  |  |
|  |  | 0.67 |  | 0.14 |  | 0.39 |  |  |  |  |
| $\mathbf{G T}+(\mathrm{sec})$ | 5.50 | -4.24 | 0.50 | -0.25 | 0.66 | -0.25 |  |  |  |  |
|  |  | 2.50 |  | 0.45 |  | 0.32 |  |  |  |  |
| GT- (sec) | -5.16 | -2.77 | 0.47 | -0.32 | 0.63 | -0.28 |  |  |  |  |
|  |  | 4.03 |  | 0.49 |  | 0.35 |  |  |  |  |
| \|PET| (sec) | 7.06 | -3.90 | 0.33 | -0.25 | 0.50 | -0.30 | - |  | - |  |
|  |  | 3.99 |  | 0.45 |  | 0.39 |  |  |  |  |
| \|GT| (sec) | 5.35 | -4.16 | 0.52 | -0.26 | 0.68 | -0.26 | - |  | - |  |
|  |  | 2.62 |  | 0.44 |  | 0.30 |  |  |  |  |
| Index <br> (function) | 0.34 | -0.23 | - |  | - |  | -0.0 |  | - |  |
|  |  | 0.20 |  |  | -0.32 | 0.46 |  |  |  |  |
| Index <br> (distribution) | 0.51 | -0.25 | - |  |  |  | - |  | - |  | 0.01 |  |
|  |  | 0.23 |  |  | -0.38 | 0.45 |  |  |  |  |

Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

Table 7.5 Summary results for different aggregation strategies for after conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user

| Conflict Indicator | Average <br> Indicator |  | Individual Index Value |  |  |  | Individual Aggregation |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Function |  | Distribution |  | Function |  | Distribution |  |
| TTC (sec) | 4.14 | -2.99 | 0.59 | -0.14 | 0.44 | -0.23 |  |  | $\begin{aligned} & \underset{\infty}{n} \\ & \vdots \\ & \ddot{\sim} \\ & \tilde{n} \\ & \underset{i}{2} \\ & \underset{n}{n} \\ & 0 \end{aligned}$ |  |
|  |  | 2.30 |  | 0.27 |  | 0.47 |  |  |  |  |
| PET+ (sec) | 7.54 | -4.17 | 0.28 | -0.21 | 0.45 | -0.27 |  |  |  |  |
|  |  | 3.77 |  | 0.47 |  | 0.42 |  |  |  |  |
| PET- (sec) | -5.81 | -3.61 | 0.38 | -0.31 | 0.56 | -0.32 |  |  |  |  |
|  |  | 4.11 |  | 0.54 |  | 0.40 |  |  |  |  |
| DST (m/s ${ }^{2}$ ) | 0.37 | -0.44 | 0.04 | -0.04 | 0.44 | -0.44 |  |  |  |  |
|  |  | 0.54 |  | 0.12 |  | 0.30 |  |  |  |  |
| $\mathbf{G T}+(\mathrm{sec})$ | 5.34 | -4.19 | 0.52 | -0.32 | 0.68 | -0.33 |  |  |  |  |
|  |  | 3.43 |  | 0.44 |  | 0.30 |  |  |  |  |
| GT- (sec) | -5.39 | -4.14 | 0.44 | -0.37 | 0.61 | -0.38 |  |  |  |  |
|  |  | 4.63 |  | 0.54 |  | 0.38 |  |  |  |  |
| \|PET| (sec) | 7.09 | -4.33 | 0.33 | -0.24 | 0.50 | -0.29 | - |  | - |  |
|  |  | 3.83 |  | 0.50 |  | 0.42 |  |  |  |  |
| \|GT| (sec) | 5.36 | -4.38 | 0.52 | -0.34 | 0.68 | -0.35 | - |  | - |  |
|  |  | 3.69 |  | 0.45 |  | 0.31 |  |  |  |  |
| Index (function) | 0.36 | -0.26 | - |  | - |  | -0.0 |  | - |  |
|  |  | 0.24 |  |  | -0.37 | 0.49 |  |  |  |  |
| Index <br> (distribution) | 0.51 | -0.27 | - |  |  |  | - |  | - |  | -0.01 |  |
|  |  | 0.26 |  |  | -0.43 | 0.47 |  |  |  |  |

Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

Table 7．6 Summary results for different aggregation strategies for all conditions．Representative statistics are drawn only from calculable values of positive DST values

| Agg． <br> Type | Freq． | Time | DST＋（sec） |  | Individual Aggregation |  | Difference individual agg． and index value |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  | Function | Distribution | Function | Distribution |
| $\underset{B}{\ddot{E}}$ |  | 这 | 0.60 | －0．44 | 0.35 | 0.52 | 0.00 | 0.01 |
|  |  |  |  | 0.44 | （－0．24：0．41） | （－0．29：0．37） | （－0．30：0．44） | （－0．34：0．41） |
|  |  | $\frac{\ddagger}{\leftarrow}$ | 0.64 | －0．43 | 0.32 | 0.50 | 0.01 | 0.02 |
|  |  |  |  | 0.42 | （－0．22：0．50） | （－0．30：0．44） | （－0．33：0．55） | （－0．37：0．49） |
|  | $\begin{aligned} & \text { \% } \\ & \text { in } \\ & 0 \end{aligned}$ | B | 0.67 | －0．42 | 0.38 | 0.55 | －0．02 | －0．01 |
|  |  |  |  | 0.44 | （－0．22：0．36） | $(-0.26: 0.32)$ | （－0．27 ：0．38） | （－0．31：0．35） |
|  |  | $\stackrel{\ddagger}{\overleftarrow{\pi}}$ | 0.67 | －0．40 | 0.41 | 0.59 | －0．04 | －0．04 |
|  |  |  |  | 0.43 | （－0．23：0．37） | $(-0.26: 0.32)$ | （－0．29：0．40） | （－0．31：0．36） |
| $\begin{gathered} \stackrel{\pi}{x} \\ \stackrel{y}{4} \\ \text { an } \end{gathered}$ |  | 这 | 0.59 | －0．49 | 0.36 | 0.53 | －0．01 | －0．02 |
|  |  |  |  | 0.51 | （－0．22 ：0．41） | （－0．30：0．36） | （－0．32：0．46） | （－0．39：0．43） |
|  |  | $\stackrel{\ddagger}{\text { む }}$ | 0.55 | －0．45 | 0.38 | 0.55 | －0．02 | －0．04 |
|  |  |  |  | 0.48 | （－0．26：0．42） | （－0．33：0．37） | （－0．37 ：0．49） | （－0．42 ：0．46） |
|  |  | 这 | 0.71 | －0．53 | 0.36 | 0.53 | －0．02 | －0．02 |
|  |  |  |  | 0.53 | （－0．21：0．39） | （－0．28：0．35） | （－0．29：0．44） | （－0．36 ：0．41） |
|  |  | $\frac{\ddagger}{\ddagger}$ | 0.64 | －0．48 | 0.37 | 0.55 | －0．03 | －0．04 |
|  |  |  |  | 0.49 | （－0．25 ：0．42） | （－0．31：0．37） | （－0．33：0．47） | （－0．39：0．43） |

Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean．

Table 7.7 Summary results for the two-sample $t$-test for the difference in mean between conflict indicators and indices for before and after time conditions. " 1 " means the test for before>after was significant at the 0.05 significance level and conversely after>before for "-1". "0" means no significant difference was found.

| Agg. <br> Type | Freq. | Indicator or Index |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | E | $\frac{+}{2}$ | $5$ | $\stackrel{+}{\circlearrowleft}$ | Function |  | Distribution |  |
|  |  |  |  |  |  | Index (max) | Index <br> (avg.) | Index (max) | Index <br> (avg.) |
| $\underset{B}{\ddot{E}}$ | Without Freq. | 1 | 1 | -1 | -1 | 1 | 1 | 1 | 1 |
|  | With <br> Freq. | 1 | 1 | -1 | -1 | 1 | 1 | 1 | 1 |
|  | Without Freq. | 1 | 0 | 1 | 0 | -1 | -1 | 0 | 0 |
|  | With <br> Freq. | 1 | 0 | 1 | 0 | -1 | -1 | 0 | 0 |

The message in Table 7.7 relating the difference of severity indices between the before and after periods appears at first to be counterintuitive. Aggregation over time and over road user provide opposite inferences on the difference between average severity index between before and after periods. An explanation is that pedestrians remain exposed to collision risk for longer times in the before period. This is a plausible explanation since pedestrian scramble provided more isolation of pedestrians compared to the traditional reliance on yield-to-pedestrian regulations.

### 7.3.3 Accounting for both Severity and Frequency

Aggregation results presented in section 7.3.1 mainly concern the average severity of all exposure traffic events. However, change in average severity between before and after periods cannot represent the change in exposure
between the same periods. For example, Figure 7.4 shows the distributions of the severity index mapped using the function mapping B1. The distributions exhibit a clear reduction in frequency of observation of traffic events at almost all severity levels. This safety improvement was not evident in Tables 7.4 and 7.5 mainly because averaging conflict indicators and indices measurements implicitly discards the effect of variant exposure. The precise definition of exposure event in this analysis includes any pair of pedestrian and vehicular road users that attain at minimum spacing closer than a spatial proximity threshold and also exhibit at some time convergent movement directions. The spatial proximity threshold was selected to be 10.0 m .

Figures 7.6 and 7.7 further demonstrate the distinct safety information obtained when normalizing various severity measurements by the number of exposure events. In Figures 7.6 and 7.7, the distributions of various conflict indicators are shown after normalizing their frequencies by the total number of exposure events. The magnitude and sign of the difference in distributions between before and after periods is mixed. Some indicators, such as $|G T|$ exhibit stable severity for every instance of road user exposure in before and after conditions. PET exhibits different trends for positive and negative values, with positive PET exhibiting increase in severity after the treatment. Other indicators such as DST and TTC exhibit an increase in severity per instance of road user exposure after the safety treatment. The distinct information contained in severity measures normalized by the number of exposure events can be misinterpreted as all-encompassing safety cue. A more comprehensive severity index can be constructed by including the following aspects:

1. Severity of each exposure event.
2. Observed number of exposure events, and
3. Maximum number of possible exposure events,

A simple mechanism to combine the first and second aspects is the summation of all severity indices measurements, $\sum_{\text {event }}$ severity indices. In order to further incorporate the third aspect, the previous summation can be divided by the number of maximum possible exposure $E_{\max }$. This is to account for the safety differential between situations where the same summation of severities originates from different levels of traffic volume. This normalized safety measure $\bar{I}$ can be constructed as follows:

$$
\begin{equation*}
\bar{I}=\frac{\sum_{\text {event }} \text { severity indices }}{E_{\max }} \tag{7.5}
\end{equation*}
$$

Theoretically, the maximum number of possible exposure events is the product of two conflicting traffic streams. In the context of pedestrian safety, $\mathrm{E}_{\text {max }}$ is the product of the number of pedestrians and the number of vehicles present during the observational period. Another plausible surrogate for $E_{\text {max }}$ is the total pedestrian and vehicle volumes during the observational period. Results of $\bar{I}$ calculation using different $\mathrm{E}_{\text {max }}$ estimates and different aggregation approaches are presented in Tables 7.8-7.11.

It is important to note the difference between the use of $\mathrm{E}_{\text {max }}$ proposed in Equation 7.5 and its use as a surrogate for the total number of exposure events (refer to section 7.1.1 and (Greene-Roesel, Diógenes \& Ragland 2007) (Keall 1995)). The construction of the normalized safety measure presented in Equation 7.5 sets clear boundary between the estimation of maximum
possible exposure and the accurate observation of exposure represented by the number of exposure events. Putting the two quantities in perspective, or dividing them as is shown in Equation 7.5, represents the distinct safety benefit of reducing actual exposure.

Table 7.8 Summary results for before and after index values normalized by the total number of tracked road users. Indices representing an event are the maximum of all mapped conflict indicators

| Agg. <br> Type | Freq. | Distribution |  | Function |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Before | After | Before | After |
| $\underset{B}{B}$ | Without Freq. | 2.69 | 1.15 | 3.37 | 1.54 |
|  | With <br> Freq. | 47.41 | 19.95 | 56.89 | 24.05 |
|  | Without Freq. | 0.10 | 0.06 | 0.13 | 0.08 |
|  | With <br> Freq. | 0.41 | 0.16 | 0.54 | 0.22 |

Table 7.9 Summary results for before and after index values normalized by the total number of tracked road users. Indices representing an event are the average of all mapped conflict indicators

| Agg. <br> Type | Freq. | Function |  | Distribution |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Before | After | Before | After |
| $\underset{B}{\underline{E}}$ | Without Freq. | 1.58 | 0.72 | 2.34 | 1.12 |
|  | With <br> Freq. | 23.59 | 10.29 | 35.47 | 15.25 |
|  | Without Freq. | 0.07 | 0.05 | 0.11 | 0.07 |
|  | With <br> Freq. | 0.25 | 0.11 | 0.39 | 0.17 |



Figure 7.4 Severity index distributions for before and after conditions. Function mapping was used. Maximum indices were selected for every frame (upper row) and road user (bottom row).


Figure 7.5 Severity index distributions for before and after conditions. Function mapping was used. Average indices were selected for every frame (upper row) and road user (bottom row).


Figure 7.6 Conflict indicator and index distributions for before and after conditions. Maximum indicators and indices were selected for every road user. Distributions are normalized by the total number of exposure events


Figure 7.7 Conflict indicator and index distributions for before and after conditions. Average indicators and indices were selected for every road user. Distributions are normalized by the total number of exposure events

Table 7.10 Summary results for before and after index values normalized by the product of the volumes of pedestrians and vehicles in millions. Indices for every event are the maxima of all mapped conflict indicators

| Agg. <br> Type | Freq. | Distribution |  | Function |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Before | After | Before | After |
| 辟 | Without Freq. | 191 | 99.4 | 239 | 132 |
|  | With <br> Freq. | 3360 | 1710 | 4000 | 2100 |
|  | Without Freq. | 7.30 | 5.25 | 9.75 | 7.11 |
|  | With <br> Freq. | 29.7 | 14.2 | 38.7 | 19.2 |

Table 7.11 Summary results for before and after index values normalized by the product of the volumes of pedestrians and vehicles in millions. Indices for every event are the averages of all mapped conflict indicators

| Agg. <br> Type | Freq. | Function |  | Distribution |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Before | After | Before | After |
| \#̈n | Without Freq. | 111 | 62.5 | 165 | 96.7 |
|  | With <br> Freq. | 1670 | 883 | 2510 | 1300 |
| $\begin{gathered} \stackrel{\pi}{x} \\ \stackrel{y}{4} \\ \underset{\sim}{2} \end{gathered}$ | Without Freq. | 5.33 | 4.32 | 7.85 | 6.16 |
|  | With <br> Freq. | 18.2 | 9.82 | 27.8 | 14.7 |

### 7.4 Conclusions

This chapter presented a series of arguments that develop into the hypothesis that conflict indicators represent partially overlapping severity aspects. A number of approaches have been proposed in order to map into the severity dimension and to integrate conflict indicators into a severity index. In addition, aggregation of conflict indicator and severity index measurements was advocated. A number of aggregation approaches have been proposed. The chapter ends with an important proposition of an aggregate safety measure that reflects the underlying level of road safety.

An important motivation for this work has been the lack of a conflict indicator that can comprehensively represent severity of traffic events. Admittedly, such development is not foreseen in the near future. This prognosis is based on the expected model complexity for representing all uncertainties that concern the risk collision. In addition, calibration of such model will require proportionately large volume of data, mainly road user tracks in normal and conflicting situations. The proposed mapping and integration approaches are interim methodological developments until the development of a comprehensive conflict indicator.

Conflict indicators provide microscopic severity measures. However, the ultimate purpose of road safety analysis is to draw an inference on the underlying level of safety. A proposition of aggregating conflict indicator measurements was introduced in this chapter. For this purpose, three approaches were developed: aggregations over time, over road users, and over exposure events. The order of the three approaches reflects the precision
of exposure measurement. However, the data required for projecting such aggregation measures outside the period of observation is proportional to their precision. With progress in road user tracking technologies, surrogates of exposure will be gradually abandoned in favour of more precise measures of exposure.

Part of the analysis presented in this chapter dealt with average conflict indicators and severity indices per road user or time frame. Several questions remain unanswered regarding the general usefulness of average severity measures. This inquiry can be even generalized for average collision severity measured in terms of the total cost of various types of road collisions normalized by the total number of road collisions. In fact, it can be argued that average severity measures provide a distinct message that can likely be misinterpreted.

A broad objective of safety treatments is the reduction in the incidence of road collision. In the special case when only road collisions are concerned, it can be argued that a reduction in average severity measures is conceptually independent of a concomitant change in the frequency of road collisions. To explain the previous argument, consider the case of two intersections for which all types of road collisions are observed, albeit with different frequency. Furthermore, assume that for these intersections the total cost of road collisions is identical. From a social cost standpoint, the two intersections are equally dangerous and therefore should be placed at the same priority level when being considered for safety improvement. The fallacy of relying on average severity is apparent when investigating the
sensitivity of prioritization for treatment to the relative frequency of different types of road collisions. It is conceivable that for each intersection an increase in the frequency of property-damage-only (PDO) collisions will result in the reduction of average severity. The same reasoning can be applied to traffic conflicts the severity of which can be measured using the methodology proposed in this chapter. A relative increase in low-severity events may reduce average severity. There is a distinct safety message contained in the measurement of average severity. On conceptual grounds, this message is not related to the effectiveness of road safety treatments and to the prioritization for safety improvement programs.

It can be argued that the message in average severity is orthogonal to quantitative inferences on road safety. Rather, average severity measures reflect the proximity and consequences per aggregation unit, e.g., time, road user, or exposure event. It reflects the severity to which road users are likely to be exposed given that there is a chance-setup or genuine exposure for such severity to materialize. It can be argued that a reduction in total exposure as well as total severity, represented by the summation of severity indices for traffic conflicts or the total cost for road collisions, accompanied by an increase in average severity involves a positive safety improvement. Furthermore, it can be argued that such safety improvement is larger than the case when average severity is retained or even reduced after the introduction of a safety treatment!

To explain the previous argument, consider the case that a combination of road user behaviour and effect of safety treatment was characterised by the
following: 1) reduction in total exposure, 2) reduction in total severity, and 3) increase in severity per exposure event. It can be inferred that this particular treatment was not only successful in reducing the incidence of severe traffic events, but also was capable to some extent of disguising this safety benefit from road users. Therefore, if it is the case that the users of this intersection are penchant to compensate for safety improvement by accepting more risk, they are less likely to engage this behaviour at an intersection with high average severity than an intersection with low average severity. We can summarize these arguments in the following hypothesis:

> Hypothesis 7.2: Total severity of traffic events and average severity per exposure events are two orthogonal dimensions for measuring road safety. The first dimension represents the magnitude of safety from a social perspective. The second dimension represents safety perceived by the average road user of the concerned intersection.

An important distinction was made in this chapter between maximum possible exposure and actual exposure. The two quantities reflect the effectiveness of a safety treatment in limiting road user exposure to collision risk. The two quantities were augmented with the summation of all severity indices obtained from each traffic event (total severity) to produce a novel safety measure. The proposed safety measure is based on normalizing the summation of all severity indices by the maximum possible exposure. There is a well-recognized shortcoming of the naïve division of total severity by exposure. It may be the case that, similar to collision frequency, total severity independent of the underlying safety level is non-linearly related to
maximum exposure. In this case, for reasons extraneous to safety, the mere increase in traffic volume would unreasonably lead to reduction in the safety measure. This non-linearity should to be further investigated. If such nonlinearity is proven, the divisional form proposed in Equation 7.5 should be modified to reflect intrinsic relationships between exposure and total severity.

## Automated Detection of Traffic Violation Events

### 8.1 Background

The incidence of traffic violations is an important indicator of road safety. Fundamentally, traffic violations occur when road users seek an increased mobility at the expense of accepting additional collision risk due to nonconforming to traffic regulations. For example, red-light violations occur when road users accept higher risk of crossing or left-turn collisions, in favour of more perceived utility achieved by reducing travel time. Traffic violations can also be viewed as precursors to traffic conflicts, in as much as traffic conflicts are conceptually viewed as indicators of collision. Recent work
has conjectured that traffic violations can be placed in the same severity hierarchy, albeit at a lower level, along with traffic conflicts and collisions (Oh et al. 2010). Accordingly, traffic violations may be viewed as a set of traffic events which comprises both traffic conflicts and collisions. However, this argument should be qualified to traffic configurations that represent sound design practice. Under these configurations, road users conforming to traffic regulations are not expected to be involved in traffic conflicts or road collisions with other road users. The relevance of traffic violations to road safety can be more evident when the prevalent chain of events that leads to collision contains an action of traffic violation. For example, all traffic conflicts observed in the study presented in Chapter 5 were caused by motorists committing illegal left-turn maneuvers. In the previous study, it could be reasonably argued that left-turn violations are plausible surrogates for traffic conflicts.

The practical benefit of observing violations as surrogates to traffic conflicts, and consequently road collisions, is especially realized when observational periods are limited. While generally more frequent than road collisions, traffic conflicts are still less frequent than traffic violations. It is possible for observational periods to be too limited to record a representative sample of traffic conflicts. In situations where it is likely that road collisions are attributable to violation actions, traffic violations can provide a reliable surrogate road safety measure. Several studies argued on conceptual and empirical grounds that traffic violations are valid indicators of road safety, e.g., (Struckman-Johnson et al. 1989) (Elliott, Baughan \& Sexton 2007) (Ayuso, Guillén \& Alcañz 2010). Furthermore, arguments raised in previous chapters
in support of automated road safety analysis find similar ground in the detection of traffic violations.

Automated analysis enables the processing of extended observational periods while consuming limited time and staff resources. A distinct practical benefit of the automated detection of traffic violations can materialize in the case of real-time video analysis. Moreover, automated detection of traffic violations can support traffic monitoring for the purpose of identifying operational or traffic control issues. Finally, methodological developments can be directly adopted in the context of security monitoring and surveillance. The previous contexts are not addressed in this chapter, however they constitute important continuation of the work presented herein. The main focus of this chapter is on the automated detection of vehicular violations in urban settings. The video sequence used in this chapter did not include enough pedestrian volume to warrant a focused study on pedestrian violations. Conversely, there was a remarkable incidence of traffic violations committed by vehicles. The methodology developed for automated detection of vehicle violations can readily be adopted for detecting pedestrian violations. Subsequent sections present two approaches for violation detections.

The first approach presents an adaptation of a traditional clustering approach, the k-means clustering algorithm, for the purpose of automated violation detection. Inherent shortcomings in the k-means algorithm motivated the reliance on more insightful discriminative features of violation movements obtained using the Longest Common Subsequence (LCSS) similarity measure. The second approach in this chapter is based on violation detection by means
of measuring track similarity to patterns learnt for normal movements using the LCSS similarity measure. As shorthand, the second approach is referred to as LCSS-based violation detection. Compliant movements are represented by normal movement patterns or prototypes. Movement patterns are a subset of road user tracks ${ }^{1}$ that possess adequate similarity to all road user tracks and lack similarity amongst each other.

A case study is presented in subsequent sections on the automated violation detection of vehicular movement. The video data analyzed in the case study was collected for approximately 2 hours at an urban intersection in Kuwait City, Kuwait. Both approaches presented relative strength, however the reliance on LCSS matching proved an overall superiority. The chapter ends with general discussion and conclusions.

### 8.2 Methodology

The following section presents an adaptation of $k$-means clustering for the purpose of automated violation detection. Clustering features were mainly based on directional movements of road users while navigating an intersection. In order to enhance the robustness to tracking errors, piecewise linear parameterization of road user tracks was performed to extract clustering features. A subsequent section presents an adaptation of a classification technique based on the LCSS matching.

[^17]
### 8.2.1 K-means Clustering using Linear Piecewise Parameterization

 Clustering is the process of organizing points in a dataset into a number of subsets, each subset contains elements of similar characteristics. Formally, the clustering problem can be represented as the mapping of all elements in a dataset $\mathrm{T}=\left\{T_{i} \mid i=1,2, \ldots n\right\}$ onto the set of clusters $\mathrm{C}=\left\{c_{j} \mid j=1,2, \ldots, m\right\}$ such that each element $T_{i} \in \mathrm{~T}$ is uniquely assigned to a cluster $c_{j} \in \mathrm{C}$. The problem of classifying road user tracks into normal and violation movements can be directly cast as a clustering problem, with $T_{i}$ representing road user tracks and $c_{j}$ representing different movement prototypes. An informed clustering technique can theoretically be constructed to group normal and violation road user tracks into separate sets.K-means clustering is one of the most widely used algorithms for clustering analysis (Lloyd 1982). A pre-requisite to k -means clustering is the reduction in dimensionality from the original data space to a lower dimension feature space. For example, a road user track can be represented by a number of features; each describing the type of manoeuvre performed by the road user generating this track. The idea behind k -means clustering is to first select, typically at random, centroids of each cluster in the feature space. Then several iterations are performed until every data point is assigned to the cluster with the closest distance measured in the feature space.

The main requirement for effective clustering is the informed selection of features that could aid in discriminating between normal and violation tracks. A typical challenge to indirect clustering techniques is the immense dimensionality of road user tracks. In order to comprehensively represent a
road user track, the number of feature space dimensions should equal three times the number of observed positions along this track $n_{p}$. Twice the number of observed positions is required for two-dimensional representation of all positions. The third multiplication is required to represent time. The dimensionality of the feature space can be reduced to $2 n_{p}+1$ if road user positions are observed every fixed time interval. A full-dimensional representation of road user tracks for clustering analysis is computationally intensive and is susceptible to tracking noise. Furthermore, road users typically exhibit some movement patterns while navigating traffic intersections or road segments. More informed clustering can be conducted if features are selected based on an assumed movement model for road user movements.

For example, in the hypothetical situation where only through movements are allowed, the prevalent movement direction is likely a discriminative feature that entails a drastic reduction in dimensionality from $3 n_{p}$ to 1 . For turning movements, more features can be used to represent throughmovement segments and turning-movement segments of the road user track. Therefore, the prevalent direction of each track segment can be used as a discriminative feature. In subsequent analysis using k-means clustering, road user directions are adopted as the main type of feature. In order to mitigate the effect of tracking noise and improve the representation of road user movements, the profile in time of road user directions is approximated using a piecewise linear model.

A typical road user moves from an origin to a specific destination in traffic scenes. Therefore, their movement can reasonably be approximated into a finite number of maneuvers. The definition of a maneuver in this context is a sequence of positional changes that can be effectively represented by a regular geometric model. Furthermore, feature definition in terms of temporal profile of movement direction enables efficient and accurate representation of all curvilinear tracks in linear form. Figure 8.1 shows a sample vehicle track approximated by three different sequences of linear segments. Piecewise linear models based on four linear segments were used in subsequent analysis since they proved, after several preliminary trials, to provide an effective representation of road user tracks. Figure 8.2 shows sample tracks approximated by piecewise linear models using four linear segments. The features that represent a track, also called clustering parameters or variables, are the slopes of the line segments which constitute the piecewise linear model.

In order to further simplify the definition of a feature, all tracks were represented along the same horizontal axis by normalizing the measurement time of each position to the total life span of each track. For example, the first point of each track occurs at moment 0 and the last point occurs at moment 1. These features drawn from each road user track are afterwards relayed to k mean clustering. Previous techniques for feature extraction and $k$-means clustering were implemented in the MATLAB language with the aid of the Cluster Analysis Toolbox (Mathworks 2010).


Figure 8.1 A vehicular track approximated by three different piecewise linear models. The vertical axis shows the cosine of the instantaneous azimuth of road users and the horizontal axis shows the moment of measurement relative to the total life span of the track.

Subsequent to k -means clustering analysis, a heuristic rule is applied to interpret the relative size of each cluster as to whether it contains violation or normal tracks. The number of tracks in each cluster is compared to a threshold R. If the number of tracks within a cluster relative to the total number of tracks exceeds this threshold, the cluster is considered to contain normal tracks. If otherwise the relative frequency of tracks within a cluster is less than R, then all tracks comprised by this cluster are considered to be violation tracks. Assuming that the number of tracks contained in some
cluster $c_{j}$ is $N\left(c_{j}\right)$, then violation tracks are formally classified according to the following criterion:

$$
\forall \text { voilation tracks } T_{i v} \in \mathrm{~T}, \frac{\mathrm{~N}\left(\mathrm{c}_{\mathrm{j}}\right),}{\sum_{\mathrm{j}=1,2, \ldots, \mathrm{~m}, \mathrm{~N}\left(\mathrm{c}_{\mathrm{j}}\right)}}<R
$$

$$
\begin{equation*}
\forall \text { normal tracks } T_{i v} \in \mathrm{~T}, \frac{\mathrm{~N}\left(\mathrm{c}_{\mathrm{j}}\right),}{\sum_{\mathrm{j}=1,2, \ldots, \mathrm{~m}, \mathrm{~N}}\left(\mathrm{c}_{\mathrm{j}}\right)} \geq \mathrm{R} \tag{8.1}
\end{equation*}
$$



Figure 8.2 Sample tracks approximated to a fourth-degree piecewise linear model (a sequence of four linear segments). Vertical axes show the cosine of the instantaneous azimuth of road users and horizontal axes show the moment of measurement relative to the total life span of each track.

Various performance measures can be calculated given the knowledge of the true classification of each road user track. Three performance measures were used in this study: percentage correct classification (PCC), Kappa categorical similarity coefficient, and Entropy. Kappa coefficient is directly proportional to the quality of classification, i.e., agreement between predicted classification and true classification. Values of Kappa coefficient greater than 0.61 generally represent substantial agreement. Entropy is negatively proportional to the classification performance. No benchmarks were found for judging the quality of performance based on Entropy values. Therefore it is best suited for comparison between two classification approaches. One of the universal measures of system randomness or uncertainty is Entropy $H$. In clustering analysis, Entropy has the advantage of representing both the success of the clustering approach in isolating violation tracks within a few number clusters while taking into account the total number of clusters required for successful classification. An adaptation of Shannon's Entropy (Shannon 1948) (Li, Zhang \& Jiang 2004) was used to measure the randomness in clustered sets that also represents the classification accuracy:

$$
\begin{gather*}
\mathrm{H}\left(\mathrm{c}_{\mathrm{j}} \mid T_{i}\right)=-\operatorname{prob}\left(\mathrm{c}_{\mathrm{j}} \mid T_{i}\right) * \log \left(\operatorname{prob}\left(\mathrm{c}_{\mathrm{j}} \mid T_{i}\right)\right) \\
\mathrm{H}\left(\mathrm{C} \mid T_{i}\right)=\sum_{\mathrm{j}=1,2, \ldots, \mathrm{~m}} \mathrm{H}\left(\mathrm{c}_{\mathrm{j}} \mid T_{i}\right) \\
\mathrm{H}(\mathrm{C} \mid \mathrm{T})=\sum_{\mathrm{i}=1,2, \ldots, \mathrm{n}} \mathrm{H}\left(\mathrm{C} \mid T_{i}\right) * \operatorname{prob}\left(T_{i}\right) \tag{8.2}
\end{gather*}
$$

where $\operatorname{prob}($.$) is the probability function. The term \operatorname{prob}\left(T_{i}\right)$ is estimated as the relative frequency of the true type of the track in reference to the total number of tracks. For example, if $T_{i v}$ is for fact a violation track, then:

$$
\begin{equation*}
\operatorname{prob}\left(T_{i v}\right)=\frac{\text { number of violation tracks }}{\text { total number of tracks }} \tag{8.3}
\end{equation*}
$$

### 8.2.2 Violation Detection using LCSS matching

The first step of LCSS-based violation detection is to create a set of movement prototypes that represent what are considered as normal movement prototypes. Subsequently, a comparison is conducted between a given track and normal movement prototypes. Any significant disagreement between both sequences of positions is interpreted as an evidence that the given track represents the movement of a road user performing a traffic violation. More specifically, this comparison relies on an LCSS similarity measure between the movement prototypes and the trajectories to make decision about the classification.

The LCSS similarity measure is defined in a non-metric space. This property enables the successful adoption of LCSS similarity measure in several applications. The LCSS problem was originally defined for matching timeseries measurements, e.g., sequences of communication signals. The application of LCSS similarity measure for matching trajectory data has been successfully demonstrated (Vlachos, Kollios \& Gunopulos 2005) (Saunier, Sayed \& Lim 2007). LCSS similarity gives more weight to similar segments of road user tracks while allowing some parts to be unmatched. This matching strategy proves remarkable robustness to tracking noise as well as incomplete tracking of road users. Furthermore, the matching of two positions can be bound by a variety of $L_{p}$ Norms; not necessarily the computationally expensive $L_{2}$ Norm. In the following analysis, $L_{\infty}$ is used to measure the proximity of two matched positions. The following section provides a
formalized definition of LCSS similarity and describes an adaptation for the purpose of violation detection.

Let $N($.$) be a measure on a finite set that returns its number of elements. Let \tau$ be a finite set of road user tracks $\left\{T_{i} \mid i \in 1,2, \ldots, N(\tau)\right\}$. Let each road user track $T_{i}$ be composed of a set of coordinate tuples such that $T_{i}=\left\{t_{i k} \mid k \in\right.$ $\left.1,2, \ldots, N\left(T_{i}\right)\right\}$ and each coordinate tuple $t_{i k}$ be defined as $t_{i k}=\left(x_{i k}, y_{i k}\right)$. Two points $t_{i k}$ and $t_{j l}$ to be matched if $\max \left\{\left|x_{i k}-x_{j l}\right|,\left|y_{i k}-y_{j l}\right|\right\}<\epsilon$, where $\epsilon$ is some spatial proximity bound called hereafter matching distance. The LCSS of two road user tracks $T_{i}$ and $T_{j}, \operatorname{LCSS}_{\epsilon}\left(T_{i}, T_{j}\right)$ of respective lengths $m$ and $n$, is defined recursively as follows:

- $\quad 0$ if $m=0$ or $n=0$,
- $1+\operatorname{LCSS}_{\epsilon}\left(\operatorname{Head}\left(T_{i}\right), \operatorname{Head}\left(T_{j}\right)\right)$ if the points $t_{i n}$ and $t_{j m}$ match,
- $\max \left\{\operatorname{LCSS}_{\epsilon}\left(\operatorname{Head}\left(T_{i}\right), T_{j}\right\}, \operatorname{LCSS}_{\epsilon}\left(T_{i}, \operatorname{Head}\left(T_{j}\right)\right)\right.$, otherwise.
where $\operatorname{Head}\left(T_{i}\right)=\left\{t_{i k} \mid k \in 1,2, \ldots, N\left(T_{i}\right)-1\right\}$ and the definition is identical for all tracks other than $i$.

The LCSS of two road user tracks is further normalized in order to produce a non-metric similarity measure. The incremental prototype learning algorithm used in this analysis, (Saunier, Sayed \& Lim 2007), adopted normalizing LCSS by the minimum length of the two matched tracks. This normalization strategy is invariant to the difference in length between the matched tracks. This strategy is effective during the process of prototype learning since it tends to yield a parsimonious representation of road user tracks.

In the process of violation detection, the matched road user tracks have different interpretation than in the case of prototype learning. Without loss of generality, it can be assumed that the function $\operatorname{LCSS}_{\epsilon}\left(T_{i}, T_{j}\right)$ defines in order the LCSS between a road user track $T_{i}$ and a previously learnt prototype $T_{j}$. Prior to similarity matching, it is not known whether a road user track $T_{i}$ represents a normal or a violation movement. In the process of violation detection it is not plausible to treat equally the case when $\operatorname{LCSS}_{\epsilon}\left(T_{i}, T_{j}\right)<T_{j}$ and the opposite case. The first case likely involves a partial road user track while the opposite can be interpreted in different ways. The case $\operatorname{LCSS}_{\epsilon}\left(T_{i}, T_{j}\right)<T_{i}$ could occur if $T_{j}$ is a partial road user track that was included in the set of prototypes. This case can also occur if $T_{i}$ is in fact a violation track that contains subsequences which were not matched to any prototype. In order to explicate the two cases, a different normalization strategy was used for violation detection. The non-metric LCSS similarity measure DLCSS (more precisely a dissimilarity measure) used in violation detection is defined as follows:

$$
\begin{equation*}
\operatorname{DLCSS}\left(T_{i}, T_{j}\right)=1-\frac{\operatorname{LCSS}\left(T_{i}, T_{j}\right)}{N\left(T_{i}\right)} \tag{8.4}
\end{equation*}
$$

Therefore, the LCSS is normalized by the length of the sub-sequence of the tacked object. LCSS-based violation detection is conducted on all road user tracks $\left\{T_{i} \in \tau\right\}$ by matching against the set of normal prototypes $\left\{T_{j} \in \tau_{\text {normal }}\right\}$. The latter set can be created by incrementally learning prototypes for a period of time and then manually removing prototypes that represent road user violations. For a given similarity threshold $\sigma \in[0,1]$, a road user track $T_{i}$ is identified as a violation track if the following condition is met:

$$
\begin{equation*}
\min \left\{\operatorname{DLCSS}\left(T_{i}, T_{j}\right) \mid j=1,2, \ldots N\left(\tau_{\text {normal }}\right)\right\}>\sigma \tag{8.5}
\end{equation*}
$$

If the condition in Equation 8.5 is not met, then a road user track $T_{i}$ is considered to represent a normal road user movement. Furthermore, in order to take into account the similarity in movement directions between two matched prototypes, an additional condition on the directional cosine of road user movements is augmented to Equation 8.5. A minimum threshold $\alpha$ is imposed on the directional cosine of a pair of positions that belong to the same common subsequence.

A key challenge in the adaptation of the LCSS algorithm is the choice of the set of matching parameters track $\{\epsilon, \sigma, \alpha\}$ that maximizes the number of correct violation detections and minimizes the number of missed violation detections. Relevant to this challenge is the study of the sensitivity of results to the selection of the matching parameters. In the following case study, sensitivity to matching parameters was investigated by detecting violations using a sample of all feasible selections of the matching parameters.

### 8.3 Case Study

The video data analyzed in this case study was collected at the Darwaza intersection in Kuwait City, Kuwait. The total length of the observation period was 24 hours. Video tracks from only two hours at dawn were analyzed in this case study. Figure 8.3 shows the results of camera calibration conducted using the methodology presented in Chapter 3. Due to the relative scarcity of pedestrian road users during this time, only vehicle tracks were considered. Traffic violations analyzed in this study involved illegal reverse-direction
turns within the intersection. No other forms of violations were observed. Analysis was restricted to reverse-direction turns within the intersection. Figure 8.4 shows all normal and violation tracks analyzed in this case study. A total of 966 normal tracks and a total of 11 violation tracks were analyzed.

## a) Image space


b)World space


Figure 8.3 Sample grids as projected from world space (b) to the image space (a).


Figure 8.4 Violation tracks (left figure) and a sample of normal tracks (right figure) during 2 hour observations at Darwaza intersection.

Automated violation detection was conducted for various combinations of k means clustering parameters and LCSS matching parameters. For k-means clustering, the natural number of clusters for violation detection is 13 , which includes twelve main traffic movements and one additional cluster to contain all violation tracks. It is also plausible that the use of more than 13 clusters may bring enhanced detection since violating tracks themselves may not exhibit enough similarity to be grouped together within the same cluster. Therefore, analysis was conducted also using 14 and 15 clusters. A range of relative frequency threshold values, as defined in Equation 8.1, were used ranging from 0.005 to 0.1 . Relative frequency thresholds above 0.1 provided significantly high number of false detections. Conversely, relative frequency thresholds below 0.005 produced significantly high number of missed detections. For k-means clustering, road user tracks significantly outside the boundaries of the intersection were truncated in order to discard segments of road user movements irrelevant to traffic regulations within the intersection.

Movement prototypes were learnt for a period of 5,000 frames selected at random from the video sequence. A total of 189 prototypes were recorded. A total of 6 prototypes were manually removed since they belong to violation movements. Figure 8.5 displays a superimposition of all normal prototypes used in LCSS-based classification. Similar to k-means clustering, various combinations of LCSS matching parameters were used. The range for matching distance $\epsilon$ is $2-10 \mathrm{~m}$ with increment 0.5 m . The range for maximum similarity threshold $\sigma$ is $0.1-0.9$ with increment 0.05 . The range for the directional cosine threshold $\alpha$ is $0.7-0.95$ with increment 0.05 . No tracks were
truncated in the case of LCSS-based violation detection due to its complete robustness to variability in track lengths.

Figure 8.6 displays the performance LCSS-based violation detection when $\epsilon$ and $\sigma$ vary in the specified intervals while $\alpha$ is kept constant at 0.9 . There was little sensitivity of detection performance to $\alpha$. In general, for a given short matching distance $\epsilon$ and low $\alpha$, the incidence of false detection of normal tracks as violation tracks is negatively related to the value of the similarity threshold $\sigma$. The same effect on the incidence of false detection, albeit at less sensitivity, was observed for values of $\epsilon$. On the other hand, reducing the value of $\sigma$ under the previous conditions was found to increase missed detection of violation tracks. Similar but more pronounced effect was observed for the selection of $\epsilon$.


Figure 8.5 A superimposition of the 183 normal movement prototypes used for LCSS-based violation detection.


Figure 8.6 Performance of LCSS-based classification using a range of values for matching distance $\boldsymbol{\epsilon}$ in meters and similarity threshold $\boldsymbol{\sigma}$. The directional cosine threshold was set to 0.9 .

The sensitivity of k-means clustering to the initial selection of the centroid of each cluster was evident. Initial selection of each cluster centroid was performed at random. At every iteration, the percentage difference between minimum and maximum Kappa, relative to minimum Kappa value, was calculated. The minimum variation in Kappa statistic was $49 \%$ and the maximum variation was $108 \%$. Figure 8.7 provides evidence of the instability of performance measures for a total of 100 iterations for a sample of three selections of clustering parameters. The exceedingly high percentage correct
detection is mainly due to the imbalance in number between normal tracks and violation tracks. Furthermore, the high number of clusters used resulted in generally high Entropy that masked to some extent individual variations due to different clustering parameters.

The performance of the violation detection using k-means clustering and LCSS matching is shown in Figure 8.8. In order to reduce the effect of random centroid selection on the performance of k-means clustering, performance parameters of k-means clustering were selected based on the median Kappa statistic value among 100 iterations. The results show a clear superiority of LCSS-based detection when a low rate of false detection is desired. However, if a higher rate of false detection is tolerable, both approaches perform well.

In order to further enhance the performance of k-means clustering, it is possible to consider only a subsample of complete road user tracks. Complete road user tracks contain all road user positions while navigating the intersection. Road user tracks shorter than 100 frames ( 4.0 seconds) were excluded from the analysis. These tracks are mostly partial tracks which do not represent the full observed range of road user movements, except for significantly fast moving road users. The total number of tracks tested was 438 tracks including 10 violation tracks. Figure 8.9 shows the performance of automated violation detection using k -means algorithm on the reduced sample size compared with the performance of LCSS-based violation detection. Note that the complete set of tracks was used for LCSS-based violation detection in both Figures 8.8 and 8.9. A summary of different performance measures is presented in Table 8.1. Peak performances of every
performance measure were independently selected for k-means clustering on both full and reduced samples as well as LCSS-based violation detection. False detection and correct detections shown in Table 8.1 were selected for the set of LCSS matching parameters that yielded the minimum summation of these two performance measures. It is evident from all performance measures that the exclusion of partial tracks improved the performance of $k$-means clustering. Furthermore, despite the favourable experimental set-up for $k-$ means clustering using reduced sample size, LCSS-based violation detection was still superior to k-means clustering at low false-positive rates. Based on previous results, there is a significant reduction in Entropy for LCSS-based violation detection. This is largely attributable to the reduction in the number of clusters from 13-15 to 2.


Figure 8.7 Evident instability of detection results of k-means clustering.


Figure 8.8 The receiver operating characteristic curve for the two violation detection approaches presented in this chapter. K-mean clustering was conducted on a full sample size of 986 tracks.


Figure 8.9 The receiver operating characteristic curve for the two violation detection approaches presented in this chapter. K-mean clustering was conducted on a reduced sample size of 448 tracks.

Table 8.1 Summary of peak performance of different violation detection approaches

| Detection <br> Approach | Sample <br> size | No. of tracks |  | Kappa | PCC | Entropy | \% false <br> detection | \% correct <br> detection |
| :--- | :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 976 | 10 | 0.0830 | 0.9897 | 2049 | 19.97 | 90.00 |
|  | Dawn- <br> reduced | 438 | 10 | 0.5262 | 0.9843 | 813 | 8.44 | 100.00 |
| LCSS | Dawn- <br> complete | 976 | 11 | 0.6378 | 0.9906 | 3 | 8.09 | 100.00 |

### 8.4 Conclusions

This chapter presented two approaches and variations thereof for the automated detection of traffic violations. The first section discussed the relevance of traffic violations to road safety analysis. The next section described an adaptation of a traditional clustering approach, the k-means clustering algorithm, for the purpose of automated violation detection. Inherent shortcomings in the k -means algorithm motivated the reliance on more insightful discriminative features of violation movements obtained using the Longest Common Subsequence (LCSS) similarity measure. Subsequent sections presented a successful application of a modified LCSS similarity criterion for the classification of violation movements. The performance of LCSS-based violation detection was generally superior to piecewise k-means clustering, especially when low false detection rates are desirable. The main shortcoming of k-means clustering is the random selection of initial centroid positions. The detection performance proved sensitive to this random selection. One possible solution to this challenge can
be the careful selection of initial centroids to represent normal movement prototypes. While plausible, this reasoning converges to the LCSS matching approach. However, it is not foreseeable that k-means could outperform LCSS-based violation detection if normal movement prototypes were a priori provided. This expected performance differential is because LCSS matching is performed on observed road user positions, not on some approximated model.

Another advantage of the methodology presented in this chapter is the unsupervised learning of normal road user movements. Representative prototypes for normal movement patterns were automatically learnt and used to provide a concrete definition of what constitutes legal and permitted movements. The reliance on the LCSS matching provides a solid foundation for automated violation detection. Moreover, the practical appeal of LCSS based automated violation detection can be improved if normal prototypes are synthesized from prior knowledge of normal traffic movement. For example, a traffic operator may be consulted to provide sketches of normal movement patterns. These sketches can be used to synthesize movement prototypes with some prior assumption regarding operating speed within the intersection.

## Summary Conclusions and Future Work

### 9.1 Background

Safety and sustainability are the two main themes of this thesis. They are also two main pillars of a functional transportation system. Traditionally, the performance of a transportation system has been measured in terms of mobility. However, enhanced mobility often comes at the expense of reducing safety or compromising the mobility of non-motorized modes of travel. Recent studies showed that the cost of road collisions in Canada exceeds the cost of traffic congestion by almost tenfold (Cannon 2006) (Vodden et al. 2007). In addition, there has been a growing grassroots demand for building a sustainable transportation system. This emerging focus on sustainability gives impetus to the study of non-motorized modes of travel. Despite the
demand for accommodating non-motorized modes of travel, key of which is walking, the current methods of analysis and the available data related to those modes of travel suffer from a traditional bias towards motorized modes of travel. In the grand view, this thesis represents a corrective step in the direction of building a safer and more sustainable transportation system.

The epidemic of road collisions still plagues world roads, inflicting 1.3 million casualties every year (WHO 2004), with no exception of Canadian roads. The cost of road collisions in terms of health care, environmental damage, and induced traffic congestion is immense. A recent study by Transport Canada revealed that the annual cost of road collisions is estimated to be \$CDN 62.7 billion/year; a staggering 5\% of the average Canadian Gross Domestic Product (Vodden et al. 2007). The incidence of road collision causes an especially detrimental effect on non-motorized road users mainly due to their physical vulnerability. Moreover, walking, the key mode of non-motorized travel, is performed by the most vulnerable road users. Pedestrians constituted $12 \%$ of total recorded deaths - the second largest group of road user fatalities (Transport Canada 2004). What makes pedestrian safety particularly important is the over-representation of children and young adults ( $0-19$ ) and elderly (65+) road users - with the first group sustaining the highest potential years of life lost and the latter being an increasing age group in Canada (Transport Canada 2004).

Amid mounting cost of road collisions and despite the detrimental effect on the sustainability of transportation systems, the discipline of road safety analysis has not evolved to meet these daunting challenges. To this date, road
safety analysis remains dependent on the observation of road collisions. This analytical approach is challenged on three accounts:

1. High cost of road collisions. This approach warrants safety treatment after the occurrence of road collisions. This carries an especially high price for pedestrians due to the elevated risk of bodily damage compared to other road users.
2. High quality collision data are particularly difficult to obtain. It is often neglected to record the precise collision location, the mechanism of failure that leads to a collision, and the exact timing of the event. The scarcity and limited quality of collision data are major impediments to collision-based road safety analysis.
3. Accidents are rare and random events. It is often necessary to observe collisions over a long period of time in order to discard variations due to the stochastic nature of road collisions and due to confounding factors (Persaud \& Lyon 2007). It is typical for before-and-after observational periods to extend for 1-3 years after the introduction of a safety treatment in order to conduct proper evaluation.

Traffic Conflict Techniques have been advocated as an alternative to or supplementary to collision-based road safety analysis. The reliance on field observations for conducting traffic conflict survey has been challenged on two accounts. The first challenge is the cost required to train human observers and institute the field surveys. The second challenge is the unavoidable subjectivity of road users in observing traffic conflicts. The lack of an
adequate inter- and intra-observer agreement has been a stumbling block toward the development of conflict-based road safety analysis.

This thesis is centered on a number of applications of computer vision techniques for the purpose of traffic data collection and automated road safety analysis. Video sensors have been used as the main source of data in this thesis. Video sensors possess a number of advantages over other data collection technologies. For example, video cameras are relatively inexpensive to procure and operate. Video data provides rich and detailed information on road user movement within the monitored field of view. In addition, many jurisdictions are installing video camera for monitoring purpose, thus greatly facilitating video data collection for computer vision applications.

The main advantage of computer vision techniques is the potential to collect microscopic road user data at a degree of automation and accuracy that cannot be feasibly achieved by manual or semi-automated techniques. Microscopic road user data can be used to draw objective inference on their proximity to the risk of collision. The objectiveness and automation of conducting traffic conflict analysis using computer vision techniques precisely empowers the two main challenges of traditional observer-based traffic conflict analysis: cost and subjectivity. Furthermore, the observation of microscopic pedestrian data using computer vision techniques can address a long-standing challenge of the availability of pedestrian data. In the following sections, summaries of the research work and the drawn conclusions are presented for each chapter. Subsequent to the description of the work is a list of related research contributions.

### 9.1.1 Chapter One: Introduction

The first chapter of this thesis presented background materials, description of research motivation, statement of research problems, and research contribution.

### 9.1.2 Chapter Two: Literature Review

The second chapter of this thesis presented a review of the technical literature on three main topics: traffic conflict analysis, developments in the realm of computer vision on topics of road user tracking, and adoption of computer vision techniques in transportation engineering applications. The first topic was a review of developmental milestones from conceptual proposals to sophisticated implementations. Main approaches for road user tracking in the realm of computer vision were outlined. The problem of pedestrian tracking is distinctively different and arguably more challenging than vehicle tracking. Pedestrians are more prone to visual occlusion, move in a less organized fashion than vehicles, and are locally non-rigid. Two separate sections were presented in Chapter 2 on the problems of vehicle tracking and pedestrian tracking. The last section presented a review of studies conducted in the literature of transportation engineering that involve the adoption of computer vision techniques. Two important conclusions can be drawn from this review:

1. On the theoretical side of traffic conflict analysis, there is little work done on the validation of traffic conflict indicators. The majority of conflict indicators are built on intuition while not contrasted with genuine quantity to be measured. Furthermore, no successful attempt appears to have been achieved in developing a conflict indicator
capable of comprehending the severities of both road collisions and traffic conflicts. The universal severity dimension theorized by Hydén possesses intuitive validity (Hydén 1987). Its implementation however proved to be one of the long-standing challenges in traffic conflict analysis.
2. There is a technological gap between developments made in the realm of computer vision and adoptions in transportation engineering. The majority of the reviewed applications do not venture into using state-of-the-art road user tracking technologies. Much work is to be conducted for increased adoption of computer vision developments in the realm of transportation engineering.

### 9.1.3 Chapter Three: Recovering Real-world Coordinates for Points that Appear in Video Observations

Video sensors have been adopted as the main method of data collection for the analysis conducted in this thesis. The main type of data sought to be recovered by computer vision techniques is road user positions. This chapter presented a methodology for the inference of the position, the orientation, and the various intrinsic parameters of a monitoring camera in order to enable the recovery of real-world positions of points on the road surface. This estimation process is referred to in this chapter as camera calibration. Camera calibration was treated as an optimization problem. The objective function is composed of different components expressed in homogenous metric units. Each cost function component represents a distinct calibration feature. Following are the calibration features considered in the analysis presented in this chapter:

1. Point correspondences.
2. Difference between true length of line segments and their backprojected ${ }^{1}$ length.
3. Discrepancy between true angles between vectors and the angle measured between back-projections of these vectors.
4. Angular discrepancy between annotated and projected vertical line segments.
5. Difference in back-projected lengths of pairs of linear measurements of an edge that appears at two different depths from the camera.

A toolbox was developed as an implementation of the presented methodology written in the Matlab language (Mathworks 2010). Subsequent applications were supported by camera parameter estimation using this toolbox. Furthermore, this toolbox was used in several studies outside the scope of this thesis in which the estimated camera parameters proved to be successful. Robustness was proved against various challenges such as degraded image quality, lack of orthographic image of the monitored site, lack of knowledge of intrinsic camera parameters, and dearth of reliable geometric primitives used as calibration features. Following are the research contributions achieved in this chapter:

1. A reliable methodology was developed for recovering real-world coordinates of points that appear in the video sequence.

[^18]2. The previous methodology was implemented in the MATLAB language (Mathworks 2010) and was successfully used in all video analysis undertaken in this thesis (six different scenes) as well as other applications outside the scope of this thesis (nine different scenes).
3. The accuracy of camera parameter estimates was found to be superior to current practical requirements for the purpose of road users tracking.

### 9.1.4 Chapter Four: Automated Measurement of Pedestrian Walking Speed

In this chapter, it was demonstrated that the application of computer vision techniques offers an appealing solution to demands for more efficient and accurate methods of pedestrian data collection. Pedestrian walking speed has been the subject of continuous research. The motivation for the study of walking speed is the proper design of traffic signals to accommodate the changes in the characteristics of the average pedestrian, mainly due to demographic changes. The majority of commercial techniques developed for the purpose of automated observation of traffic data focus primarily on vehicular traffic. The technological aspects of automated pedestrian data collection are generally more challenging than vehicular traffic. The majority of walking speed studies in the literature do not adopt automated video analysis for collecting pedestrian data.

In this chapter, an automated system for collecting pedestrian walking speed using video analysis was developed and tested. The video analysis system was tested on real video data collected at the Downtown area of Vancouver,

British Columbia, during day- and night-time conditions. Validation of walking speed measurements against manual observations proved the accuracy of automated measurements. The following conclusions were drawn from the research work presented in this chapter:

1. Based on the review of relevant studies it can be argued that the literature of pedestrian observational studies is yet to benefit from automated video analysis techniques. It is expected that the system presented in this study will be further improved by adding other appearance-based techniques.
2. Walking speed measurements were found to be sensitive to the estimation of camera parameters. Estimation of camera parameters using few point correspondences did not support accurate measurement of walking speed.
3. An application of the camera calibration methodology, as presented in Chapter 3, provided sufficiently accurate estimates of camera parameters.
4. Pedestrians walk faster at marked crosswalks than sidewalks.
5. Night-time conditions proved to be the most difficult as expected because of the obscurity of pedestrian outlines and video recording noise. A special set of detection parameters was used for night videos and results obtained were satisfactory.
6. Walking speed was more variable at unmarked crosswalks compared to marked crosswalks.
7. Road surface gradient and lighting conditions were identified as statistically significant variables that influence walking speed.

The research work presented in Chapter 4 was problem driven. The main objective was to adapt a feature-based tracking system developed for vehicle tracking for the purpose of measuring pedestrian walking speed. In meeting this objective, the following research contributions were achieved:

1. Average walking speed was measured from a relatively large sample of pedestrian movements and at adequate accuracy. Pedestrians were observed for a total of two hours while moving to a local gathering held annually in Vancouver, British Columbia.
2. The aggregate estimate of average walking speed obtained in this chapter may serve as a key design variable in crowd management, traffic signal design, and design of pedestrian facilities.
3. Statistical analysis of the measurements was conducted in order to investigate the variance of walking speed under different conditions such as time of the day, type of pedestrian facility, movement direction, and longitudinal pavement slope. The result of this analysis provides useful insight into the considerations required for the design of pedestrian facilities under different operational and physical conditions.

### 9.1.5 Chapter Five: Automated Detection of Pedestrian-vehicle Conflicts

The work performed in Chapters 3 and 4 provided a solid basis for pursuing an important focus of this thesis; the study of pedestrian-vehicle conflicts. By the conclusion of Chapter 4, it was evident that the video analysis system reached a level of development that supports accurate recovery of pedestrian
positions in real-world coordinates. Consequent to this functionality is to measure the severity of traffic events that involve pedestrian and vehicles.

The work presented in Chapter 5 demonstrated the feasibility of using automated video analysis for achieving the following objectives: 1) detect and track road users in a traffic scene, and classify them as pedestrian and motorized road users using a maximum speed threshold; 2) identify important events that may lead to collisions; 3) calculate several severity conflict indicators. The functionality of the system was demonstrated on a video dataset collected over two days at an intersection in Downtown Vancouver, British Columbia. Four conflict indicators were automatically computed for all pedestrian-vehicle events and provide detailed insight in the conflict process.

The quality of four conflict indicators, Time-to-Collision, Post-Encroachment Time, Gap Time, and Deceleration-to-Safety Time, were assessed in regard to their ability to comprehend the severity of traffic conflicts. None of the conflict indicators were individually capable of capturing all dangerous interactions between road users. However, a combination of the four indicators proved to be useful in the identification of important events and traffic conflicts. For this purpose, simple detection rules defined over the four conflict indicators were tested to classify traffic events. This study was successful in the attempt to extract conflict indicators from video sequences in a fully automated way. In tackling this research problem, a number of research contributions have been achieved:

1. Successful application of feature-based tracking for the purpose of detecting, tracking, and correctly classifying pedestrians involved in traffic conflicts with crossing vehicles.
2. The development of various methods of analysis for the automated and objective measurement of the severity of pedestrian-vehicle conflicts. In order to measure the accuracy of detecting pedestrianvehicle conflicts, system output was contrasted with observer-based traffic conflict analysis. The results demonstrated the effectiveness of the developed methodology for detecting pedestrian-vehicle conflicts.

### 9.1.6 Chapter Six: Automated Safety Analysis in a Before-and-After Context

Data availability has been a common challenge to pedestrian studies; especially for proactive safety analysis. The traditional reliance on collision data has many shortcomings in terms of the quality and quantity of collision record. The feasibility of automated analysis of pedestrian-vehicle conflicts has been demonstrated in Chapter 5. One of the focus areas of pedestrian safety that will greatly benefit from vision-based road user tracking is before-and-after (BA) evaluation of safety treatments. One of the main shortcomings of collision-based road safety analysis is the extended observational period required to identify stable trends. For example, typical observational period for before-and-after studies of safety treatments is 1-3 years. A key advantage of the method of analysis developed in the previous problem is to shorten this observational period from a time span of 1-3 years possibly to few weeks.

This chapter demonstrated the feasibility of conducting BA analysis using video data collected from a commercial-grade camera in Chinatown, Oakland, California. Video sequences for a period of two hours before and two hours after the introduction of a pedestrian scramble phase were automatically analyzed. The before-and-after results of the automated analysis exhibit a pattern of declining conflict frequency, a reduction in the spatial density of conflicts, and a shift in spatial distribution of conflicts further from crosswalks. From the analysis presented in this chapter several conclusions can be drawn:

1. The reliance on motion prototypes demonstrated a clear advantage over classification methods used in previous studies.
2. The context of this study was the evaluation of the safety benefits of introducing the pedestrian scramble phase. A two-hour video sequence was analyzed for pre- and post-scramble. Despite that the video analyzed in this study was not collected for the purpose of automated analysis, tracking accuracy was satisfactory. The automated analysis of four conflict indicators showed a reduction in conflict frequency. In addition, there was a general reduction in the spatial density of conflicts after the safety treatment.
3. It was not attempted in this chapter to draw a statistical inference regarding the safety benefit of the pedestrian scramble phase. The only work on this subject found in the literature involved statistical tests for the difference in the shapes of conflict indicator distributions between before and after periods. This research need was the motivation behind the research work presented in Chapter 7 in which a severity index
was developed. This severity index captures both the differences in severity and exposure to collision risk between before and after periods.
4. A random sample of 366 automatically detected traffic events was considered for manual review. Among the 266 events in the before period, 17 events were found to involve classification error. Among the 100 events reviewed for the after period, 12 events were found to involve misclassification of pedestrians into vehicles.

In order to meet the research objective set forth in this chapter, the following key contributions were achieved:

1. The development of a new methodology for road user classification into pedestrian and non-pedestrians. The performance of the classification approach was found to be superior to the maximum speed classification adopted in Chapters 4 and 5.
2. The novel application of automated analysis of pedestrian-vehicle conflicts in the context of before-and-after evaluation of safety treatments. This contribution component represents the first attempt to probe the severity of all traffic events that involve a pedestrians and non-pedestrians in an automated fashion. The results of the automated system were to a satisfactory degree consistent with findings in the literature.

### 9.1.7 Chapter Seven: Development of an Aggregate Safety Index

The research work presented in Chapter 7 was mainly motivated by the lack of a mechanism for drawing a safety inference based on the microscopic
traffic conflict data obtained using computer vision techniques. The various objective severity measures and conflict indicators calculated in Chapters 5 and 6 were hypothesized to be of different and sometimes of independent nature. Empirical findings were in support of this hypothesis.

A number of approaches have been proposed in order to map measurements of four conflict indicators into the severity dimension and to integrate conflict indicators into a severity index. The first approach assumes that the percentile ranking of each conflict indicator measurement is representative of its severity. The second approach is based on severity benchmarks found in the literature. In addition, aggregation of conflict indicator and severity index measurements was advocated. A number of aggregation approaches have been proposed: aggregation over time, over road user, and over traffic events. In this chapter, an important proposition was presented for an aggregate safety measure that reflects the underlying level of road safety. At the end, in discussing the findings of this chapter it was hypothesized that total severity measures and average severity measures reflect two different and orthogonal dimensions for measuring road safety. While total severity reflects the magnitude of road safety, average severity possible represents the safety level perceived by road users.

The following contributions were achieved in this chapter:

1. The development of an objective methodology for integrating various severity aspects measured by different conflict indicators into a single severity index. The proposed methodology was tested on the video
data used for developing the application on before-and-after safety evaluation presented in Chapter 6.
2. Two hypotheses were raised, Hypotheses 7.1 and 7.2. The first hypothesis states that conflict indicators provide partial severity aspects. The second hypothesis conjectures that total severity, measured by the summation of severity measurements over all traffic events, and average severity, expressed as severity per events, are orthogonal dimensions of measuring safety. The first hypothesis was supported by empirical evidence obtained in this study. Hypothesis 7.2 was demonstrated on conceptual grounds.
3. A number of aggregation approaches and variations thereof were proposed.
4. A safety index was proposed that captures both total severity of traffic events and the maximum possible level of exposure.

### 9.1.8 Chapter Eight: Automated Detection of Traffic Violations

A major focus of this thesis was on traffic conflicts as a data type for measuring road safety. However, observational periods may be too limited to record a representative sample of traffic conflicts. In this case, traffic violations may provide surrogate safety information which may be especially relevant if road collisions are attributable to some occurrence of traffic violations.

This chapter presented an approach for the automated detection of traffic violations. Two approaches were tested for this purpose, k -means clustering and Longest Common Subsequence (LCSS) similarity measure. Inherent
shortcomings in the k -means algorithm motivated the reliance on more insightful discriminative features of violation movements obtained using the LCSS similarity measure. The two approaches were used to detect violating reverse-direction turning vehicular movements at an intersection in Kuwait city. The length of the analyzed video sequence was 2 hours. The focus of the analysis in this chapter was on vehicular violations. There was not adequate number of pedestrians using the intersection in order to analyze their violations. The developed methodology can be readily applied for detecting pedestrian violations marked by distinct spatial or directional characteristics.

The performance of LCSS-based violation detection was generally superior to piecewise k-means clustering, especially when low false detection rates are desirable. The main shortcoming of k-means clustering is the random selection of initial centroid positions. The detection performance proved to be sensitive to this random selection. By addressing the research problem related to this chapter, the following research contributions were achieved:

1. Proposal for a methodology for extracting features of road user tracks for clustering analysis. A piecewise linear model was used to represent road user tracks. The parameters of each model were adopted for subsequent k-means clustering analysis.
2. Successful adaption of LCSS similarity measure for the purpose of automated violation detection.

### 9.2 Recommendation for Future Work

The research work presented in this thesis lies at two frontiers: pedestrian tracking using computer vision techniques and automated road safety analysis. The broad research problem of pedestrian tracking has attracted increasing attention in the realm of computer vision. However, many aspects of this research objective remain unsolved. A universally accurate pedestrian tracking technology remains an elusive target and is likely to remain so for a number of years to come. It is foreseeable that substantial developments are to be achieved on this subject. Since these developments are likely to take place in the realm of computer vision, this section focuses on future work in the realm of road safety analysis.

Automated road safety analysis empowers the traditional weaknesses brought about by the reliance on collision data and traffic conflict observations gathered by human observers. This research is among the few steps taken to achieve the long-sought objective of conducting automated and objective road safety analysis. Future work can be classified into three categories: validation of automated traffic conflict analysis, development of a probabilistic conflict indicator, and a number of different extensions of the work presented in this thesis.

### 9.2.1 Validation of Automated Traffic Conflict Analysis

In general, validation of a safety analysis technique concerns its potential to measure road safety. Traditionally, validation of traffic conflict techniques was conducted by contrasting the frequencies of traffic conflicts and road
collisions. The advocated approach for traffic conflict analysis can be validated through the following research work:

1. Collecting video observations of a representative sample of traffic intersections, in the order of hundreds of intersections. This video data should be automatically analyzed in order to measure the severity of observed traffic events.
2. Historical collision data should be collected for the monitored intersections as well as a reference group of intersections.
3. It is preferable to disaggregate road collision and traffic conflict observations by type, location, and time. Should this level of disaggregation result in few data points per aggregate category, data points in different categories can be combined.
4. Contrasting severity measurements obtained from automated traffic conflict analysis with human expert's evaluation. The severity mapping approaches presented in this thesis can potentially be calibrated based on expert evaluations.
5. The hypothesis of validity to be tested is whether the integration of safety cues obtained from automated traffic conflict analysis results in improvement in the prediction of road collisions. Furthermore, the degree of improvement in predictive power should be investigated.
6. Benchmark of severity distributions should be established from the previous large-scale video data observations. Such severity distributions can be used to identify abnormalities in the incidence or severity of traffic conflicts observed at traffic intersections with similar characteristics.

### 9.2.2 Potential Developments of Conflict Indicators

The shortcomings of traditional conflict indicators have been discussed in this thesis. Recent developments on this topic suggests the extrapolation of road user positions based on a current road user track as well as records of previous road user tracks traversing the intersection in a fashion similar to the current road user. While initial attempts have been performed to develop probabilistic conflict indicators, in general they fell short of achieving this goal. This future work concerns the development of a severity measure in terms of the probability of collision between two conflicting road users. Probability of collision is therefore the joint probability of two road users to remain on a collision course and the probability that both of them fail to undertake a successful evasive action. This future work takes a step forward from previous work to crystallize the representation of the uncertainty regarding future positions of road users in probabilistic terms.

The ambitious modelling of future road user positions and evasive actions in a probability model is supported by the extensive observation of road user tracks that is possible using computer vision techniques. This development represents a leap forward from observer-based and subjective methods of traffic conflict analysis as to the proximity evaluation of road user positions. Figure 9.1 shows an example of the tracks of two road users; $U_{1}$ and $U_{2}$. Future positions of every road user are represented as a sequence of bivariate Gaussian. The variance of every Gaussian reflects the precision of positional predictions. The probability that the two road users become closer than some threshold can be evaluated at any defined moment in the future. Furthermore, the conditional probability of the two road users being unable
to make a successful evasive action should be integrated to produce a single proximity measure. Finally, this measure should be augmented with a measure of the severity of the impending collision.

One particular shortcoming of previous probabilistic conflict indicator developments is the absence of a proper modelling of the uncertainty in road user destinations. The purpose of estimating road user trajectory is to predict how road users might have driven to end up in a collision. Precisely, knowledge is lacking as to the destination of each road user and the deviation from some defined typical course of movement. The first source of uncertainty can however be eliminated by observing a road user further in time. In previous work, the two types of uncertainties were indiscriminatingly used and produced inaccurate probabilities of collision based on road user trajectories that would never be followed. A possible development to overcome this challenge is to retroactively remove from the sample space of possible road user positions those points that correspond to a road user destination that is significantly different from the actual destination targeted by the tracked road user.


Figure 9.1 A hypothetical case of two road users $U_{1}$ and $U_{2}$ predicted to follow two trajectories $\mathrm{H}_{12}$ and $\mathrm{H}_{22}$ that lead to a potential collision. Predicted positions of road users are approximated using a bivariate Gaussian.

### 9.2.3 Other Extensions to Work Presented in this Thesis

Following is a list of future studies that represent potential continuation of the work presented in Chapters 3 to 8:

1. A general extension of the work presented in this thesis is the continuation of creating larger video libraries of road user movement.
2. Formulation of the cost function presented in Equation 3.4 in Chapter 3 using geometric algebra. In this algebra, different geometric elements can be uniformly represented in harmonious coordinate system. Further improvements to the proposed methodology should consider
the inclusion of additional non-linear parameters other than radial lens distortion.
3. Investigating the effect of camera set-up, hardware configuration, and camera jitter on the accuracy of road user tracking.
4. Investigating the use of computer vision techniques for collecting other forms of pedestrian data such as density, flow, and level of service.
5. Investigating on empirical ground Hypotheses 7.1 and 7.2 posed in Chapter 7. In particular, while it was conjectures in Hypothesis 7.2 that total severity and average severity represent two orthogonal dimensions for measuring road safety, it is not known how a combination of their measurement should be performed.
6. Investigating the reliance on synthetic road user prototypes for automatically detecting traffic violations. For example, a traffic operator may be consulted to provide sketches of normal movement patterns. These sketches can be used to synthesize these movement prototypes with some prior assumption regarding operating speed within the intersection.

## References

- AASHTO, 2001, A Policy on Geometric Design of Highways and Streets, Washington, D.C.
- Abrams, C \& Smith, S 1977, 'Selection of Pedestrian Signal Phasing', Transportation Research Record: Journal of the Transportation Research Board, vol 692, pp. 1-6.
- Acharjee, S, Kattan, L \& Tay, R 2009, 'Pilot Study on Pedestrian Scramble Operations in Calgary', Transportation Research Record: Journal of the Transportation Research Board, vol 2140, pp. 79-84.
- Agarwal, S, Awan, A \& Roth, D 2004, 'Learning to Detect Objects in Images via a Sparse, Part-based Representation', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 26, no. 11, pp. 14751490.
- Ahmed, K 1999, 'Modeling Divers'acceleration and Lane Changing Behavior', Sc.D., MIT, Boston, MA.
- Akin, D \& Sisiopiku, V 2007, 'Modeling Interactions Between Pedestrians and Turning Vehicles at Signalized Crosswalks Operating Under Combined Pedestrian-Vehicle Interval', Transportation Research Board Annual Meeting, Washington, DC.
- Al-Azzawi, M \& Raeside, R 2007, 'Modeling Pedestrian Walking Speeds on Sidewalks', Journal of Urban Planning and Development, vol 133, no. 3, pp. 211-219.
- AlGhadi, S, Mahmassani, H \& Herman, R 2002, 'A Speed-concentration Relation for Bi-directional Crowd Movements with Strong Interaction', Internaltional Conference on Pedestrian and Evacuation Dynamics, Springer, Berlin, Germany.
- Al-Haji, G 2005, 'Towards a Road Safety Development Index', PhD Thesis, Linköpings University.
- Allen, B, Shin, B \& Cooper, P 1978, 'Analysis of Traffic Conflicts and Collisions', Transportation Research Record: Journal of the Transportation Research Board, vol 667, pp. 67-74.
- Al-Masaeid, H, Al-Suleiman, T \& Nelson, B 1993, 'Pedestrian Speed-flow Relationship for Central Business District Areas in Developing Countries', Transportation Research Record: Journal of the Transportation Research Board, vol 1396, pp. 69-74.
- Alonso, I, Llorca, D, Sotelo, M, Bergasa, L, de Revenga, P, Nuevo, J, Ocana, M \& Garrido, M 2007, 'Combination of Feature Extraction Methods for SVM Pedestrian Detection', IEEE Transactions on Intelligent Transportation Systems, vol 8, no. 2, pp. 292-307.
- Amundsen, F \& Hydén, C 1977, , First Workshop on Traffic Conflicts, Institute of Transport Economics, Oslo, Norway.
- Antonini, G, Martinez, V, Bierlaire, M \& Thiran, J 2006, 'Behavioral Priors for Detection and Tracking of Pedestrians in Video Sequences', International Journal of Computer Vision, vol 69, no. 2, pp. 159-180.
- Archer, J 2004, 'Methods for the Assessment and Prediction of Traffic Safety at Urban Intersections and their Application in Micro-simulation Modelling', PhD Thesis, Royal Institute of Technology.
- Ayuso, M, Guillén, M \& Alcañz, M 2010, 'The Impact of Traffic Violations on the Estimated Cost of Traffic Accidents with Victims', Accident Analysis and Prevention, vol 42, no. 2, pp. 709-717.
- Badenas, J, Sanchiz, J \& Pla, F 2000, 'Using Temporal Integration for Tracking Regions in Traffic Monitoring Sequences', 15th International Conference on Pattern Recognition, Barcelona, Spain.
- Ballard, D \& Brown, C 1982, Computer Vision, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, p.2.
- Baro, X, Escalera, S, Vitria, J, Pujol, O \& Radeva, P 2009, 'Traffic Sign Recognition Using Evolutionary Adaboost Detection and Forest-ECOC Classification', IEEE Transactions on Intelligent Transportation Systems, vol 10, no. 1, pp. 113-126.
- Bechtel, A, MacLeod, K \& Ragland, D 2003, 'Oakland Chinatown Pedestrian Scramble: An Evaluation', Final Report, UC Berkeley Traffic Safety Center.
- Bechtel, A, MacLeod, K \& Ragland, D 2004, 'Pedestrian Scramble Signal in Chinatown Neighborhood of Oakland, California: an Evaluation', Transportatoin Research Record: Journal of the Transportation Research Board, vol 1878, pp. 19-26.
- Beymer, D, McLauchlan, P, Coifman, B \& Malik, J 1997, 'A Real-time Computer Vision System for Measuring Traffic Parameters', Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Computer Society.
- Beymer, D, McLauchlan, P, Coifman, B \& Malik, J 1998, 'A Real-time Computer Vision System for Measuring Traffic Parameters', Transportation Research Part C: Emerging Technologies, vol 6, no. 4, pp. 271288.
- Bi, L, Tsimhoni, O \& Liu, Y 2009, 'Using Image-Based Metrics to Model Pedestrian Detection Performance With Night-Vision Systems', IEEE Transactions on Intelligent Transportation Systems, vol 10, no. 1, pp. 155-164.
- Boillot, F, Midenet, S \& Pierrelée, J-C 2006, 'The Real-time Urban Traffic Control System CRONOS: Algorithm and Experiments', Transportation Research Part C: Emerging Technologies, vol 14, no. 1, pp. 18-38.
- Boles, W \& Hayward, S 1978, 'Effects of Noise and Sidewalk Density upon Pedestrian Cooperation and Tempo', Journal of Social Psychology, vol 104, no. 1, pp. 29-35.
- Bougnoux, S 1998, 'From Projective to Euclidean Space Under any Practical Situation, a Criticism of Self-Calibration', International Conference on Computer Vision, Washington, DC.
- Bowerman, W 1973, 'Ambulatory Velocity in Crowded and Uncrowded Conditions', Perceptual and Motor Skills, vol 36, no. 1, pp. 107-111.
- Bowman, B \& Vecellio, R 1994, 'Pedestrian Walking Speeds and Conflicts at Urban Median Locations', Transportation Research Record: Journal of the Transportation Research Board, vol 1438, pp. 67-73.
- Bradski, G \& Pisarevsky, V 2000, 'Intel's Computer Vision Library: Applications in Calibration, Stereo Segmentation, Tracking, Gesture, Face and Object Recognition', IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, SC.
- 'British Columbia Traffic Collision Statistics' 2005, 1203-8008, Insurance Corporation of British Columbia.
- Cameron, R 1977, 'Pedestrian Volume Characteristics', Traffic Engineering, vol 47, no. 1, pp. 36-37.
- Cannon, L 2006, 'The Cost of Urban Traffic Congestion in Canada', Transport Canada, Environmental Affairs.
- Cao, X-B, Qiao, H \& Keane, J 2008, 'A Low-Cost Pedestrian-Detection System with a Single Optical Camera', IEEE Transactions on Intelligent Transportation Systems, vol 9, no. 1, pp. 58-67.
- Caprile, B \& Torre, V 1990, 'Using Vanishing Points for Camera Calibration', International Journal of Computer Vision, vol 4, no. 2, pp. 127-140.
- Chae, K \& Rouphail, N 2008, 'Empirical Study of Pedestrian-Vehicle Interactions in the Vicinity of Single-Lane Roundabouts', Transportation Research Board Annual Meeting, Washington, DC.
- Cheek, M, Hawkins, H \& Bonneson, J 2008, 'Improvements to Queue and Delay Estimation Algorithm Utilized in Video Imaging Vehicle Detection Systems', Transportation Research Board Annual Meeting, Washington, DC.
- Chin, H-C \& Quek, S-T 1997, 'Measurement of Traffic Conflicts', Safety Science, vol 26, no. 3, pp. 169185.
- Chitturi, M, Medina, J \& Benekohal, R 2010, 'Effect of Shadows and Time of Day on Performance of Video Detection Systems at Signalized Intersections', Transportation Research Part C: Emerging Technologies, vol 18, no. 2, pp. 176-186.
- Choudhury, C \& Ben-Akiva, M 2008, 'Lane Selection Model for Urban Intersections', Transportation Research Record: Journal of the Transportation Research Board, vol 2088, pp. 167-176.
- Conche, F \& Tight, M 2006, 'Use of CCTV to Determine Road Accident Factors in Urban Areas', Accident Analysis and Prevention, vol 38, no. 6, pp. 1197-1207.
- Cooper, J 1983, 'Experience with Traffic Conflicts in Canada with Emphasis on 'Post-encroachment Time' Techniques'. International Calibration Study of Traffic Conflict Techniques, pp. 75-96.
- Costa, M \& Shapiro, L 2000, '3D Object Recognition and Pose with Relational Indexing', Computer Vision and Image Understanding, vol 79, no. 3, pp. 364-407.
- Cui, J, Zhao, H \& Shibasaki, R 2006, 'Fusion of Detection and Matching Based Approaches for Laser Based Multiple People Tracking', International Conference on Computer Vision and Pattern Recognition, New York, NY.
- Cunto, F \& Saccomanno, F 2008, 'Calibration and Validation of Simulated Vehicle Safety Performance at Signalized Intersections', Accident Analysis and Prevention, vol 40, no. 3, pp. 1171-1179.
- Cynecki, M 1980, 'Development of Conflicts Analysis Technique for Pedestrian Crossings', Transportation Research Record: Journal of the Transportation Research Board, vol 743, pp. 12-20.
- Dahlstedt, S 1978, 'Walking Speed and Walking Habits of Elderly People', National Swedish Road and Traffic Research Institute, Stockholm, Sweden.
- Dalal, N \& Triggs, B 2005, 'Histograms of Oriented Gradients for Human Detection', IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA.
- Davis, G 2003, 'Bayesian Reconstruction of Traffic Accidents', Law Probability and Risk, vol 2, no. 2, pp. 69-89.
- Davis, G, Hourdos, J \& Xiong, H 2008, 'Outline of Causal Theory of Traffic Conflicts and Collisions', Transportation Research Board Annual Meeting, Washington, DC.
- Davis, S, King, L \& Robertson, A 1988, 'Predicting Pedestrian Crosswalk Volumes', Transportation Research Record: Journal of the Transportation Research Board, vol 1168, pp. 25-30.
- de-Leur, P \& Sayed, T 2003, 'A Framework to Proactively Consider Road Safety within the Road Planning Process', Canadian Journal of Civil Engineering, vol 30, no. 4, pp. 711-719.
- Dickmanns, E 2002, 'The Development of Machine Vision for Road Vehicles in the Last Decade', IEEE Intelligent Vehicle Symposium, Versailles, France.
- Doiron, R 2008, 'Annual Canadian Vehicle Survey', Online, Statistics Canada.
- Douma, F, Frooman, S \& Deckenbach, J 2008, 'What You Need to Know - and Not Know: Current and Emerging Privacy Law for ITS', Transportation Research Board Annual Meeting, Washington, DC.
- Dubrofsky, E \& Woodham, R 2008, 'Combining Line and Point Correspondences for Homography Estimation', Lecture Notes in Computer Science, vol 5359, pp. 202-213.
- Elliott, M, Baughan, C \& Sexton, B 2007, 'Errors and Violations in Relation to Motorcyclists' Crash Risk', Accident Analysis and Prevention, vol 39, no. 3, pp. 491-499.
- Elmitiny, N, Yan, X, Radwan, E, Russo, C \& Nashar, D 2010, 'Classification Analysis of Driver's Stop/Go Decision and Red-light Running Violation', Accident Analysis and Prevention, vol 42, no. 1, pp. 101-111.
- EnvCanada 2008, 'National Inventory Report 1990-2006: Greenhouse Gas Sources and Sinks in Canada', Greenhouse Gas Division, Environment Canada.
- Enzweiler, M \& Gavrila, D 2008, 'A Mixed Generative-discriminative Framework for Pedestrian Classification', IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK.
- Enzweiler, M \& Gavrila, D 2009, 'Monocular Pedestrian Detection: Survey and Experiments', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 31, no. 12, pp. 2179-2195.
- Fang, C, Fuh, C, Yen, P, Cherng, S \& Chen, S 2004, 'An Automatic Road Sign Recognition System Based on a Computational Model of Human Recognition Processing', Computer Vision and Image Understanding, vol 96, no. 2, pp. 237-268.
- Fan, L, Sung, K-K \& Ng, T-K 2003, 'Pedestrian Registration in Static Images with Unconstrained Background', Pattern Recognition, vol 36, no. 4, pp. 1019-1029.
- Fitzpatrick, K, Turner, S, Brewer, M, Carlson, P, Ullman, B, Trout, N, Park, E-S, Whitacre, J, Lalani, N \& Lord, D 2006, 'Improving Pedestrian Safety at Unsignalized Crossings', TCRP Report 112/NCHRP Report 562, Transportation Research Board.
- Forsyth, D, Arikan, O, Ikemoto, L, O'Brien, J \& Ramanan, D 2005, 'Computational Studies of Human Motion: Part 1, Tracking and Motion Synthesis', Foundations and Trends in Computer Graphics and Vision, vol 1, no. 2/3, pp. 77-254.
- Fozard, J 1981, 'Person-environment Relationships in Adulthood: Implications for Human Factors Engineering', Human Factors, vol 23, no. 1, pp. 7-27.
- Fruin, J 1970, 'Designing for Pedestrians - a Level of Service Concept', PhD Dissertation, The Polytechnic Institute of Brooklyn.
- Fugger, T, Randles, B, Stein, A, Whiting, W \& Gallagher, B 2000, 'Analysis of Pedestrian Gait and Perception-reaction at Signal-controlled Crosswalk Intersections', Transportation Research Record: Journal of the Transportation Research Board, vol 1705, pp. 20-25.
- Gårder, P 1989, 'Pedestrian Safety at Traffic Signals: A Study Carried out with the Help of a Traffic Conflicts Technique', Accident Analysis and Prevention, vol 21, no. 5, pp. 435-444.
- Gates, T, Noyce, D, Bill, A \& Van Ee, N 2006, 'Recommended Walking Speeds for Timing of Pedestrian Clearance Intervals Based on Characteristics of the Pedestrian Population', Transportation Research Record: Journal of the Transportation Research Board, vol 1982, pp. 38-47.
- Gavrila, D 2007, 'A Bayesian, Exemplar-Based Approach to Hierarchical Shape Matching', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 29, no. 8, pp. 1408-1421.
- Gavrila, D \& Munder, S 2007, 'Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle', International Journal of Computer Vision, vol 73, no. 1, pp. 41-59.
- Glauz, W \& Migletz, D 1984, 'The Traffic Conflict Technique of the United States of America', NATO ASI Series, F5, Springer-Verlag, Heidelberg, Germany.
- Goh, P \& Lam., W 2004, 'Pedestrian Flows and Walking Speed: A Problem at Signalized Crosswalks', Institute of Transportation Engineers Journal, vol 74, pp. 28-33.
- Golani, A \& Damti, H 2007, 'Model for Estimating Crossing Times at High-Occupancy Crosswalks', Transportation Research Record: Journal of the Transportation Research Board, vol 2002, pp. 125-130.
- Grayson, G \& Hakkert, A 1987, 'Accident Analysis and Conflict Behaviour', Road User and Traffic Safety, Van Gorcum Publishers, The Netherlands.
- Grayson, G, Hydén, C, Kraay, J, Muhlrad, N \& Oppe, S 1984, 'The Malmö Study: a Calibration of Traffic Conflict Techniques', Technical Report, Institute of Road Safety Research, R-84-12, Leidschendam, the Netherlands.
- Greenberg, E 2005, 'Toward a New Urbanist Transportation Agenda', Congress for the New Urbanism, Kansas City, KA.
- Greene-Roesel, R, Diógenes, M \& Ragland, D 2007, 'Estimating Pedestrian Accident Exposure: Protocol Report', Institute of Transportation Studies. UC Berkeley Traffic Safety Center, California PATH.
- Greene-Roesel, R, Diógenes, M, Ragland, D \& Lindau, L 2008, 'Effectiveness of a Commercially Available Automated Pedestrian Counting Device in Urban Environments: Comparison with Manual Counts', Transportation Research Board Annual Meeting, Washington, DC.
- Guerrier, J \& Sylvan, C 1998, 'Give Elderly Pedestrians More Time to Cross Intersections', Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Santa Monica, CA.
- Guerrier, J \& Sylvan, C 1998, 'Give Elderly Pedestrians More Time To Cross Intersections', Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting.
- Hamdar, S, Treiber, M \& Mahmassani, H 2009, 'Calibration of a Stochastic Car-following Model Using Trajectory Data: Exploration and Model Properties', Transportation Research Board Annual Meeting, Washignton, DC.
- Harkey, D 1995, 'Intersection Geometric Design Issues for Older Drivers and Pedestrians', Institute of Transportation Engineers 65 th Annual Meeting, Denver, Colorado.
- Harkey, D, Srinivasan, R, Baek, J, Council, C, Eccles, K, Lefler, N, Gross, F, Persaud, B, Lyon, C, Hauer, H \& Bonneson, J 2008, 'Accident Modification Factors for Traffic Engineering and ITS Improvements', NCHRP 617, Transportation Research Board, Washington, DC.
- Hauer, E 1982, 'Traffic Conflicts and Exposure ', Accident Analysis and Prevention, vol 14, no. 5, pp. 359364.
- Hauer, E \& Gårder, P 1986, 'Research into the Validity of Traffic Conflict Techniques', Accident Analysis and Prevention, vol 18, no. 6, pp. 471-481.
- Hayhurst, E 1932, 'Review of Industrial Accident Prevention: a Scientific Approach', American Journal of Public Health Nations Health, vol 22, no. 1, p. 119-120.
- Hayward, J 1968, 'Near-miss Determination through Use of a Scale of Danger', Highway Research Record, vol 384, pp. 24-34.
- HCM 2000, Highway Capacity Manual, Transportation Research Board, National Research Council, Washington, DC.
- Heap, T \& Hogg, D 1998, 'Wormholes in Shape Space: Tracking through Discontinuous Changes in Shape', International Conference on Computer Vision, Bombay, India.
- Herman, L \& Nadella, S 2005, 'Accuracy of a Traffic Noise Model Using Data from Machine Vision Technology', Transportation Research Record: Journal of the Transportation Research Board, vol 1941, pp. 155160.
- Hermans, E, Van den Bossche, F \& Wets, G 2009, 'Uncertainty Assessment of the Road Safety Index', Reliability Engineering \& System Safety, vol 94, no. 7, pp. 1220-1228, Special Issue on Sensitivity Analysis.
- Hoogendoorn, S \& Daamen, W 2003, 'Extracting Microscopic Pedestrian Characteristics from Video Data', Transportation Research Board Annual Meeting, Washington, DC.
- Hoogendoorn, S \& Daamen, W 2006, 'Microscopic Parameter Identification of Pedestrian Models and Implications for Pedestrian Flow Modeling', Transportation Research Record: Journal of the Transportation Research Board, vol 1982, pp. 57-64.
- Hoogendoorn, S, Zuylen, H, Schreuder, M, Gorte, B \& Vosselman, M 2002, 'Microscopic Traffic Data Collection by Remote Sensing', Transportation Research Record: Journal of the Transportation Research Board, vol 1885, pp. 121-128.
- Hoogendorn, S, Daamen, W \& Bovy, P 2003, 'Extracting Microscopic Pedestrian Characteristics from Video Data: Results from Experimental Research into Pedestrian Walking Behavior', Transportation Research Board Annual Meeting, Washington, DC.
- Hoseini, M \& Vaziri, M 2006, 'Using Image Processing for Microsimulation of Freeways in Developing Countries', Transportation Research Board Annual Meeting, Washington, DC.
- Hoxie, R \& Rubenstein, L 1994, 'Are Older Pedestrians Allowed enough Time to Cross Intersections Safely?', Journal of the American Geriatrics Society, vol 42, no. 3, pp. 241-244.
- Hsieh, J-W, Yu, S-H, Chen, Y-S \& Hu, W-F 2006, 'Automatic Traffic Surveillance System for Vehicle Tracking and Classification', IEEE Transactions on Intelligent Transportation Systems, vol 7, no. 2, pp. 175187.
- Hua, J, Gutierrez, N, Banerjee, I, Markowitz, F \& Ragland, D 2009, 'San Francisco PedSafe II Project Outcomes and Lessons Learned', Transportation Research Board Annual Meeting, Washington, DC.
- Hubbard, S, Bullock, D \& Day, C 2008, 'Opportunities to Leverage Existing Infrastructure to Integrate Real-Time Pedestrian Performance Measures into Traffic Signal System Infrastructure', Transportation Research Board Annual Meeting, Washington, DC.
- Hui, X, Jian, L, Xiaobei, J \& Zhenshan, L 2007, 'Pedestrian Walking Speed, Step Size, and Step Frequency from the Perspective of Gender and Age: Case Study in Beijing, China', Transportation Research Board Annual Meeting, Washington, DC, Source Data: Transportation Research Board Annual Meeting 2007 Paper No. 07-1486.
- Hupfer, C 1997, 'Deceleration to Safety Time (DST)-a Useful Figure to Evaluate Traffic Safety', ICTCT, Department of Traffic Planning and Engineering, Lund University, Lund, Sweden.
- Hu, H, Qu, Z, Li, Z \& Wang, D 2008, 'Research on Recognition and Classification of Moving Objects in Mixed Traffic Based on Video Detection', Transportation Research Board Annual Meeting, Washington, DC.
- Hu, Z \& Tsai, Y 2009, 'Homography-Based Vision Algorithm for Traffic Sign Attribute Computation', Computer-aided Civil and Infrastructure Engineering, vol 24, no. 6, pp. 385-400.
- Hu, W, Xiao, X, Xie, D, Tan, T \& Mayban, S 2004, 'Traffic Accident Prediction Using 3D Model Based Vehicle Tracking', IEEE Transactions on Vebicular Tecbnology, vol 53, no. 3, p. 677-694.
- Hu, W, Xie, D \& Tan, T 2004, 'A Hierarchical Self-organizing Approach for Learning the Patterns of Motion Trajectories', IEEE Transactions on Neural Networks, vol 15, no. 1, pp. 135-144.
- Huybers, S, Van Houten, R \& Malenfant, J 2004, 'Reducing Conflicts between Motor Vehicles and Pedestrians: the Separate and Combined Effects of Pavement Markings and a Sign Prompt', Journal of Applied Behavior Analysis, vol 37, no. 4, pp. 445-456.
- Hydén, C 1987, 'The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique', Department of Traffic Planning and Engineering, Lund University, Lund, Sweden.
- Hydén, C, Garder, P \& Linderholm, L 1982, 'An Updating of the Use of Further Development of the Traffic Conflict Technique'.
- ICBC 2006, 'British Columbia Traffic Collision Statistics', Annual Report, Insurance Corporation of British Columbia.
- Ismail, K, Sayed, T \& Saunier, N 2009, 'Automated Collection of Pedestrian Data Using Computer Vision Techniques', Transportation Research Board Annual Meeting, Washington, DC, 09-1122.
- Ismail, K, Sayed, T \& Saunier, N 2010, 'Automated Analysis of Pedestrian-Vehicle Conflicts: A Context for Before-and-After Studies', (In-print) Transportation Research Record: Journal of the Transportation Research Board.
- Ismail, K, Sayed, T \& Saunier, N 2010, 'Automated Safety Analysis Using Video Sensors: Technology and Case Studies', Canadian Multidisciplinary Road Safety Conference, Niagara Falls, Ontario.
- Ismail, K, Sayed, T, Saunier, N \& Lim, C 2009, 'Automated Analysis of Pedestrian-Vehicle Conflicts Using Video Data', Transportation Research Record: Journal of the Transportation Research Board, vol 2140, pp. 44-54.
- Iwasaki, Y \& Kurogi, Y 2007, 'Real-time Robust Vehicle Detection through the Same Algorithm both Day and Night', International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China.
- Izadpanah, P, Hellinga, B \& Fu, L 2009, 'Automatic Traffic Shockwave Identification Using Vehicles' Trajectories', Transportation Research Board Annual Meeting, Washington, DC.
- James, S \& Walton, C 2000, 'Platoon Pedestrian Movement Analysis: a Case Study Utilizing the Market Street Station in Denver, Colorado', University of Texas, Austin; Southwest Region University Transportation Center.
- Kalinke, T, Tzomakas, C \& Seelen, WV 1998, 'A Texture-based Object Detection and an Adaptive Model-based Classification', IEEE Intelligent Vehicles Symposium, Stuttgart, Germany.
- Kanhere, N, Birchfield, S \& Sarasua, W 2008, 'Automatic Camera Calibration Using Pattern Detection for Vision-Based Speed Sensing', Transportation Research Record: Journal of the Transportation Research Board, vol 2086, pp. 30-39.
- Kanhere, N, Birchfield, S, Sarasua, W \& Whitney, T 2007, 'Real-Time Detection and Tracking of Vehicle Base Fronts for Measuring Traffic Counts and Speeds on Highways', Transportation Research Record: Journal of the Transportation Research Board, vol 1993, pp. 155-164.
- Keall, M 1995, 'Pedestrian Exposure to Risk of Road Accident in New Zealand', Accident Analysis and Prevention, vol 27, no. 5, pp. 729-740.
- Kerridge, J, Armitage, A, Binnie, D, Lei, L \& Sumpter, N 2004, 'Using Low-Cost Infrared Detectors to Monitor Movement of Pedestrians: Initial Findings', Transportation Research Record: Journal of the Transportation Research Board, vol 1878, pp. 11-18.
- Kerridge, J \& Chamberlain, T 2005, 'Collecting, Processing and Calculating Pedestrian Flow Data in Real-time', Transportation Research Board Annual Meeting, Washington, DC.
- Khan, Z, Balch, T \& Dellaert, F 2005, 'MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 27, no. 11, pp. 1805-1918.
- Khanloo, B, Stefanus, F, Ranjbar, M, Li, Z-N, Saunier, N, Sayed, T \& Mori, G 2010, 'Max-Margin Offline Pedestrian Tracking with Multiple Cues', Seventh Canadian Conference on Computer and Robot Vision, Ottawa, Ontario.
- Kim, K, Hallonquist, L, Settachai, N \& Yamashita, E 2007, 'Walking in Waikiki: Measuring Impact of Street Performers on Pedestrian Level of Service in Urban Resort District', Transportation Research Record: Journal of the Transportation Research Board, vol 1982, pp. 104-112.
- Kim, S, Oh, S, Kim, K, Park, S, Park, K \& Kim, J 2005, 'Front and Rear Vehicle Detection and Tracking in the Day and Night Time Using Vision and Sonar Sensors', World Congress on Intelligent Transport Systems, San Francisco, CA.
- Kim, K \& Teng, H 2004, 'Time Separation Control between Pedestrians and Turning Vehicles at Intersections', Transportation Research Board Annual Meeting, Washington, DC.
- Knoblauch, R, Pietrucha, M \& Nitzburg, M 1996, 'Field Studies of Pedestrian Walking Speed and Startup Time', Transportation Research Record: Journal of the Transportation Research Board, vol 1538, pp. 27-38.
- Kovvali, V, Alexiadis, V \& Zhang, L 2007, 'Video-based Vehicle Trajectory Data Collection', Transportation Research Board Annual Meeting, Washington, DC.
- Krotosky, S \& Trivedi, M 2007, 'On Color-, Infrared-, and Multimodal-Stereo Approaches to Pedestrian Detection', IEEE Transactions on Intelligent Transportation Systems, vol 8, no. 4, pp. 619-629.
- Kyte, M, Abdel-Rahim, A, Dixon, M, Li, J-M \& Urbanik, T 2009, 'Distance Gap as a Detection Design and Operations Tool', Transportation Research Board Annual Meeting, Washington, DC.
- Lam, W \& Cheung, C 2000, 'Pedestrian Speed Flow Relationships for Walking Facilities in Hong Kong', Journal of Transportation Engineering, vol 126, no. 4, pp. 343-349.
- Lam, W, Morrall, J \& Ho, H 1995, 'Pedestrian Flow Characteristics in Hong Kong', Transportation Research Record: Journal of the Transportation Research Board, vol 1487, pp. 56-62.
- LaPlante, JN \& Kaeser, TP 2004, 'The Continuing Evolution of Pedestrian Walking Speed Assumptions', Institute of Transportation Engineers Journal, vol 74, pp. 32-40.
- Laureshyn, A 2010, 'Application of Automated Video Analysis to Road User Behaviour', PhD Thesis, Faculty of Engineering, Lund University, Lund Sweden.
- Laureshyn, A \& Ardö, H 2006, 'Automated Video Analysis as a Tool for Analysing Road User Behaviour', IET Intelligent Transport Systems, vol 3, no. 3, pp. 345-357.
- Lee, J \& Lam, W 2006, 'Variation of Walking Speeds on a Unidirectional Walkway and on a Bidirectional Stairway', Transportation Research Record: Journal of the Transportation Research Board, vol 1982, pp. 122-131.
- Leibe, B, Cornelis, N, Cornelis, K \& Gool, L 2007, 'Dynamic 3D Scene Analysis from a Moving Vehicle', IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, Minnesota.
- Litman, T 2003, Non-Motorized Transportation Demand Management, Sustainable Transport: Planning for Walking and Cycling in Urban Environments, Woodhead Publishing Ltd.
- Liu, M, Wu, C \& Zhang, Y 2008, 'A Review of Traffic Visual Tracking Technology', International Conference on Audio, Language and Image Processing, Shanghai, China.
- Li, H, Zhang, K \& Jiang, T 2004, 'Minimum Entropy Clustering and Applications to Gene Expression Analysis', IEEE Computational Systems Bioinformatics Conference, Stanford, CA.
- Li, Y, Zhu, F, Ai, Y \& Wang, F 2007, 'On Automatic and Dynamic Camera Calibration based on Traffic Visual Surveillance', Intelligent Vehicles Symposium, IEEE, Istanbul.
- Lloyd, S 1982, 'Least Squares Quantization in PCM', IEEE Transactions on Information Theory, vol 28, no. 2, pp. 129-137.
- Lord, D 1996, 'Analysis of Pedestrian Conflicts with Left-turning Traffic', Transportation Research Record, Journal of the Transportation Research Board, vol 1538, pp. 61-67.
- Lord, D, Washington, S \& Ivan, J 2005, 'Poisson, Poisson-gamma and Zero-inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory', Accident Analysis and Prevention, vol 37, no. 1, pp. 35-46.
- Lou, J, Tan, T, Hu, W, Yang, H \& Maybank, S 2005, '3-D Model-based Vehicle Tracking', IEEE Transactions on Image Processing, vol 14, no. 10, pp. 1561-1569.
- Lowe, D 2004, 'Distinctive Image Features from Scale-invariant Keypoints', International Journal of Computer Vision, vol 60, no. 2, pp. 91-110.
- Lucas, B \& Kanade, T 1981, 'An Iterative Image Registration Technique with an Application to Stereo Vision', 7th International Joint Conference on Artificial Intelligence, Vancouver, BC.
- Wei, G \& Ma, S 1994, 'Implicit and Explicit Camera Calibration: Theory and Experiments', IEEE Transactions on Pattern Analysis and Macbine Intelligence, vol 16, no. 5, pp. 469-480.
- Malinovskiy, Y, Wu, Y \& Wang, Y 2008, 'Video-based Monitoring of Pedestrian Movements at Signalized Intersections', Transportation Research Record: Journal of the Transportation Research Board, vol 2073, pp. 11-17.
- Malinovskiy, Y, Zheng, J \& Wang, Y 2008, 'Model-Free Video Detection and Tracking of Pedestrians and Bicyclists', Computer-Aided Civil and Infrastructure Engineering, vol 24, no. 3, pp. 157-168.
- Malkhamah, S, Miles, T \& Montgomery, F 2005, 'The Development of an Automatic Method of Safety Monitoring at Pelican Crossings', Accident Analysis and Prevention, vol 37, no. 5, pp. 938-946.
- Marquardt, D 1963, 'An Algorithm for Least-Squares Estimation of Nonlinear Parameters', SLAM Journal on Applied Mathematics, vol 11, no. 2, pp. 431-441.
- Martel, L \& Malenfant, É 2006, 'Portrait of the Canadian Population in 2006, by Age and Sex: Findings', Demography Division, Statistics Canada.
- Masoud, O \& Papanikolopoulos, N 2007, 'Using Geometric Primitives to Calibrate Traffic Scenes', Transportation Research Part C, vol 15, no. 6, pp. 361-379.
- Mathers, C \& Loncar, D 2005, 'Updated Projections of Global Mortality and Burden of Disease, 20022030: Data Sources, Methods and Results.', World Health Organization.
- Mathworks 2010, 'MATLAB® - The Language of Technical Computing. © 1994-2010 The MathWorks, Inc.'.
- McClurg, A 1995, 'Bringing Privacy Law out of the Closet: A Tort Theory of Liability for Intrusions in Public Places', N.C. L. Review, vol 73, p. 989.
- Medina, J, Benekohal, R \& Chitturi, M 2009, 'Changes in Video Detection Performance at Signalized Intersections over Different Illumination Conditions', Transportation Research Record: Journal of the Transportation Research Board, vol 2129, pp. 111-120.
- Medina, J, Benekohal, R \& Wang, M-H 2008, 'In-Street Pedestrian Crossing Signs and Effects on Pedestrian-Vehicle Conflicts at University Campus Crosswalks', Transportation Research Board Annual Meeting, Washington, DC.
- Mehmood, A, Saccomanno, F \& Hellinga, B 2001, 'Simulation of Road Crashes by Use of Systems Dynamics', Transportation Research Record: Journal of the Transportation Research Board, vol 1746, pp. 37-46.
- Michalopoulos, P 1991, 'Vehicle Detection Video through Image Processing: The Autoscope System', IEEE Transactions on Vehicular Technology, vol 40, no. 1, pp. 21-29.
- Middleton, D, Park, E, Charara, H \& Longmire, R 2008, 'Evidence of Unacceptable Video Detector Performance for Dilemma Zone Protection', Transportation Research Record: Journal of the Transportation Research Board, vol 2080, pp. 100-110.
- Migletz, D, Glauz, W \& Bauer, K 1985, 'Relationships between Traffic Conflicts and Accidents', FHWA/RD-84/042, US Department of Transportation, Federal Highway Administration.
- Milam, R \& Mitchell, C 2008, 'Conventional Level of Service Analysis, Thresholds, and Policies Get a Failing Grade', Transportation Research Board Annual Meeting, Washington, DC.
- Minderhoud, M \& Bovy, P 2001, 'Extended Time-to-collision Measures for Road Traffic Safety Assessment', Accident Analysis and Prevention, vol 33, no. 1, pp. 89-97.
- Mohan, A, Papageorgiou, C \& Poggio, T 2001, 'Example-based Object Detection in Images by Components', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 23, no. 4, pp. 349-361.
- Montufar, J, Michelle, J \& Nakagawa, S 2007, 'Pedestrians’ Normal Walking Speed and Speed When Crossing a Street', Transportation Research Record: Journal of the Transportation Research Board, vol 2002, pp. 90-97.
- Moon, H, Chellappa, R \& Rosenfeld, A 2002, 'Optimal Edge-based Shape Detection', IEEE Transactions on Image Processing, vol 11, no. 11, pp. 1209-1227.
- Morrall, J, Ratnayake, L \& Seneviratne, P 1991, 'Comparison of Central Business District Pedestrian Characteristics in Various Areas', Transportation Research Record: Journal of the Transportation Research Board, vol 1294, pp. 57-61.
- Muhlrad, N 1993, 'Traffic Conflict Technique and other Forms of Behavioural Analysis: Application to Safety Diagnosis', ICTCT, Salzburg, Austria.
- Munder, S \& Gavrila, DM 2006, 'An Experimental Study on Pedestrian Classification', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 28, no. 11, pp. 1863-1868.
- Nelder, J \& Mead, R 1965, 'A Simplex Method for Function Minimization', Computer Journal, vol 7, no. 4, p. 308-313.
- O’Kelly, M, Matisziw, T, Li, R, Merry, C \& Niu, X 2005, 'Identifying Truck Correspondence in Multiframe Imagery', Transportation Research Part C: Emerging Technologies, vol 13, no. 1, pp. 1-17.
- Odero, W, Garner, P \& Zwi, A 1997, 'Road Traffic Injuries in Developing Countries: A Comprehensive Review of Epidemiological Studies', Tropical Medicine and International Health, vol 2, no. 5, pp. 445-460.
- Odobez, J \& Bouthemy, P 1995, 'Robust Multiresolution Estimation of Parametric Motion Models', Journal of Visual Communication and Image Representation, vol 6, no. 4, pp. 348-365.
- Oh, J, Kim, E, Kim, M \& Choo, S 2010, 'Development of Conflict Techniques for Left-turn and Crosstraffic at Protected Left-turn Signalized Intersections', Safety Science, vol 48, no. 4, pp. 460-468.
- Ossen, S \& Hoogendoorn, S 2005, 'Car-following Behavior Analysis from Microscopic Trajectory Data', Transportation Research Record: Journal of the Transportation Research Board, vol 1934, pp. 13-21.
- Pang, C, Lam, W \& Yung, N 2004, 'A Novel Method for Resolving Vehicle Occlusion in a Monocular Traffic-image Sequence', IEEE Transactions on Intelligent Transportation Systems, vol 5, no. 3, pp. 129-141.
- Papageorgiou, C \& Poggio, T 2000, 'A Trainable System for Object Detection', International Journal of Computer Vision, vol 38, no. 1, pp. 15-33.
- Paragios, N \& Deriche, R 2000, 'Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects', IEEE Transactions on Pattern Analysis and Macbine Intelligence, vol 22, no. 3, pp. 266-280.
- Paragios, N \& Tziritas, G 1999, 'Adaptive Detection and Localization of Moving Objects in Image Sequences', Signal Processing and Image Communication, vol 14, no. 4, pp. 277-296.
- Parker, M \& Zegeer, C 1989, 'Traffic Conflict Techniques for Safety and Operations Observer's Manual', US Department of Transportation, Federal Highway Administration, FHWA-IP-88-027, Federal Highway Administration, McLean, Virginia.
- Pauls, J 2008, 'Demographic Changes Leading to Deterioration of Pedestrian Capabilities Affecting Safety and Crowd Movement', Transportation Research Board Annual Meeting, Washington, DC, Source Data: Transportation Research Board Annual Meeting 2007 Paper No. 08-1029.
- Pengfei, CZAS 2004, 'Efficient Method for Camera Calibration in Traffic Scenes', Electronic Letters, vol 40, no. 6, pp. 1-2.
- Perkins, R \& Harris, J 1968, 'Traffic Conflict Characteristics: Accident Potential at Intersections', Highway Research Record, vol 225, pp. 35-43.
- Persaud, B \& Lyon, C 2007, 'Empirical Bayes Before-after Safety Studies: Lessons Learned from Two Decades of Experience and Future Directions', Accident Analysis and Prevention, vol 39, no. 3, pp. 546-555.
- Peterfreund, N 1999, 'Robust Tracking of Position and Velocity with Kalman Snakes', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 21, no. 6, pp. 564-569.
- Phong, T, Horaud, R, Yassine, A \& Tao, P 2005, 'Object Pose from 2-D to 3-D Point and Line Correspondences', International Journal of Computer Vision, vol 15, no. 3, pp. 225-243.
- Prevedouros, P, Ji, X, Papandreou, K, Kopelias, P \& Vegiri, V 2006, 'Video Incident Detection Tests in Freeway Tunnels', Transportation Research Record: Journal of the Transportation Research Board, vol 1959, pp. 130-139.
- Pulugurtha, S \& Repaka, S 2008, 'Assessment of Models to Measure Pedestrian Activity at Signalized Intersections', Transportation Research Record: Journal of the Transportation Research Board, vol 2073, pp. 39-48.
- Qi, Y, Tang, Y \& Smith, B 2006, 'Video-Based Automated Identification of Freeway Shoulder Events', Transportation Researcb Record: Journal of the Transportation Research Board, vol 1972, pp. 46-53.
- Ramanan, D, Forsyth, D \& Zisserman, A 2005, 'Strike a Pose: Tracking People by Finding Stylized Poses', Computer Vision and Pattern Recognition, San Diego, CA.
- Ran, Y, Chellappa, R \& Zheng, Q 2006, 'Finding Gait in Space and Time', 18th International Conference on Pattern Recognition, IEEE Computer Society, Hong Kong.
- Remondino, F \& Fraser, C 2006, 'Digital Camera Calibration Methods: Consideration and Comparisons', International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol 36, no. B5, pp. 266272.
- Rodriguez-Seda, J, Benekohal, R \& Morocoima-Black, R 2008, 'Pedestrian, Bicycle, and Vehicle Traffic Conflict Management in Big Ten University Campuses', Transportation Research Board Annual Meeting, Washington, DC.
- Ronold, K \& Christensen, C 2001, 'Optimization of a Design Code for Wind-turbine Rotor Blades in Fatigue', Engineering Structures, vol 23, no. 8, pp. 993-1004.
- Sabzmeydani, P \& Mori, G 2007, 'Detecting Pedestrians by Learning Shapelet Features', Computer Vision and Pattern Recognition, Minneapolis, Minnesota.
- Saunier, N \& Sayed, T 2006, 'A Feature-based Tracking Algorithm for Vehicles in Intersections', The 3rd Canadian Conference on Computer and Robot Vision, IEEE.
- Saunier, N \& Sayed, T 2007, 'Automated Road Safety Analysis Using Video Data', Transportation Research Record: Journal of the Transportation Research Board, vol 2019, pp. 57-64.
- Saunier, N \& Sayed, T 2008, 'A Probabilistic Framework for Automated Analysis of Exposure to Road Collisions', Transportation Research Record: Journal of the Transportation Research Board, vol 2083, pp. 96-104.
- Saunier, N, Sayed, T \& Ismail, K 2009, 'An Object Assignment Algorithm for Tracking Performance Evaluation', Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2009).
- Saunier, N, Sayed, T \& Ismail, K 2010, 'Large Scale Automated Analysis of Vehicle Interactions and Collisions', Transportation Research Record: Journal of the Transportation Research Board, 10-4059. Recommended for publication.
- Saunier, N, Sayed, T \& Lim, C 2007, 'Probabilistic Collision Prediction for Vision-based Automated Road Safety Analysis', 10th International IEEE Conference on Intelligent Transportation Systems, Seattle, WA.
- Sayed, T \& Zein, S 1999, 'Traffic Conflict Standards for Intersections', Transportation Planning and Technology, vol 22, no. 4, pp. 309-323.
- Schoepflin, T \& Dailey, D 2003, 'Dynamic Camera Calibration of Roadside Traffic Management Cameras for Vehicle Speed Estimation', IEEE Transactions on Intelligent Transportation Systems, vol 4, no. 2, pp. 90-98.
- Shankar, V, Ulfarsson, G, Pendyala, R \& Nebergall, M 2003, 'Modeling Crashes Involving Pedestrians and Motorized Traffic', Safety Science, vol 41, no. 7, pp. 627-640.
- Shannon, C 1948, 'A Mathematical Theory of Communication', Bell System Techical Journal, vol 27, pp. 379-423.
- Shelby, S, Bullock, D, Gettman, D, Ghaman, R, Sabra, Z \& Soyke, N 2008, 'An Overview and Performance Evaluation of ACS Lite - A Low Cost Adaptive Signal Control System', Transportation Research Board Annual Meeting, Washington, DC.
- Shi, J, Chen, Y, Ren, F \& Rong, J 2007, 'Research on Pedestrian Behavior and Traffic Characteristics at Unsignalized Midblock Crosswalk', Transportation Research Record: Journal of the Transportation Research Board, vol 2038, pp. 23-33.
- Shinar, D 1984, 'The Traffic Conflict Technique: A Subjective vs. Objective Approach', Journal of Safety Research, vol 15, no. 4, pp. 153-157.
- Shrestha, L 2006, 'The Changing Demographic Profile of the United States', Congressional Research Service. The Library of Congress, National Council for Science and Environment.
- Sidenbladh, H \& Black, M 2003, 'Learning the Statistics of People in Images and Video', International Journal of Computer Vision, vol 54, no. 1-3, pp. 183-209.
- Sminchisescu, C \& Triggs, B 2001, 'Covariance Scaled Sampling for Monocular 3D Body Tracking', IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI.
- Smith, R \& Knowles, E 1979, 'Affective and Cognitive Mediators of Reactions to Spatial Invasions', Journal of Experimental Social Psychology, vol 15, no. 4, pp. 437-452.
- Solove, D 2006, 'A Taxonomy of Privacy', University of Pennsylvania Law Review, vol 154, pp. 477-570.
- Songchitruksa, P \& Tarko, A 2006, 'Practical Method for Estimating Frequency of Right-angle Collisions at Traffic Signals', Transportation Research Record: Journal of the Transportation Research Board, vol 1953, pp. 89-97.
- Songchitruksa, P \& Tarko, A 2006, 'The Extreme Value Theory Approach to Safety Estimation', Accident Analysis and Prevention, vol 38, no. 4, pp. 811-822.
- Sperling, D \& Gordon, D 2008, 'Two Billion Cars: Tranforming a Culture', TR News, vol 259, pp. 3-9.
- Stauffer, C \& Grimson, W 1999, 'Adaptive Background Mixture Models for Real-time Tracking', IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY.
- Stenger, B, Thayananthan, A, Torr, P \& Cipolla, R 2006, 'Model-based Hand Tracking Using a Hierarchical Bayesian Filter', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 28, no. 9, pp. 1372-1384.
- Stollof, E, McGee, H \& Eccles, K 2007, 'Pedestrian Signal Safety for Older Persons', AAA Foundation for Traffic Safety.
- Struckman-Johnson, D, Lund, A, Williams, A \& Osborne, D 1989, 'Comparative Effects of Driver Improvement Programs on Crashes and Violations', Accident Analysis and Prevention, vol 21, no. 3, pp. 203-215.
- Sturm, P 1997, 'Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction', IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico.
- Sun, C \& Rescot, R 2008, 'Feasibility of an Automated Solution for Deriving Roundabout Turning Movements', Transportation Research Board Annual Meeting, Washington, DC.
- Sun, C, Ritchie, S, Tsai, K \& Jayakrishnan, R 1999, 'Use of Vehicle Signature Analysis and Lexicographic Optimization for Vehicle Reidentification on Freeways', Transportation Research Part C: Emerging Tecbnologies, vol 7, no. 4, pp. 167-185.
- Suri, S \& Parr, M 2004, 'The Hidden Epidemic-War on Roads', Indian Journal of Critical Care Medicine, vol 8, no. 2, p. 69-72.
- Svensson, $\AA 1992$, 'Vidareutveckling och Validering av den Svenska Konflikttekniken (in Swedish). Further Development and Validation of the Swedish Traffic Conflicts Technique.', Department of Traffic Planning and Engineering, Lund University.
- Svensson, $\AA 1998$, 'A Method for Analysing the Traffic Process in a Safety Perspective', PhD Thesis, Department of Traffic Planning and Engineering, Lund University.
- Svensson, A\& Hydén, C 2006, 'Estimating the Severity of Safety Related Behaviour', Accident Analysis and Prevention, vol 38, no. 2, pp. 379-385.
- Szarvas, M, Yoshizawa, A, Yamamoto, M \& Ogata, J 2005, 'Pedestrian Detection with Convolutional Neural Networks', IEEE Symposium on Intelligent Vehicles, Las Vegas, US.
- Tan, T \& Baker, K 2000, 'Efficient Image Gradient based Vehicle Localization', IEEE Transactions on Image Processing, vol 9, no. 8, pp. 1343-1356.
- Tarawneh, M 2001, 'Evaluation of Pedestrian Speed in Jordan with Investigation of Some Contributing Factors', Journal of Safety Research, vol 32, no. 2, pp. 229-236.
- Thiemann, C, Treiber, M \& Kesting, A 2008, 'Estimating Acceleration and Lane-Changing Dynamics from Next Generation Simulation Trajectory Data', Transportation Research Record: Journal of the Transportation Research Board, vol 2088, pp. 90-101.
- Tiwari, G, Mohan, D \& Fazio, J 1998, 'Conflict Analysis for Prediction of Fatal Crash Locations in Mixed Traffic', Accident Analysis and Prevention, vol 30, no. 2, pp. 207-215.
- Topp, H 1998, 'Traffic Safety Work with Video-Processing', University Kaiserslautern.
- Tourinho, L \& Pietrantonio, H 2003, 'A Decision-based Criterion for Selecting Parameters in the Evaluation of Pedestrian Safety Problems with the Traffic Conflict Analysis Technique', Transportation Planning and Technology, vol 29, no. 3, pp. 183-216.
- Transport Canada 2004, 'Pedestrian Fatalities and Injuries, 1992-2001', Road Safety Fact Sheet, Road Safety and Motor Vehicle Regulation, Transport Canada, RS-2004-01E.
- Transportation Association of Canada 2002, 'Manual of Uniform Traffic Control Devices for Canada (MUTCD)'.
- Tsai, R 1987, 'A Versatile Camera Calibration Technique for High-accuracy 3D Machine Vision Metrology Using off-the-shelf TV Cameras and Lenses', IEEE Journal on Robotics and Automation, vol 3, no. 4, pp. 323-344, http://www.cs.cmu.edu/~rgw/TsaiCode.html.
- Tsai, L-W, Hsieh, J-W \& Fan, K-C 2007, 'Vehicle Detection Using Normalized Color and Edge Map', IEEE Transactions on Image Processing, vol 16, no. 3, pp. 850-864.
- Tsai, Y, Wu, J \& Wang, Z 2010, 'Horizontal Roadway Curvature Computation Algorithm Using Vision Technology', Computer-Aided Civil and Infrastructure Engineering, vol 25, no. 2, pp. 78-88.
- Turner, S, Fitzpatrick, K, Brewer, M \& Park, E 2006, 'Motorist Yielding to Pedestrians at Unsignalized Intersections: Findings from a National Study on Improving Pedestrian Safety', Transportation Research Record: Journal of the Transportation Research Board, vol 1982, pp. 1-12.
- UN 2003, 'Global Road Safety Crisis: Report of the Secretary-General', General, A/58/228, United Nations General Assembly.
- USDOT 2008, 'National Transportation Statistics 2008', U.S. Department of Transportation : Bureau of Transportation Statistics, Washington, DC.
- Van Houten, R, Malenfant, J, Houten, J \& Retting, R 1997, 'Using Auditory Pedestrian Signals to Reduce Pedestrian and Vehicle Conflicts', Transportation Research Record: Journal of the Transportation Research Board, vol 1578, pp. 20-22.
- Vanderbilt, T 2008, Traffic: Why We Drive the Way We Do (and What It Says About Us), 1st edn, Alfred A. Knopf, Random House, New York, Toronto, p. 64.
- Várhelyi, A 1996, 'Dynamic Speed Adaptation Based on Information Technology - A Theoretical Background', Lund University, Lund, Sweden.
- Vaziri, B 1998, 'Exclusive Pedestrian Phase for the Business District Signals in Beverly Hills: 10 Years Later', Institute of Transportation Engineers District 6 Meeting, Washington, DC.
- Veeraraghavan, H, Masoud, O \& Papanikolopoulos, N 2003, 'Computer Vision Algorithms for Intersection Monitoring', IEEE Transactions on Intelligent Transportation Systems, vol 4, no. 2, pp. 78-89.
- Velastin, S, Boghossian, B \& Vicencio-Silva, M 2006, 'A Motion-based Image Processing System for Detecting Potentially Dangerous Situations in Underground Railway Stations', Transportation Research Part C: Emerging Technologies, vol 14, no. 2, pp. 96-113.
- Viola, P, Jones, M \& Snow, D 2003, 'Detecting Pedestrians Using Patterns of Motion and Appearance', International Conference on Computer Vision, IEEE Computer Society, Nice, France.
- Virkler, M \& Elayadaph, S 1994, 'Pedestrian Speed-flow-density Relationships', Transportation Research Record: Journal of the Transportation Research Board, vol 1438, pp. 51-58.
- Vlachos, M, Kollios, G \& Gunopulos, D 2005, 'Elastic Translation Invariant Matching of Trajectories', Machine Learning, vol 58, no. 2/3, pp. 301-334.
- Vodden, K, Smith, D, Eaton, F \& Mayhew, D 2007, 'Analysis and Estimation of the Social Cost of Motor Vehicle Collisions in Ontario', Final Report, Transport Canada.
- Walmsley, D \& Lewis, G 1989, 'The Pace of Pedestrian Flows in Cities', Environment and Behavior, vol 21, no. 2, pp. 123-150.
- Wang, K, Hou, Z \& Gong, W 2010, 'Automated Road Sign Inventory System Based on Stereo Vision and Tracking', Computer-Aided Civil and Infrastructure Engineering, pp. 1-10, in-print.
- Wang, C-C \& Lien, J-J 2008, 'Automatic Vehicle Detection Using Local Features - A Statistical Approach', IEEE Transactions on Intelligent Transportation Systems, vol 9, no. 1, pp. 83-96.
- Wang, G, Xiao, D \& Gu, J 2008, 'Review on Vehicle Detection based on Video for Traffic Surveillance', IEEE International Conference on Automation and Logistics, Qingdao, China.
- Weinstein, A \& Schimek, P 2005, 'How Much do Americans Walk? An Analysis of The 2001 NHTS', Transportation Research Board 84th Annual Meeting, Washington, D.C.
- Weng, J, Cohen, P \& Herniou, M 1992, 'Camera Calibration with Distortion Models and Accuracy Evaluation', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 14, no. 10, pp. 965-980.
- WHO 2004, 'The Global Burden of Disease: 2004 Update', World Health Organization.
- WHO-UNEP 2007, 'Climate Change 2007: Mitigation of Climate Change', Intergovermental Panel on Climate Change. IPCC Fourth Assessment Report, Working Group III Summary for Policymakers, Intergovernmental Panel on Climate Change.
- Williams, M 1981, 'Validity of the Traffic Conflicts Technique', Accident Analysis and Prevention, vol 13, no. 2, pp. 133-145.
- Willis, A, Gjersoe, N, Havard, C, Kerridge, J \& Kukla, R 2004, 'Human Movement Behaviour in Urban Spaces: Implications for the Design and Modelling of Effective Pedestrian Environments', Environment and Planning B: Planning and Design, vol 31, no. 6, p. 805-828.
- Woehler, C \& Anlauf, J 1999, 'An Adaptable Time-Delay Neural-Network Algorithm for Image Sequence Analysis', IEEE Transactions on Neural Networks, vol 10, no. 6, pp. 1531-1536.
- Worrall, A, Sullivan, G \& Baker, K 1994, 'A Simple, Intuitive Camera Calibration Tool for Natural Images', 5 th British Machine Vision Conference, The University of Birmingham.
- Wu, B \& Nevatia, R 2007, 'Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors', International Journal of Computer Vision, vol 75, no. 2, pp. 247-266.
- Wu, J \& Tsai, Y 2006, 'Enhanced Roadway Inventory Using a 2-D Sign Video Image Recognition Algorithm', Computer-Aided Civil and Infrastructure Engineering, vol 21, no. 5, pp. 369-382.
- Wu, Y \& Yu, T 2006, 'A Field Model for Human Detection and Tracking', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 28, no. 5, pp. 753-765.
- Xin, W, Hourdos, J, Michalopoulos, P \& Davis, G 2008, 'The Less-than-perfect Driver: A Model of Collision-Inclusive Car-following Behavior', Transportation Research Record: Journal of the Transportation Research Board, vol 2088, pp. 126-137.
- Xu, F, Liu, X \& Fujimura, K 2005, 'Pedestrian Detection and Tracking with Night Vision', IEEE Transactions on Intelligent Transportation Systems, vol 6, no. 1, pp. 63-71.
- Ye, J, XiaoHong, C, Yang, C \& Wu, J 2008, 'Walking Behavior and Pedestrian Flow Characteristics for Different Types of Walking Facilities', Transportation Research Record: Journal of the Transportation Research Board, vol 2048, pp. 43-51, Source Data: Transportation Research Board Annual Meeting 2007 Paper No. 08-1991.
- Yu, X, Jiang, N, Cheong, L-F, Leong, H \& Yan, X 2009, 'Automatic Camera Calibration of Broadcast Tennis Video with Applications to 3D Virtual Content Insertion and Ball Detection and Tracking', Computer Vision and Image Understanding, vol 113, no. 5, pp. 643-652, Computer Vision Based Analysis in Sport Environments.
- Zhang, Z 1994, 'Estimating Motion and Structure from Comspondences of Line Segments between two Perspective Images', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 17, no. 12, pp. 11291139.
- Zhang, Z 2000, 'A Flexible New Technique for Camera Calibration', IEEE Transactions on Paterrn Analaysis and Machine Intelligence, vol 22, no. 11, pp. 1330-1334.
- Zhang, L \& Prevedouros, P 2003, 'Signalized Intersection Level of Service Incorporating Safety Risk', Transportation Research Record, Journal of the Transportation Research Board, vol 1852, pp. 77-86.
- Zhang, L, Wu, B \& Nevatia, R 2007, 'Detection and Tracking of Multiple Humans with Extensive Pose Articulation', IEEE 11th International Conference on Computer Vision, Janeiro, Brazil.
- Zhao, T \& Nevatia, R 2004, 'Tracking Multiple Humans in Complex Situations', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 26, no. 9, pp. 1208-1221.
- Zhou, J, Gao, D \& Zhang, D 2007, 'Moving Vehicle Detection for Automatic Traffic Monitoring', IEEE Transactions on Vehicular Technology, vol 56, no. 1, pp. 51-59.
- Zhu, Q, Yeh, M-C, Cheng, K-T \& Avidan, S 2006, 'Fast Human Detection Using a Cascade of Histograms of Oriented Gradients', IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Setúbal, Portugal.


## Appendices

## Appendix A

Table A. 1 Summary of video data

| Video Sequence | Location | Video Sensor | Context | Duration |
| :---: | :---: | :---: | :---: | :---: |
| $\begin{aligned} & \text { BR-1, BR-2, ..., } \\ & \text { BR-7 } \end{aligned}$ | Traffic intersection at Broughton and Robson streets in the Downtown area of Vancouver, BC | High-definition digital camera setup on the Empire Landmark high-rise building. | Inbound and outbound crowd movement to and from the Celebration of Light local event. | 2 hours |
| PG-1 and PG-2 | Traffic intersection between Pender and W. Georgia streets in the Downtown area of Vancouver, BC | Two video cameras with external storage. | Two Crosswalks with conflicting left-turn vehicular traffic. | $\begin{aligned} & 2 \times 2 \\ & \text { days } \end{aligned}$ |
| Cairo-1 to <br> Cairo-10 | Several intersections in Cairo, Egypt. | Hand-held video camera. | Typical traffic operations in a driving culture significantly different from North America. | 1 hour |
| Berk-1 to Berk-10 | 8th street and Webster, Oakland, California. | Mounted analog camera. Tape digitization has been performed. | Pre-scramble and postscramble phase road user movement. | $\begin{aligned} & 10 \times 1 \\ & \text { hour } \end{aligned}$ |
| KW1 to KW9 | Darwaza intersection, Kuwait City, Kuwait | Mounted, standard definition digital camera. | Mixed traffic navigating a busy signalized fourleg intersection. | 24 hours |
| COV-1-6 | W. Georgia street, Vancouver, BC | Mounted surveillance camera | Cross walk on Georgia street on Nov 252006 during a Rolling Stones Concert. | 10 hours |
| Ed-2 Ed-3 | Yellowhead and Victorial Trail, Edmonton, AB | Before-and-after analysis | Right-turn movement at a signalized intersection | 8 hours |

## Appendix B

Table B. 1 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index at each frame

| Conflict <br> Indicator | Mean |  | Individual Index Value |  |  |  | Individual Aggregation |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Function |  | Distribution |  | Function |  | Distribution |  |
| TTC | 5.43 | -3.39 | 0.50 | -0.15 | 0.29 | -0.19 |  |  |  |  |
|  |  | 2.93 |  | 0.26 |  | 0.48 |  |  |  |  |
| PET+ | 6.48 | -4.24 | 0.39 | -0.29 | 0.56 | -0.34 |  |  |  |  |
|  |  | 4.17 |  | 0.49 |  | 0.39 |  |  |  |  |
| PET- | -6.05 | -2.99 | 0.35 | -0.26 | 0.53 | -0.27 |  |  |  |  |
|  |  | 3.07 |  | 0.44 |  | 0.34 |  |  |  |  |
| DST | 0.28 | -0.63 | 0.03 | -0.02 | 0.36 | -0.36 |  |  |  |  |
|  |  | 0.62 |  | 0.13 |  | 0.37 |  |  |  |  |
| GT+ | 6.77 | -4.64 | 0.36 | -0.21 | 0.53 | -0.24 | $\begin{aligned} & \underset{\sim}{7} \\ & \dot{0} \\ & \ddot{\sim} \\ & \tilde{0} \\ & \dot{i} \\ & \underset{\sim}{n} \\ & 0 \end{aligned}$ |  | 8 <br>  <br> 0 <br> $\vdots$ <br> $\vdots$ <br> $\vdots$ <br> $\vdots$ <br> $i$ |  |
|  |  | 2.79 |  | 0.53 |  | 0.42 |  |  |  |  |
| GT- | -5.49 | -3.26 | 0.43 | -0.32 | 0.60 | -0.31 |  |  |  |  |
|  |  | 3.69 |  | 0.49 |  | 0.36 |  |  |  |  |
| \|PET| | 6.22 | -3.54 | 0.42 | -0.28 | 0.59 | -0.31 | - |  | - |  |
|  |  | 3.51 |  | 0.42 |  | 0.34 |  |  |  |  |
| \|GT| | 6.18 | 6.18 | 0.42 | -0.25 | 0.59 | -0.27 | - |  | - |  |
|  |  | -4.24 |  | 0.48 |  | 0.37 |  |  |  |  |
| Index (function) | 0.36 | -0.18 | - |  | - |  | 0.016 |  | - |  |
|  |  | 0.16 |  |  | -0.30 | 0.45 |  |  |  |  |
| Index <br> (distribution) | 0.53 | -0.18 | - |  |  |  | - |  | - |  | 0.05 |  |
|  |  | 0.17 |  |  | -0.34 | 0.43 |  |  |  |  |

Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

Table B. 2 Summary results for different aggregation strategies for after conditions. Representative statistics are drawn only from calculable values of each indicator or index at each frame


Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

Table B. 3 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index at each frame taking into account frequency of observation


Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

Table B. 4 Summary results for different aggregation strategies for after conditions. Representative statistics are drawn only from calculable values of each indicator or index at each frame taking into account frequency of observation


Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

Table B. 5 Summary results for different aggregation strategies for before conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user taking into account frequency of observation

| Conflict Indicator | Mean |  | Individual Index Value |  |  |  | Individual Aggregation |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Function |  | Distribution |  | Function |  | Distribution |  |
| TTC | 4.85 | -2.80 | 0.54 | -0.14 | 0.35 | -0.20 | $\because$ |  |  |  |
|  |  | 2.52 |  | 0.22 |  | 0.42 |  |  |  |  |
| PET+ | 7.40 | -4.24 | 0.30 | -0.22 | 0.47 | -0.28 |  |  |  |  |
|  |  | 3.88 |  | 0.48 |  | 0.42 |  |  |  |  |
| PET- | -6.37 | -3.28 | 0.31 | -0.25 | 0.50 | -0.27 |  |  |  |  |
|  |  | 3.24 |  | 0.45 |  | 0.36 |  |  |  |  |
| DST | 0.41 | -0.69 | 0.05 | -0.05 | 0.47 | -0.47 |  |  |  |  |
|  |  | 0.69 |  | 0.15 |  | 0.33 |  |  |  |  |
| GT+ | 6.11 | -4.07 | 0.43 | -0.23 | 0.60 | -0.24 |  |  |  |  |
|  |  | 2.59 |  | 0.47 |  | 0.36 |  |  |  |  |
|  |  | -2.43 | 0.46 | -0.29 | 0.63 | -0.25 |  |  |  |  |
|  |  | 3.51 |  | 0.45 |  | 0.33 |  |  |  |  |
| \|PET| | 6.92 | -3.78 | 0.34 | -0.25 | 0.51 | -0.29 | - |  | - |  |
|  |  | 3.76 |  | 0.44 |  | 0.38 |  |  |  |  |
| \|GT| | 5.75 | -3.88 | 0.47 | -0.24 | 0.64 | -0.24 | - |  | - |  |
|  |  | 2.51 |  | 0.44 |  | 0.33 |  |  |  |  |
| Index <br> (function) | 0.33 | -0.19 | - |  | - |  | -0.01 |  |  |  |
|  |  | 0.18 |  |  | -0.29 | 0.44 |  |  |  |  |
| Index <br> (distribution) | 0.51 | -0.22 | - |  |  |  | - |  |  |  | 0.00 |  |
|  |  | 0.20 |  |  |  |  |  |  | -0.37 | 0.42 |

Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

Table B. 6 Summary results for different aggregation strategies for after conditions. Representative statistics are drawn only from calculable values of each indicator or index for each road user taking into account frequency of observation


Values in italic are the $15^{\text {th }}$ percentile value minus the mean and the $85^{\text {th }}$ percentile value minus the mean.

## Appendix C

This thesis is constituted by published work, or papers submitted for review. Chapters 3, 4, 5, and 6 appeared in the following publications:

1. Ismail, K, Sayed, T \& Saunier, N 2009, 'Automated Collection of Pedestrian Data Using Computer Vision Techniques', Transportation Research Board Annual Meeting, Washington, DC, 09-1122.
2. Ismail, K, Sayed, T \& Saunier, N 2010, 'Automated Analysis of Pedestrian-Vehicle Conflicts: A Context for Before-and-After Studies', (In-print) Transportation Research Record: Journal of the Transportation Research Board.
3. Ismail, K, Sayed, T \& Saunier, N 2010, 'Automated Safety Analysis using Video Sensors: Technology and Case Studies', Canadian Multidisciplinary Road Safety Conference, Niagara Falls, Ontario.
4. Ismail, K, Sayed, T, Saunier, N \& Lim, C 2009, 'Automated Analysis of Pedestrian-Vehicle Conflicts Using Video Data', Transportation Research Record: Journal of the Transportation Research Board, vol. 2140, pp. 44-54.
5. Ismail, K, Sayed, T \& Saunier, N 2009, 'Camera Calibration for Urban Traffic Scenes: Practical Issues and a Robust Approach', Transportation Research Board Annual Meeting, Washington, DC, 102715.

[^0]:    ${ }^{1}$ Assuming an annual registration rate of 1.2 million vehicles based on the average from the period 1997-2007. If the rate of increase is otherwise taken, the period for doubling the number of vehicles becomes approximately 7 years based on an average annual growth rate of $10 \%$ calculated for the period 2002-2007.

[^1]:    ${ }^{2}$ This sentence refers to the well-established discipline of intelligent transportation systems.

[^2]:    ${ }^{3} \ldots$ or system safety as described by (Hauer 1982).

[^3]:    ${ }^{4}$ To the best of the author's knowledge, the type of high-precision positional data referred to was first used in (Saunier \& Sayed 2007).

[^4]:    ${ }^{1}$ The precise definition of traffic conflicts will be discussed in subsequent sections.

[^5]:    ${ }^{2}$ Refer to section 2.1.3b for a detailed description of the term "validity".

[^6]:    ${ }^{3}$ The injury referred to here is the rare occurrence of bodily harm to road users avoiding a collision.

[^7]:    ${ }^{4}$ In this thesis, the term road user trajectories, opposed to tracks, is used when referring to the extrapolation of road user positions.

[^8]:    ${ }^{5}$ Processing time shorter than 250 millisecond for every evaluation of a pedestrian image.
    ${ }^{6}$ Number of correct detections (positive detection of pedestrian images and negative detection of nonpedestrian images) per unit time.

[^9]:    ${ }^{7}$ NG-VIDEO: Next Generation Vehicle Interaction and Detection Environment for Operations.

[^10]:    ${ }^{8}$ Delft University of Technology: Department of Transport \& Planning

[^11]:    ${ }^{9}$ Public Law108-495

[^12]:    ${ }^{1}$ Refer to section "Benchmark Evaluation" and evaluation work in (Enzweiler \& Gavrila 2009).

[^13]:    ${ }^{1}$ This number is proportional to the total number of events in the before and after periods.

[^14]:    ${ }^{1}$ Single-user collisions are excluded by this argument mainly because an important focus of this chapter is on pedestrian safety. Pedestrian-involved collisions cannot be of the single-user type.

[^15]:    ${ }^{2}$ As shorthand, "objective conflict indicators" is replaced in this chapter with "conflict indicators".

[^16]:    ${ }^{3}$ In this thesis PET is assumed to be either positive or negative depending on whether the encroaching vehicle passes in front of or after the conflicted pedestrian. This is a slight variation of notation since PET is originally intended to be only positive.

[^17]:    ${ }^{1}$ As shorthand, road user tracks is used to mean the tracks of different road users.

[^18]:    ${ }^{1}$ Back-projection refers to the mapping of various features from image coordinates to world coordinates.

