5-1-2016

Vision-Based Intersection Monitoring: Behavior Analysis & Safety Issues

Mohammad Shokrolah Shirazi
University of Nevada, Las Vegas

VISION-BASED INTERSECTION MONITORING: BEHAVIOR ANALYSIS & SAFETY ISSUES

By

Mohammad Shokrolah Shirazi

Bachelor's Degree in Computer Engineering
Ferdowsi University of Mashhad, Khorasan, Iran
2005

Master's Degree in Computer Engineering
Sharif University of Technology, Tehran, Iran
2007

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy – Electrical Engineering

Department of Electrical and Computer Engineering
Howard R. Hughes College of Engineering
The Graduate College

University of Nevada, Las Vegas
May 2016

UNLV | GRADUATE COLLEGE

Dissertation Approval

The Graduate College
The University of Nevada, Las Vegas

April 19, 2016

This dissertation prepared by

Mohammad Shokrolah Shirazi

entitled

Vision-Based Intersection Monitoring: Behavior Analysis & Safety Issues

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Engineering - Electrical Engineering
Department of Electrical and Computer Engineering

Brendan Morris, Ph.D.
Examination Committee Chair

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean

Pushkin Kachroo, Ph.D.
Examination Committee Member

Emma Regentova, Ph.D.
Examination Committee Member

Venkatesan Muthukumar, Ph.D.
Examination Committee Member

Alexander Paz-Cruz, Ph.D.
Graduate College Faculty Representative

Abstract

The main objective of my dissertation is to provide a vision-based system to automatically understands traffic patterns and analyze intersections. The system leverages the existing traffic cameras to provide safety and behavior analysis of intersection participants including behavior and safety. The first step is to provide a robust detection and tracking system for vehicles and pedestrians of intersection videos. The appearance and motion based detectors are evaluated on test videos and public available datasets are prepared and evaluated. The contextual fusion method is proposed for detecting pedestrians and motion-based technique is proposed for vehicles based on evaluation results. The detections are feed to the tracking system which uses the mutual cooperation of bipartite graph and enhance optical flow. The enhanced optical flow tracker handles the partial occlusion problem, and it cooperates with the detection module to provide long-term tracks of vehicles and pedestrians. The system evaluation shows 13% and 43% improvement in tracking of vehicles and pedestrians respectively when both participants are addressed by the proposed framework. Finally, trajectories are assessed to provide a comprehensive analysis of safety and behavior of intersection participants including vehicles and pedestrians. Different important applications are addressed such as turning movement count, pedestrians crossing count, turning speed, waiting time, queue length, and surrogate safety measurements. The contribution of the proposed methods are shown through the comparison with ground truths for each mentioned application, and finally heat-maps show benefits of using the proposed system through the visual depiction of intersection usage.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor *Dr. Brendan Morris* for his continuous support on my dissertation. I thank him for his caring, motivating, and guidance. His guidance and immense knowledge helped me during the period of research and writing of this dissertation. I appreciate his help.

I would like to thank *Dr. Pushkin Kachroo*, *Dr. Emma Regentova*, *Dr. Venkatesan Muthukumar* and *Dr. Alexander Paz*, for being part of the committee and providing their insightful comments and encouragement. Special thanks to Nevada Department of Transportation (NDOT) for their support and funding my project during the last three semesters of my Ph.D. education.

At last, I would like to express sincere thanks to my parents, specially my mother for encouraging and supporting me towards my higher studies. I would like to thanks all of my friends here in UNLV, who have made my stay at UNLV a memorable one. I thank you for your wonderful company.

Mohammad Shokrolah Shirazi

University of Nevada, Las Vegas
May 2016

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Vehicle drivers and pedestrian safety are the most important transportation concerns in the world and this make intersections the interesting target for monitoring. According to traffic studies, 50 percent of all reported crashes and 22 percent of the total fatalities happened at intersections [9]. Besides safety, behavior analysis of vehicles and pedestrians is useful for intersection design, planning and signal optimization [10, 11].

Behavior and safety analysis are built from the collected data which are usually provided manually by human observations. It is quite obvious that manual monitoring of traffic events from several different screens require high labor cost and special concentration of the traffic operators and there are always possibilities to miss the urgent events in the scene [12, 13]. Automatic Intersection monitoring is an important problem in the context of intelligent transportation systems (ITS). A real-time scene monitoring system capable of identifying situations giving rise to accidents would be very useful.

Different sensing modalities such as loop detectors, radars, LIDARs and optical cameras provide automated collection of safety measurements. Inductive loop detectors are the most common equipments used in traffic management but they are not popular in the area of safety due to high installation cost. Radar-based systems are usually mounted on poles or embedded in vehicles to detect surrounding vehicles and pedestrians. Although, radar works fairly well for different situations regarding weather and illumination, its field of view (FOV) is narrow. LIDARs provide clearer measurements and wider FOV than radars but they higher hardware cost limits their usage. Table 1.1 compares different sensors for data collection at intersections.

Table 1.1: Comparison of common sensors used for data collection

Sensor	**Perceived Energy**	**Raw Measurement**	**Hardware Cost**	**Units**	**FOV**
Radar	Millimeter wave radio signal	Distance	Medium	Meters	Small
Lidar	Nanometer wave laser signal	Distance	High	Meters	Wide
Camera	Visible light	Light intensity	Low	Pixels	Wide

1.1 Motivation

Among the different data collection methods, traffic cameras are commonly used since they have already been installed for monitoring purposes by transportation management centers, and they can be leveraged by computer vision techniques to assess participants' behavior and their safety. Figure 1.1 shows the snapshot of the RTC/FAST traffic cameras which provide an access at different intersections and highways of Las Vegas. The main motivation behind this work is to leverage these existing cameras (see Figure 1.1) to provide useful information for traffic safety. Since automated system is developed to understand through cameras, the safety evaluation process will be cheap. However, there are different challenging situations regarding tracking road users and estimating their behavior that system needs to address them.

Vehicle and pedestrian detection and tracking is the foundational step which should be able to deal with challenging problems like different environment situations, occlusions and low resolution of images [14]. Vehicle and Pedestrian tracking in intersections entails new problems in comparison to highways regarding the presence of multiple road user flows including turning movements [12]. Since most vision-based systems use motion for detection and tracking of road users, they are not appropriate for intersection monitoring. Long-term stationary road users would merge into a background leading to missing targets. Appearance-based detection techniques also provide false alarms and they limit real-time capability due to lower detection speed. As a result, a robust tracking system which leverages cooperation of appearance and motion is required to provide a long-term track of vehicles and pedestrians regardless of traffic phase signals. Long-term track of participants enables system to analyze participant behaviors and estimate surrogate safety measurements such as DTI, and TTI.

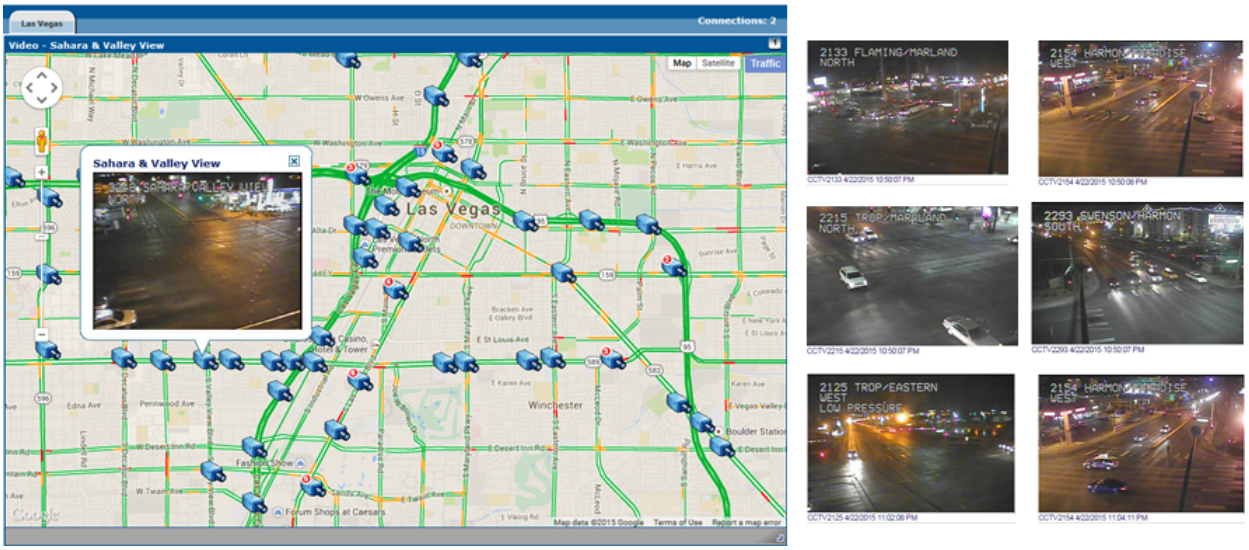


Figure 1.1: RTC/FAST traffic cameras [1]

1.2 Objective

The main objective of this dissertation is to develop a video-based data collection system that can automatically connects to cameras and provides high level of useful information for intersection analysis. The typical framework is proposed to provide a comprehensive safety and behavior analyses of participants with small human effort and high accuracy. The transportation applications addressed in this dissertation include turning movement count and speed analysis, queue length estimation, and conflict assessment. The main contributions of the dissertation are as below:

1. The proposed vision-based intersection monitoring system benefits robust detection and tracking by fusing of appearance and motion cues of intersection participants (i.e. vehicles and pedestrians). This fusion is used during detection and tracking benefits cooperation of enhanced optical flow and bipartite graph to provide long term trajectories of participants.

2. Different public available datasets of vehicles and pedestrians are prepared, trained and evaluated based on accuracy and speed. The best one is chosen for fusion with the motion-based detection. The UNLV dataset is finally introduced to better address detection difficulties of test videos.

3. The automated system provides long term behavior analysis of intersection participants by learning, modeling, matching and examining trajectories. Learning trajectories provide com-

prehensive behavior analyses of turning vehicles and crossing pedestrians. The real time turning movement count is proposed that benefits high accuracy using complementary counting modules.

4. The proposed systems provides a comprehensive behavior analysis of pedestrian at intersections. Waiting time, crossing speed and count are addressed by the proposed system. The efficiency of the system for each application is compared with the ground truth.

5. The proposed system evaluates pedestrians and vehicles safety by recognizing the observed trajectory of participants and predicting conflicts and their severity. The comprehensive safety measurements are estimated with high accuracy due to successful tracking of moving and stationary participants.

1.3 Outline

In Chapter 1, we provided a brief introduction to the area of research. We discussed about our motivation to choose this particular topic area for this dissertation work.

In Chapter 2, existing methods and algorithms in machine vision for object detection and tracking are reviewed. The critical discussion and comments regarding existing approaches for intersection monitoring is provided and different methods of behavior and safety analysis are provided. The chapter provide a comprehensive methods of behavior analysis for pedestrians and vehicles and safety analysis including gap, risk, threat, conflict and accident analysis. The methods are discussed and compared based on their potential of getting utilized by computer vision techniques.

In Chapter 3, we will give a brief overview to our approach for intersection monitoring. We provide our intuition for utilizing motion and appearance of objects to better detecting road users and our proposed tracking system as well.

In Chapter 4, we will discuss about our method for behavior analysis of vehicles and pedestrians. The objectives are turning movement count, turning speed and waiting time of vehicles. The same approach can concurrently provide crossing count, speed and waiting time of pedestrians. The obtained information help for design and planing intersections and signals.

In Chapter 5, we will present our approach for safety analysis through investigating obtained trajectories of road users . The surrogate safety measurements including Distance to Intersection (DTI), Time to Intersection (TTI), and Time to Collision (TTC) are addressed in this chapter.

In Chapter 6, we will show the results of our work. The proposed system is first evaluated at detection and tracking steps. The reason for using appearence for pedestrians and not vehicles is highlighted. The important mentioned behavior and safety parameters in Chapters 4, 5 are estimated and provided in this Chapter. The discussion about th results and the comparison with other methods are provided to show the efficiency of our system in measuring traffic information.

In Chapter 7, conclusion about our work is provided and future perspective is presented.

Chapter 2

Literature Review

Intersections are well-known targets for monitoring because of the high number of reported accidents and collisions. Around two million accidents and 6,700 fatalities occur at intersections every year in the United States [15], constituting 26% of all collisions [16]. Moreover, one fourth of all fatal accidents happen near intersections [16]. As a consequence, accident avoidance and safety improvement is a prime focus addressed by advanced driver assistance and safety systems.

2.1 Automated Monitoring of Intersection Participants

Automatically detection and tracking of vehicles and pedestrians have greatly benefited from vision-based techniques and have seen significantly advances in recent years. However, more traditional data collections often are preferred since robust detection remains a challenging problem at intersections due to several reasons:

1. Vehicles and pedestrians are not in continuous motion at intersections. Vehicles might stop behind the stop bars, and pedestrians stand waiting for a red traffic signal to change. Consequently, detection by using motion is not reliable [14].

2. Generally, there are more vehicles and pedestrians at intersections as compared to highways, which increases the possibility for interactions and occlusion, a well-known and challenging problem with vision-based monitoring systems.

3. Pedestrian detection by motion is harder than a vehicle in video surveillance because of their low quality and small size [14]. Consequently, many vision-based studies only consider vehicle tracking at intersections.

Since there is a greater trends toward automated data collection, this paper reviews vehicle and pedestrian detection as well as tracking based on available studies and systems for intersection monitoring purposes.

2.1.1 Detection

A variety of sensors have been used for vehicle and pedestrian detection at intersections [17], including global positioning systems (GPS), light detection and ranging (LIDAR), radar and cameras. Using cameras are more common since they are cheap and easy to install; in addition, there has been significant advancements in vision-based techniques.

GPS

Current collection trends have utilized trajectory data from GPS receivers in dash navigation systems of vehicles and mobile smart devices. GPS satellites provide localization messages for a receiver, which work well in open areas. However, accuracy is limited in areas with obstructed views of the sky (the urban canyon effect). Its usage is limited for safety applications because the spatial resolution, that is required, needs additional expensive hardware. Further, it is not feasible currently to expect that all intersection participants carry such a device.

LIDAR

Recently, LIDARs have become more popular ever since their usage-related costs have been improved, including for the lasers, sensor arrays and computations. LIDARs provide clearer measurements than radars and they have wider field of view. They work based on emitting laser at wavelengths normally between 600 and 1000 nm and scanning area at 10-15 Hz [18]. Several studies use laser scanners inside an ego-vehicle to detect and track other vehicles and enhance driver safety at intersections by warning them about the possible collisions and risks [19–23].

Radar

Radar technology usually is mounted on pols or embedded in vehicles for detecting surrounding vehicles [18] and pedestrians [17]. Radar works based on emitting frequency modulated signals. The frequency content of the demodulated received signal is analyzed and the distance to the detected object is determined using a frequency shift in the reflected signal. Although, radar works fairly

well for different situations of weather and illumination, its field of view is narrow. For instance, four radars were used by Aoude et al. [24] to detect vehicles and measure speed, range, and lateral position at a rate of 20 Hz with distance of 150 m away from intersection. In addition, several studies demonstrated the limitations of Doppler radar in the field of view as well as for detecting stopped vehicles [25–27].

Vision-based Cameras

Vision-based object detection is performed by using motion or the appearance of a target object. For video surveillance, background subtraction is a common method for detecting moving objects. Although this method is a traditional way of object detection, it is used a great deal because of its speed, simplicity, and good performance when there is continuous motion and normal congestion [12]. The method works by estimating the background from the sequence of images, and keep updating that to deal with changes in the background scene. However, its performance decreases when moving objects get close to each other and occlusion occurs. Since this method relies on motion, stationary objects can not be detected, such as vehicles behind the stop bars at intersections [14].

Appearance-based methods recognize a target object directly from images by evaluating the pixel values. Therefore, they do not need a sequence of frames; instead, they use a great many positive and negative samples to train the classifiers. Two basic appearance-based object detection methods that are very well known are presented in [28, 29]. Viola and Jones [28] used Haar like features to construct small and efficient boosted classifiers that could be cascaded to detect object of interest with high accuracy. Dalal and Triggs [29] demonstrated that a histogram of oriented gradients along with an Support Vector Machine (SVM) classifier showed good performance for pedestrian detection. This feature counted the occurrences of gradient orientation, computed on a dense grid of uniformly spaced cells to characterize an edge-like appearance.

2.1.2 Tracking

Tracking aims to generate the trajectory of an object by locating its position over the time. There are numerous methods for tracking (e.g. [17, 18], and [35]); this section briefly discusses those that have been used at intersection monitoring systems shown in Table 2.1.

Table 2.1: Vision-based tracking at intersections

Study	**Goal**	**Detection Method**	**Tracking Method**
Barth and Frank, 2010 [30]	Tracking oncoming and turning vehicles	Motion	Stereo vision tracking using kalman filter with interacting multiple model likelihood
Messelod et al., 2005 [31]	Tracking/ classification	Motion (background subtraction)	Feature and region-based tracking with model-based matching for classification
Saunier and Sayed, 2006 [32]	Tracking	Motion (optical flow)	Feature-based tracking
Veeraraghavan et. al., 2002 [33]	Tracking/ Classification for vehicles and pedestrians	Motion (background subtraction)	Region-based tracking using oriented bounding box overlap of detections
Veeraghavan and Papanikolopoulos, 2004 [34]	Vehicle Tracking, occlusion handling	Motion (background subtraction)	Region-based tracking using multiple cues with extended Kalman filter

Region-based Tracking

This method usually comes after background subtraction to track foreground regions or blobs. Shape characteristics of blobs (i.e., area, perimeter, color, and texture) usually are followed by using the Kalman filter in the tracking step. Since vehicles move in different directions at intersections, oriented bounding boxes and their overlapping areas between two frames can be used as a good metric for tracking [33, 34].

Feature-based Tracking

Selecting the right features for object representation is an important step for tracking. For example, color is used for histogram-based tracking [36] and edges are used for contour-based tracking [37]. Scale-invariant feature transform (SIFT) is shown to be a strong feature set as it is invariant under scale and rotation [38]. Among recent studies, Messelodi [31] chose edge-based features and Saunier and Sayed [32] relied on corners. Features are tracked over the time and they should be grouped inside moving objects and tracked based on some criteria, such as movement displacement and direction. One great advantage about feature-based tracking is that some features still remain when the object is partially occluded; however, it suffers from the problem of grouping features inside the object. Feature detection for an object varies for different sizes and image quality, which affects the performance of the tracking.

Optical Flow

Optical flow is the pattern of a displacement vector implying each pixel translation for the observer (e.g., an eye or a camera). Optical flow assumes brightness constancy of corresponding pixels over frames, and can be computed by two well known methods Lucas-Kanade [39] and Horn-Schunck [40]. Although optical flow is used for vehicle and pedestrian tracking at intersections [32, 41–45], it is computationally expensive and it doesn't work for stationary objects.

Particle Filters

Particle filters or Sequential Monte Carlo (SMC) methods are a set of on-line algorithms that estimate the posterior density of the state-space by directly implementing the Bayesian recursion equations. They are used for object tracking [46] in videos based on two separated steps. First, an object curve is built from its attributes such as color or edge, by using so many particles for sampling. As result of sampling, any arbitrary curve of an object can be tracked. Particles move based on object dynamics, including deterministic and non-deterministic factors. By using an observation model, particles get weighted and high weighted samples are chosen at re-sampling step. This process continues for each image sequence. Tracking by particle filters is robust toward partial occlusions and it can track stationary objects if their dynamics are well estimated. The algorithm requires sampling for each object, and periodic re-sampling of particles, which can be computationally expensive and prohibit real-time performance. In addition, initialization and learning each object dynamic is a challenging problem in particle-based tracking.

Spatial-temporal Markov Random Field

This was introduced by Kamijo et al. [47] for robust vehicle tracking at intersections. A Markov random field is a graph of random variables as nodes and it was used to generate vehicle labels of the image blocks [47]. Graph edges represent statistical dependency between nodes and adjacent blocks and blocks in consecutive frames are considered neighbors for the model. After forming and modeling the graph over the blocks of the image, nodes weights get updated by means of the back propagation process [48] and by obtaining an inference. Markov random fields show good tracking performance for highly congested intersections with high volumes of occlusion [47].

Table 2.2: Vehicle Behavior Analysis at Intersections

Study	**Data Source & Features**	**Goal**	**Classification, Inference or Approach**
Viti et al. 2008 [50]	Video: Trajectory	Analyzed speed and acceleration behavior	Probability density functions using trajectories
Kumar et al. 2005 [51]	Video: Trajectory, velocity, convex hull	Behavior recognition	Bayesian network
Kafer et al. 2010 [52]	GPS: Position, velocity, yaw rate	Long term trajectory prediction	Quaternion-based rotationally invariant LCS (QRLCS)
Hulnhagen et al. 2010 [53]	Observer: Trajectory, velocity, acceleration, steering angle	Maneuver recognition	Fuzzy logic system with a Bayesian filter

Stereo-based Tracking

Multi-view geometry provides 3D information that helps researchers with a better understanding of the scene, motion characteristics, and physical measurements. Vehicle tracking by stereo vision works by measuring and estimating the position and velocity in a metric space. The state vector mostly includes the vehicle's weight, height, lateral and longitudinal position, and velocity. Kalman filtering can be used for estimation when the noise is Gaussian with a linear motion constraint. Since real vehicle motion is nonlinear, an extended Kalman filter (EKF) often is used to estimate nonlinear parameters by linearizing the motion equations [30, 49].

2.2 Behavior Analysis

A behavior that belongs to the object of interest can be a single event (e.g., braking behavior) or a sequence of events indicating an action (e.g., turning right behavior includes a declaration as well as braking behavior). Behavior analysis provides an answer for some questions, such as what an object (vehicle or pedestrian) is doing right now or what is that object going to do in the upcoming seconds. In this section, we review important studies of behavior analysis according to three major groups of vehicles, drivers, and pedestrians. The reason for grouping based on intersections participants is due to characteristic of typical behavior which can exclusively occur by each group. For example, turning recognition belongs to vehicles while crossing count is commonly used to study pedestrians' behavior.

2.2.1 Vehicle Behavior

Trajectory Prediction, Vehicle Speed, and Acceleration

This includes measuring and learning vehicle kinematics, dynamics, and making predictions [13]. Assessing vehicle speed and acceleration [50, 51] are major areas of research. They are organized in a same group since they all investigate extracted trajectories of a vehicle after smoothing them by some filters such as Kalman. As a result, estimated behaviors can be compared against the models for defined areas. For instance, Kumar et al. [51] contextually defined check-post areas and evaluated vehicles speed on these predefined locations. The performance of kinematic trajectory evaluation is relatively strong for typical traffic but degrades during congestion. This is due to difficulties dealing with occlusion using background subtraction techniques for region-based tracking.

Turning Recognition

Turning movements are of particular interest for safety since these paths intersect with one another and may lead to hazardous situations. Predicting turning behavior involves learning turning patterns, building a model and finding a match for observed vehicle patterns with the model. There are two major methods to recognize turnings. The deterministic method recognizes turnings using learned prototypes such as longest common subsequence (LCSS) [12, 52] in video surveillance. The more elaborate methods consider variations in human driver behaviors within probabilistic frameworks. Since vehicle behavior is affected by the driver, a human having complex behavioral models, probabilistic frameworks are preferred to describe the future vehicle motion due to the uncertainty problem. This perspective of vehicle behavior mainly focuses on safety regarding other vehicles using various applications in Advanced Driver Assistance Systems (ADAS).

Kafer et al. [52] predicts probability values of turning left, turning right, and going straight by using a quaternion-based rotationally invariant longest common subsequence (QRLCS) to compare an observed trajectory with typical path models. This work predicts the long-term future trajectory of other vehicles, using particle filters and motion databases. Particles indicate motion states, which are weighted by comparison with the prototypes. Particle filters are efficient methods, in this context, and they can be applied for different scenarios. For example, Tran and Firl [23] improved upon their previous research [22] in trajectory prediction by using particle filters instead of an extended Kalman filter in order to recognize a maneuver by finding the best match from

Gaussian regression models. However, the number of particles and observation models should be carefully introduced since sampling from each trajectory pair can be computationally expensive and prohibits the real-time performance.

There is a trend towards using a probabilistic framework as a simple map to explain a vehicle behavior by case-based reasoning [53] and state-based diagrams [54] since they are simple to understand and use by drivers. These methods aim to imitate a person's natural reasoning process; however, they become challenging when burdensome states, conducted by a driver, become complicated. In addition, most accidents occur in a quite short amount of time without rigidly following predefined rules. Therefore, they should benefit on-line learning with probabilistic methods [53] to deal with different driving models with the uncertainty problem. For instance, Hulnhagen et al. [53] used a probabilistic finite state machine to model the compound driving maneuvers by breaking them into simple fundamental states such as braking, and following a vehicle. A Bayesian filter approach infers the driving maneuver by computing the probability of each basic element in the context of the 'maneuver' model.

Table 2.2 shows representative works for vehicle behavior analysis with respect to learning algorithms, applications, and major features. The behavior assessment studies [50, 51] provide medium level of output accuracy due to tracking limitation which can be improved by optical flow method. Safety assessment studies [52, 53] do not address real-time processing since they analyze trajectories off-line.

2.2.2 Driver Behavior

Driver behavior analysis includes measuring, learning, and predicting driver intention [6, 24, 55–59], perception reaction time (PRT) [25, 60, 61], and braking response at intersection junctions and dilemma zones.

Turning Intention & Aggressive Behavior

Predicting the drivers' intention including turnings is different from turning prediction methods used for vehicles because it requires not only vehicle dynamics but other high resolution data sources as well such as human factor measures (e.g., look direction). The uncertainty of the driver's intention demands probabilistic methods such as Bayesian networks [56] and hidden Markov models [57]. In a few recent studies, SVM was used as a learning framework for binary classification [24, 55] due to two major reasons:

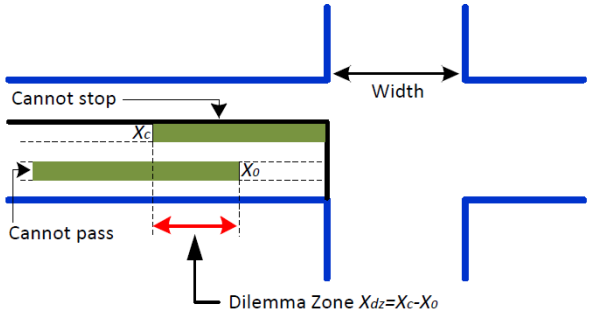


Figure 2.1: A graphical illustration of a dilemma zone at a signalized intersection [2]

1. Training SVMs involves an optimization problem of a convex function, thus the optimal solution is a global one.

2. The upper bound on the generalization error does not depend on the dimensionality of the space. Hence, the dimensional variation does not affect the classification accuracy.

Aoude et al. [24] used SVM to classify violating drivers. However, the method only is limited to recognizing drivers that do not stop behind the stop bar. However, this can happen by compliant drivers at unsignalized intersections with no stop bar. A variety of violating behaviors should be considered in the literature.

Other traditional learning methods- Bayesian networks and hidden Markov models (HMM)- still are used extensively since they are simple, capable methods to model dynamic stochastic processes and they do not need large quantities of data for their training. For instance, a model developed by Streubel and Hoffman [57] learned behavior HMMs by using the calculated speed, acceleration, and yaw rate for each direction: straight, left, and right. Although this method was used to evaluate a large database of real driving data, it could not be applied to arbitrary intersections. Lefevre et al. [56] used a Bayesian network to address arbitrary intersections by incorporating knowledge about the layout of the intersection using contextual information from a digital map.

Perception Reaction Time (PRT)

PRT is defined as the time difference between the onset of the yellow phase and the activation of the vehicle's brake lights, and it is a critical measure for the design of intersections and timers.

Moreover, PRT of drivers is usually studied to estimate a dilemma zone. A dilemma zone is a range in which a vehicle approaching the intersection during the yellow phase can neither safely clear the intersection nor stop comfortably at the stop line (see Figure 2.1). Dilemma zones have been the subject of much attention for observing behaviors since there is a high possibility of collisions due to hesitancy by a driver to stop or pass the traffic signal. The National Highway Traffic Safety Administration (NHTSA) reported in 2005 that 805 of the fatalities were the result of red light running [62].

The collisions within a dilemma zone can be avoided if the location and the length of a dilemma zone is known for an intersection. The dilemma zone can be estimated using stopping x_c and clearing x_0 distances, as shown in Figure 2.1,

$$x_c = v\sigma + \frac{v^2}{2a},\ x_0 = v\tau - (W + L) \tag{2.1}$$

where v is the speed of the approaching vehicle, σ is the perception reaction time of the driver, a is the maximum deceleration rate of the vehicle, τ is the length of the amber signal, W is the width of the intersection, and L is the length of the vehicle. A segment of a dilemma zone approaching the stop-line is said to exist when $x_c > x_0$. In literature, the critical PRT value is normally 1-s which is $\geq 85\%$ of the PRT cumulative distribution [63–65].

There are two major problems regarding to dilemma zone estimation:

1. A dilemma zone distribution varies for different drivers regarding their age and gender [66]. As a result, some studies incorporate these measurement to find a model for dilemma zone distribution [64, 66].

2. The dynamic nature of a dilemma zone is difficult to measure [2, 67] since the acceleration/deceleration rate and the PRT should be captured at the onset time of the yellow light. As a result, naturalistic observation of videos is preferred since other methods, such as simulators and controlled roads, produce practice effects. Although, difficult to obtain generally, traffic phase data could be collected by video recordings, and drivers' characteristics could be obtained by some complementary methods, such as the communication between vehicles and infrastructure-based systems at intersections.

There are various ways to control the dilemma zone for purposes of avoiding a collision. For instance, Tarko et al. [68] used a probabilistic approach to control dilemma occurrence by developing a dilemma likelihood function and finding the optimal green extension to minimize this likelihood.

Table 2.3: Driver Behavior Analysis at Intersections

Study	Application	Data Source	Description	Important Findings
Gates, et al. 2007 [60]	Behavior assessment at dilemma zone.	Video	Deceleration rates and brake response times were estimated by manual calculation of related parameters such as approaching speed, brake response time, and headway.	PRT is significantly affected by approach speed, distance to intersection at onset of yellow, and deceleration rate.
Rakha et al. 2007 [64]	Behavior assessment at the onset of the yellow-phase.	Instrumented vehicle	Distribution of PRT and the probability of stopping/running versus the distance to the stop bar was estimated.	Males have low probability of stopping. 1 sec PRT is valid and older drivers less likely clear intersections at short yellow trigger distances.
Liu and Tao 2006 [2]	Behavior assessment to estimate dilemma zone.	Video	Velocities, acceleration/deceleration rates, distance, and expected time to the stop line were measured to find the dilemma-zone distribution	Dilemma zone is dynamic in nature and its location varies with the driving populations.
Caired et al. 2007 [69]	Behavior assessment to estimate PRT.	Driver simulator	Dependent variables including the stop/go percentage, velocity, PRT, deceleration, stopping accuracy, and intersection clearance were evaluated using statistic software package (i.e. UNIONOVA).	PRT of 1 sec is sufficient to accommodate older drivers.
Kaysi and Abbany 2007 [70]	Behavior assessment to model aggressive behavior.	Observer	Distribution of aggressive driving estimated for driver characteristics (gender and age), car characteristics (type and model year), and traffic attributes. Aggression defined for turning vehicle that forces major road to slow.	Indicators of aggressive behavior are age, car performance, and average speed on major road.

To be truly effective, a green phase extension requires driver information regarding sex, age, and time to intersection.

Table 2.3 shows the representative works of driver behavior regarding their goals, applications, and important findings. Our evaluation of literature shows that behavior analysis requires a high data collection cost. However, Gates et al. [60], and Liu and Tao [2] showed reasonable results and cost due to usage of video recording to estimate parameters used in Eq. (2.1) to calculate dilemma zone distribution. This shows the effectiveness of using video recordings to estimate PRT and dilemma zones.

2.2.3 Pedestrian Behavior

Pedestrians behavior studies can be grouped into "prediction-based" and "observation-based" methods. The early prediction of a future trajectory is the main focus for prediction-based methods since it can be used in safety systems (e.g., ADAS) to avoid collisions. The question, "Will the pedestrian cross?" should be answered early and promptly [71]. However, the question is difficult to answer since it is not simple to infer the current state of pedestrians due to the localization problem [72]; pedestrians' behavior is highly dynamic as well.

Motion Prediction

One way to perform a prediction relies on the closed-form solutions of a Bayesian filter, such as Kalman filtering; however, non-parametric stochastic models are preferred due to non-linearity of pedestrians' motion patterns. Possible trajectories are generated by Monte Carlo simulations using dynamical models. For instance, Abramson and Steux [73] combined a constant motion model with particle filtering but constant motion model loses its accuracy for standing pedestrians around the crosswalk.

Markovian models also have been used to model pedestrians' motion [72], using fours states of a Markov chain, corresponding to 'stand still', 'walking', 'jogging' and 'running'. However, the model does not explicitly address the 'pedestrian crossing' state in the traffic safety domain.

Predicting pedestrian future states as crossing or not requires a binary classifier with more features than just positions and velocities. As a result, SVM is ideal for this task, and the most meaningful features can be verified to make a reliable prediction about the pedestrians' future behavior. Pedestrians' distance to the curb and the crosswalk should be considered as important features for training models and making predictions.

Waiting Time, Walking Speed, Crossing Speed and Choices

Measuring waiting times and speeds are used in numerous studies [74, 80] since they are valuable metrics for intersection design and signal control. Walking and crossing speeds are two major elements to characterize pedestrians' behaviors [76, 79, 81], and contributing factors include age, gender, and group size [79, 81]. Pedestrian routes and crossing choices are two other important factors considered during behavior analysis since they might lead to hazardous situations, such as walking or crossing at unmarked locations [75, 77, 78].

Table 2.4: Pedestrian Behavior Analysis at Intersections

Study	Data Source	Description	Important Findings
Hamed 2001 [74]	Observers, interview	A risk function models waiting times and Maximum Likelihood Estimate (MLE) is used to model parameters.	A pedestrian's expected waiting time significantly affect required attempts to make a successful crossing.
Sisiopiku and Akin 2003 [75]	Video, survey forms	Sidewalk cameras for movement data and survey for complementary information. Compliance rate (ratio of ped. count over total peds. including jaywalkers) is used for comparison.	Mid-block crosswalk is most influential facility. Distance of crosswalk to desired destination most influential factor in decision to cross.
Lam and Cheung 2000 [76]	Video	Estimated walking speeds and flow relationships for commercial and shopping areas.	Pedestrians tend to walk faster in commercial areas than in shopping areas. Pedestrians generally walk slower on crosswalks with mid-block than those crosswalks without mid-block.
Keegan and Mahony 2003 [77]	Video, survey forms	Extracted relationship between green cycle length and number of pedestrian crossings. A survey was used to determine pedestrians' characteristics, such as age and gender.	The count-down timer induced a reduction in the number of individuals who crossed during the red signal.
Bernhoft and Carstensen 2008 [78]	Questionnaire forms	The χ^2 test showed a significant difference between old versus young respondents and men versus women. A linear regression was used to build a model.	The younger group generally finds it important to move fast but older group shows more cautious behavior.
Tarawneh 2001 [79]	Observers	A four-factor analysis of variance (ANOVA) was performed with the mean speed as the dependent variable and age, gender, group size, and street width as independent variables.	Age, gender, group size and street width significantly contribute to a pedestrian speed.

The methods require evaluating over a long period with high accuracy, followed by statistical evaluations including t and χ^2 tests. The statistical methods look at the aggregates from the samples of large populations to determine the significant differences between studying behaviors and factors.

Table 2.4 shows key factors used for pedestrian behavior analysis, and various findings in representative studies. Most pedestrian behavior studies suffer from high data collection cost. However, this extensive data collection resulted in important findings. The usage of automatic data collection as a simplified analysis tool is suggested which can be supplemented with survey forms, interviews,

and recorded videos as complementary methods to decrease data collection cost and provide a comprehensive analysis. For example, if automatic data collection is used, recorded videos help to verify incidents and correct observations. In addition, pedestrian age and gender information can be obtained by using high resolution cameras.

2.3 Safety Analysis

Safety inference refers to two different tasks. One is an assessment process undertaken by pedestrians or drivers at the decision making level to avoid the accidents as below:

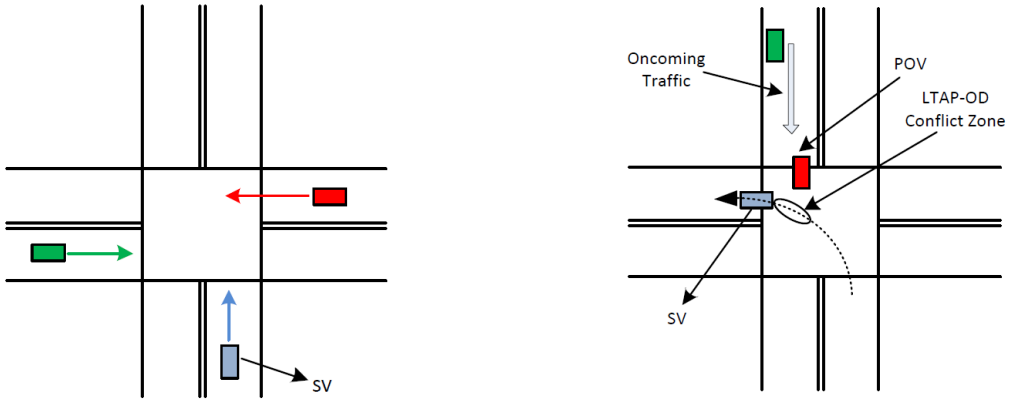
- Gap: An available time/space for a maneuver or between leading and trailing vehicles in a car following model.
- Risk: The uncertain level of danger introduced by other vehicles for the subject vehicle under a specific mission (e.g., turning risk, or crossing risk).
- Threat: The possible danger of imminent collisions introduced by other vehicles.

The second task involves models and predictions for conflicts (i.e., near-accidents) and accidents built from safety measurements, crash datasets and police reports.

2.3.1 Gap

While pedestrians and drivers are making the decision to pass or cross at intersections, they check the sequence of gaps that can be rejected or accepted. An accident or near-accident may be caused when the accepted gap is small. Gap analysis consists statistical inference and modeling of accepted gaps, which are studied at intersections due to two major reasons:

1. The first and second most common types of accidents at controlled intersections having two stop signs involve scenarios for a) a straight crossing path (at over 45%) and b) a left turn access path/opposite direction (at around 25%) [25]. They happen when drivers in the subject vehicle are not able to accurately judge the speed of the approaching vehicles and the available time gap to complete their turning or crossing maneuver. A conflict involving the left turn access path/opposite direction (LTAP/OD) occurs when a subject vehicle (SV), while making a left turn, encounters a threat presented by an approaching principal other vehicle (POV). POV refers to the opposing vehicle that most likely might be in conflict with the SV due to its closeness in terms of distance or time. The scenarios are shown in Figure 2.2.



(a) Straight crossing path [82] (b) LTAP/OD [25, 83]

Figure 2.2: Two common accident scenarios

2. If gap acceptance behavior is accurately estimated for drivers or pedestrians, intersection decision support (IDS) systems can correctly predict the turning or crossing time by monitoring approaching vehicles.

Gap inference includes estimating the frequency of accepted gaps, the median acceptable gap size, and gap orders [84, 85]; usually, these are modeled by probability distribution functions, regression models, and logistic functions [86]. However, it is difficult to implement a gap inference for real traffic scenarios since it should be customized based on age and gender of the drivers and pedestrians. For instance, the probability distribution of accepted gaps were studied in [84] [85] for different driver characteristics (e.g. females versus males) using driving simulators; it was found that males more likely accept smaller gaps than females. Other important factors were observed on gap acceptance behavior, such as the velocity of the oncoming vehicles. Drivers would accept a smaller gap size for scenarios of higher speeds on major roads [85].

Few studies have investigated pedestrians gap acceptance (PGA) due to the difficulty in data collection and the lack of supporting safety systems for pedestrians. For instance, Sun et al. [87] modeled PGA by estimating the probability of the acceptable gap size, and Banerjee et al. [88] evaluated the presence of pedestrians on drivers gap acceptance in the LTAP/OD scenario. Sun et al. [87] studied the probabilistic gap acceptance behavior as a random variable that best fits the training data. Since the pedestrian decision is to reject or accept the available gap, the binary logit model with such factors as waiting time and number of waiting pedestrians on the curb side was

Table 2.5: Gap Analysis at Intersections

Study	Application	Data Source	Subject	Important Findings
Alexander et al. 2002 [84]	Safety assessment: building a model by regression to predict incidents (i.e., an accident or near-accident) from gap and driver characteristics.	Driving simulator, survey forms	Driver	Male and young drivers more likely accept smaller gaps in comparison with female and elderly drivers.
Yan et al. 2007 [85]	Safety assessment: investigation of vehicle speed, driver age, and gender on left- turn gap acceptance.	Driving simulator, survey forms	Driver	The major road traffic speed and driver age and gender have significant effects on the gap acceptance maneuver.
Ragland et al. 2005 [86]	Safety systems: modeling gap acceptance of left-turner subject vehicle (SV) in order to design Intersection Decision Support (IDS) systems.	Video	Driver	Presented gap to the driver of SV has a log-normal distribution. Accepted gap ranged from 3 to 12 seconds.
Chan et al. 2004 [89]	Safety systems: extracting time gap acceptance (TGA) to design IDS systems.	Video	Driver	Warnings should issue in the range $t = -5$ to $t = -3$ of driver decision. Pedestrians' presence significantly affect driver behavior.
Leung and Starmer 2005 [90]	Safety assessment: Evaluation of age and alcohol use on gap acceptance behavior.	Driving simulator, survey forms	Driver	Young drivers demonstrated and increased tendency to engage in risky tactics.
Sun et al. 2003 [87]	Safety assessment: modeling pedestrian gap acceptance using deterministic, probabilistic, and binary logit methods.	Video	Pedestrian	The minimum accepted gap by younger pedestrians is less. The mean of accepted gap increases marginally with waiting time.

used in their study. The waiting time was incorporated in [87] since it was shown by Hamed et al. [74] that high waiting times make pedestrians impatient to cross.

Table 2.5 describes gap analysis studies with more details including goals and important findings. Most studies expose high cost except those use manual observation of videos. Although driving simulators facilitate the gap estimation process (e.g., they provide the location and speed of oncoming vehicles), they suffer from the sickness problem. In particular, simulator sickness is correlated with turning maneuvers and braking behaviors [85]. Without a real safety risk as a field test, sickness in a simulator experiment may significantly harm drivers' gap acceptance decisions because the drivers would try to complete the experiment as soon as possible to reduce the discomfort level.

2.3.2 Threat

Some studies evaluated threat assessment [91–93] by predicting the possible threats of other vehicles [91] or by calculating the safety measurements [93]. Predictions were made based on combining the intention predictor and an efficient threat assessor by using rapidly-exploring random trees. The threat assessor computed the threat level, and the corresponding maneuver for the best escape route. Chan et al. [93] assessed the threat of opposing traffic by finding the time to intersection (TTI) value, which indicates a practical way of using safety measurements for threat analysis. When TTI is greater than a threshold, a time window is open for the SV to take its maneuver.

Other threat assessment studies [91–93] collected data using instrumented vehicles and vision based systems. A threat assessment method has real applicability in ADAS systems as long as collected trajectories are evaluated in real-time fashion to determine the possible threat levels of opposing traffic.

2.3.3 Risk

Risk assessment is the process of detecting a dangerous situation that might cause by drivers' error, for example, perception failure, misunderstanding of the situation, or a wrong decision. Drivers mostly evaluate the risk of collision with other vehicles by assessing the other drivers intentions. Driver intention should be predicted and compared against the expectation [94, 95] in order to determine the probability of the risk. As an alternative approach, the collision risk of an intended path could be compared against all possible paths of other vehicles, according to probabilistic models [19]. The Bayesian network and HMM are two popular learning frameworks in this context [19, 94, 95].

Pedestrians' risk exposures were estimated using data-mining techniques on observed data, reported accidents, and collision datasets. For example, King et al. [96] used crash datasets and observation data for illegal crossings, such as walking against the red light and crossing away from the signals but within 20 m. In their studies, relative risk ratios were calculated for these categories using annual crash reports, and experimental results indicated eight times higher risk of illegal crossings.

Tiwari et al. [97] observed video recordings of seven intersections to determine the number of safe and unsafe crossings, waiting times, and survival times. An unsafe pedestrian crossing is defined when the traffic signal indicates a green or yellow light for vehicles. Survival analysis

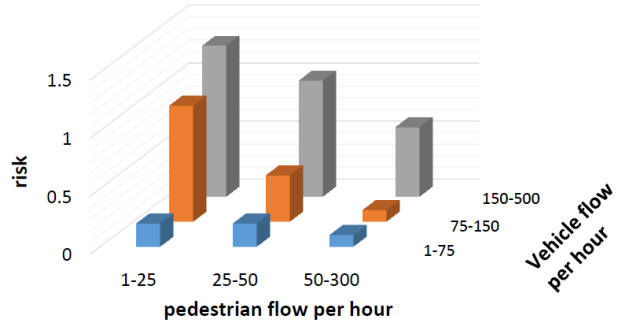


Figure 2.3: Risk versus vehicle and pedestrian flow [3]

leverages waiting time, the number of waiting pedestrians, and unsafe crossings in order to provide the probability of initiating an unsafe crossing. Survival analysis is a popular method used in medical science to study the effectiveness of different drugs on cancer patients. Tiwari et al. [97] applied this idea to investigate the variant nature of pedestrian risk as a function of time.

Leden [3] investigated police reports of accidents to determine the correlation between pedestrians' risk and vehicles' flow. As shown in Figure 2.3, risk decreased with increasing pedestrian flow and increased with increasing vehicle flow. In addition, left-turning vehicles caused higher risk for pedestrians than right-turning vehicles.

Unfortunately, risk analysis studies usually introduce their own definition regarding near-accidents or unsafe situations. Further efforts are required to codify universally consistent definitions. Since accident dataset is a valuable source for risk analysis, a comprehensive dataset is necessary regarding pedestrians and drivers with their accident records. In addition, accident records, specifically for pedestrians and drivers, are critical for risk analysis. The accuracy of risk analysis is thus subject to the the accuracy of collected accident data. Currently, these are obtained manually from video recordings and automated vision-based systems could greatly facilitate this branch of safety studies.

2.3.4 Conflict

As long as there is a correlation between accidents and conflict-based safety measurements, and the safety measurements are reliable and consistent in definition, they can be proven as practical metrics for safety analysis [98]. A traffic conflict mostly is used in safety analysis defined by

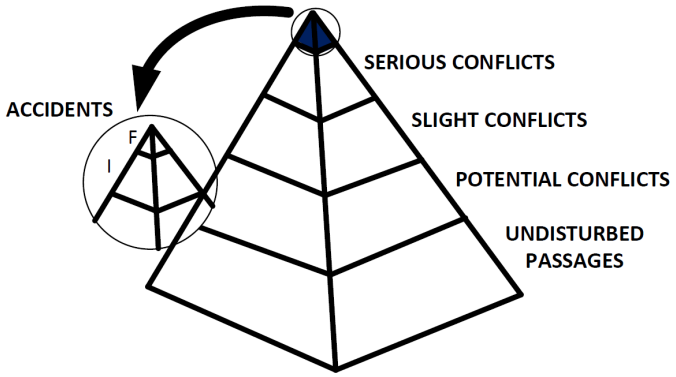


Figure 2.4: Traffic safety pyramid measurement showing the hierarchy of traffic events (F= Fatal, I=Injury) [4]

Amundsen and Hyden [99] 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". Figure 2.4 shows a hierarchical concept of using conflicts that implies critical observations used in safety analysis. Associated severity can be estimated for each traffic event in the hierarchy, thus, the severity represents its location in the hierarchy. In this section, surrogate safety measurements are explained with their application for safety quantification and finally conflict-based studies at intersections are presented.

Surrogate Safety Measurements

Important safety measurements used in intersection studies are shown in Table 2.6. Low values of time to collision (TTC) and post-encroachment time (PET) are used in literature to imply a high probability of collisions. Low TTC values also can be mapped to a severity index with the formula defined in (2.2),

$$SI = \exp\left(-\frac{TTC^2}{2PRT^2}\right) \tag{2.2}$$

where SI is the severity index, ranging from 0 to 1, and PRT typically is 1-s as shown in Section 2.2.2. The distribution of the severity index is an important measurement for safety evaluation [42].

Headway is another safety indicator in a car-following model [103]. Time headway is measured

Table 2.6: Safety Measurements Used in Intersection Studies

Parameter	**Definition**
Time To Collision (TTC)	The time for two vehicles (or a vehicle and pedestrian) to collide if they continue at their present speeds on their paths. [6, 7, 26, 42, 44, 45, 90, 100]
Distance To Intersection (DTI)	The distance until a vehicle reaches to stop bar with current speed. Stop bar is used as reference point. [2, 6, 7, 44, 45, 64–66, 69, 89, 93, 100, 101]
Time To Intersection (TTI)	The time remains until a vehicle reaches the stop bar with its current speed. Stop bar is used as a reference point. [2, 26, 57, 64, 65, 69, 89, 93]
Time Headway	Elapsed time between the front of the lead vehicle passing a point on the roadway and the front of the following vehicle passing the same point. [60, 61, 90]
Post-Encroachment Time (PET)	Time lapse between end of encroachment of a turning vehicle and the time that the through vehicle (or pedestrian) actually arrives at the potential point of collision. [42, 44, 45, 100, 102]

as the difference of time that leading vehicle i and following vehicle $i-1$ reach a same location in a car-following model. TTC in car following model is defined when $V_i(t)$ is higher than $V_i(t-1)$. Otherwise, a vehicle can have large or infinite TTC value but still have relatively small headway, which indicates potential danger. This is the major reason that headway is more appropriate in a car following model.

Departure headway is an important variant of headway used to measure the intersection capacity and time of traffic signals. It is usually defined as the time that elapses between consecutive vehicles when vehicles in a queue start crossing the stop line (or any other reference line) at a signalized intersection, after the light turns green [104,105]. Since the inaccurate estimation of departure flow often leads to an inappropriate signal-timing plan, many investigations had been carried out to study the statistics of departure headways regarding such external influence factors as the number of lanes and vehicle types. For instance, Jin et al. [104] and Yin et al. [106] showed that departure headways follow a certain log-normal distribution for each vehicle position in queue with different mean and variance values. This distribution is suggested intuitively to interpret the outcome of the interactions between the vehicles in the discharging queue.

Since availability and quality problems are associated with collision data, some studies have relied upon traffic conflict analysis as an alternative or a complementary approach to analyzing traffic safety. Traffic safety analysis has been investigated separately for vehicle-vehicle and vehicle- pedestrian conflicts. For vehicle-vehicle conflicts, safety measurements include distance to intersection (DTI) [7, 26, 27] Headway [61], and TTC [100].

Chan et al. [26, 27] addressed the LTAP/OD scenario (see Figure 2.2) to assess the left-turn conflicts. The data acquisition system included radars to capture vehicle position and speed relative to the intersection and a video camera to provide complementary data. This group showed the effectiveness of PET and TTC values to compare and quantify safety at three different intersections. However, the safety quantification method should be conducted for all scenarios since various turning behaviors are manifested from vehicles at a typical intersection.

Automated vision-based systems address all turning scenarios by providing reliable trajectories of intersection participants and directly estimating the surrogate safety measurements during tracking. For instance, Sayed et al. [100] leveraged TTC measurements to identify the safety deficiencies at regions of interest. The accuracy of the direct methods were affected by the reliability of individual vehicle and pedestrian tracks which would be noisy for occluded and stopped vehicles. As an alternative way (i.e., indirect conflict detection), trajectories are clustered to learn the typical models using probabilistic frameworks. For instance, statistical sequence clustering by HMM was used in [43] to learn models of conflicting trajectories. However, HMMs require considerable data for the reliable estimation of the model parameters. Determining the model parameters, especially the number of model components in HMM-based clustering, is a complicated and uncertain process.

Conflicts at intersections are defined in some studies as abnormal or unexpected behavior of vehicles. Vehicles approaching an intersection follow a certain model that helps to determine their abnormal behavior. Akamutsu et al. [7] developed the Bayesian framework to learn a model and detect conflicts on the onset of braking using vehicle velocity and DTI. However, the model requires improvement using on-line updating methods, since a variety of driving patterns are introduced by different drivers.

Turning Conflicts

For vehicle-pedestrian conflicts, violations by pedestrians and their conflicts with right- or left-turning vehicles were studied. Ismail et al. [44, 45] extracted DTI, PET, and TTC; and Zaki et al. [41] investigated pedestrian violations by comparing a given track and normal movement prototypes. Sayed et al. [42] relied on the development of a database to cover all interactions between road users, including TTC and other measurements. Most vehicle-pedestrian conflict analyses use cameras and a tracking system with optical flow to provide trajectories of pedestrians and vehicles. Although feature-based tracking with optical flow can tackle the partial occlusion problem, it is challenging for stopped pedestrians with small size. In addition, grouping the features

regarding road users, is another problem which is addressed by hierarchical clustering. Hierarchical clustering requires perfect criteria for the segmentation process; however, road users are prone to the over segmentation problem [32].

In general, the main threat to pedestrian safety rises from the interaction with turning vehicles since crossing pedestrians and turning vehicles share the common phase of signal. Since left-turn conflicts with pedestrians frequently occur, some studies addressed that issue by extracting vehicle speed and PET [102] or by using regression models to find the greatest contributing factors [107]. The conflict point was determined in [102] by predicting future trajectories using predefined models, such as the speed profile, gap acceptance and stopping and clearing profiles. The accuracy of the system is affected by cubic functions used to estimate model parameters and situation results do not reflect real driving scenarios.

Table 2.7 shows representative studies of conflict based safety analysis methods and highlights their applications and classification. Among the presented methods, those with automated vision-based methods using optical flow (OF) [42–45] demonstrated better performance. The tracking accuracy is high due to robustness against partial occlusions and the limitations of stopped vehicles does not affect TTC and PET calculations.

2.3.5 Accident

Accident-based safety analysis includes various methods used to learn and model accident patterns, making predictions to prevent accidents and using data mining techniques on accident reports.

Automated Accident Detection

Automated vision-based methods are available that can address accidents and collisions, based on predicting the future state of the vehicle by using vehicle dynamics. Collisions are detected if there is an overlap between the predicted 3D cubic models of vehicles at the same time. As a typical example of vision-based work, Kamijo et al. [47] addressed three types of accidents 1) bumping accidents, 2) stop and start in tandem, and 3) passing. These types of collisions were detected using HMM to learn the crash patterns. Akoz [108] used continuous HMM for clustering paths, and linear regression for recognizing the severity of an accident.

The major difficulty of vision-based systems is to provide a robust tracker against occlusion, which is an undesired effect of accidents. This manifests as well when vehicles get too close and the segmentation of vehicles becomes highly challenging. Region-based tracking using background

Table 2.7: Conflict-based Safety Analysis at Intersections

Study	**Application**	**Data Source**	**Participants**	**Classification, Inference or Approach**
Chan and Bougler 2005 [27]	Safety system (CAS): Assessed conflicts by cooperative roadside and vehicle-based data collection.	Video, vehicle, radar, loops	Vehicle-Vehicle	Estimated the distribution of some measurements, such as TTC, steering wheels and angles.
Saunier and Sayed 2006 [43]	Safety assessment: Traffic conflict detection.	Video	Vehicle-Vehicle	Used HMM for clustering trajectories of traffic conflicts.
Ismail et al. 2010 [44, 45]	Safety assessment: Localized conflicts using heat map.	Video	Vehicle-Pedestrian	Optical flow tracking, classification of pedestrians and vehicles by speed and trajectory prototypes.
Alhajyaseen et al. 2012 [102]	Safety assessment: Conflict analysis with left turning vehicles.	Video	Vehicle-Pedestrian	Manual video observations; estimated vehicle speed profiles, PET.
Sayed et al. 2012 [42]	Safety assessment: Localized rear-end and merging conflicts by heat map.	Video	Vehicle-Pedestrian	Optical flow tracking; estimated distribution of TTC and the severity index.

subtraction is still a good solution for accident prediction since 3D model of object can be inferred using foreground mask [109].

As an alternative way, extracting different features regarding shape and motion tracks (e.g., variation rate of velocity, position, area, and direction) are quite common for accident detection [110]. For instance, Atev et al. [111, 112] inferred all possible pairs of rectangles that intersect in current and future time steps, based on the estimated position, orientation, and size of the vehicles. In studies by Hu et al. [109, 113] motion patterns were learned by neural networks, and the probability of accidents was calculated for partial trajectories obtained by 3D model tracking of vehicles [109].

Since automatic accident detection from videos is a complicated task, non-vision-based techniques rely on other sensors for vehicle detection. Harlow and Wang [114] used acoustic signals to automatically detect accidents by creating a database from traffic features and accident sounds. In a study by Streib et al. [20] LIDAR data was used to detect vehicles, and the severity of collisions was detected using a 3D model estimation of the target by an extended Kalman filter. Salim et al. [9, 115, 116] used a simulation environment. So, they did not need to address tracking problems

and collision patterns were stored in a knowledge base for statistical inference operations.

Data Mining

Data mining algorithms work on data of accident reports to find cause of accidents and contributing factors. Vehicle-pedestrian accidents are addressed through the non-vision based studies using real crash datasets and police reports. Since real datasets are used, statistical inferences are more accurate and valuable. Finding reasons for the accidents [117, 118] with regard to the types of vehicles [119], time, location, and injury [120] are the common subjects for these intersection studies. Crash datasets is also used to build a model based on pedestrian intersection safety indices ($PED\ ISI$), used to determine the safety index score for a single pedestrian crossing. The model is defined in Eq. (2.3),

$$PED\ ISI = 2.372 - 1.867(S) - 1.807(St) + 0.335(Tl) + 0.018(Sp) + 0.238(Cm) + 0.006(Ma)(S) \quad (2.3)$$

where S, St, and Cm are binary values indicating signal controlled, stop signs, and predominantly commercial areas. Tl is the number of through lanes, Sp is the 85% speed of the street being crossed, and Ma is the main street traffic volume.

Table 2.8 shows safety analysis methods based on accident reports including important findings and classification methods. The major problem involves limited availability of real crash datasets which are required for informative predictions. Moreover, the update of estimated models and new inferences undergo a long time period since real accident data is the rare event which firstly need to be collected and reported by police. Vision-based systems have better performance for collecting accident events due to lower cost and automated analysis.

Table 2.8: Accident-based Safety Analysis at Intersections

Study	Application	Data Source	Participants	Classification, Inference or Approach	Important Findings
Ki and Lee 2007 [110]	Safety assessment: Accident detection and reporting model.	Video	Vehicle-Vehicle	Extracted features with a predefined threshold value	The proposed system detected some accidents which were not reported by police in two weeks evaluation.
Lee 2005 [117]	Safety assessment: Analysis of vehicle-pedestrian crashes.	Crash dataset	Vehicle-Pedestrian	Developed two types of models to analyze frequency and injury severity of pedestrian crashes	Middle age (25-64) and male drivers are more involved in crashes. Higher average traffic volume increase the number of pedestrian crashes.
Al-ghamdi 2002 [119]	Safety assessment: Association analysis between crash severity and such variables as age, gender and nationality.	Crash dataset	Vehicle-Pedestrian	Chi-square and odd ratio techniques	77% of pedestrians were struck while crossing a roadway outside of crosswalk area. More than one third of fatal injuries were located on the head and chest.
Preusser et al. 2003 [118]	Safety assessment: Extracted crash type versus culpability pedestrians and drivers	Crash dataset	Vehicle-Pedestrian	Extracting information from crash datasets reported by the police	Pedestrians were slightly more likely to be judged culpable. Turning vehicle crashes typically involves driver's failure to yield pedestrians.
Zeeger et al. 2007 [121]	Safety assessment: Safety effects of installing crosswalks at uncontrolled locations.	Video	Vehicle-Pedestrian	Manual observation, Poisson and negative binomial regressive models	Pedestrians' crash severity of marked versus unmarked locations didn't differ significantly for two-lane roads but it did for multi-lane roads.
LaScala et al. 2000 [122]	Safety assessment: The effects of environmental and demographic characteristics on pedestrian injury collisions.	Crash dataset	Vehicle-Pedestrian	Spatial analysis	Injuries in pedestrian crashes are greater in the areas with higher population density and average daily traffic.

Chapter 3

System Overview

A typical framework for automated vision-based intersection analysis is shown in Figure 3.1. Video data (i.e., a sequence of images) is provided as input to the system in either an online or offline manner. The system front-end deals with detecting and classifying objects as road users (i.e., vehicles or pedestrians) while the back-end provides higher-level traffic activity analysis.

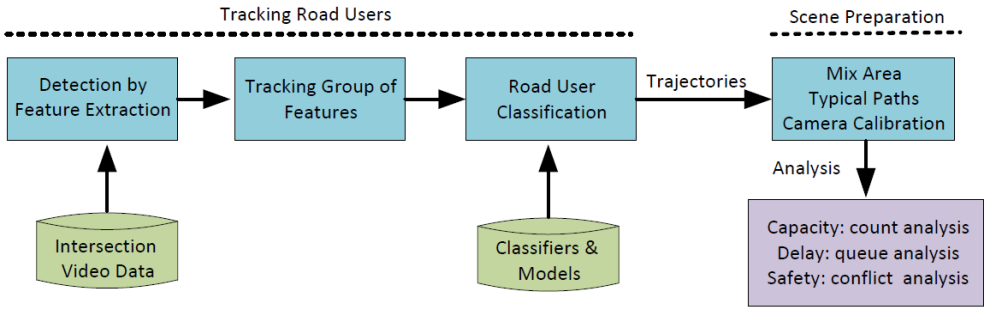


Figure 3.1: Intersection monitoring system overview

Feature extraction methods are common for detecting objects in a dense scene. The salient features (e.g., corners) identify parts of an object and motion features identify the whole of moving object. The good point about the salient features is that they can be reliably detected even with occlusion. These features are identified in consecutive video frames and matched through tracking. The features are grouped (e.g., using spatial proximity) to form object proposals for road users. These object proposals are passed to a classification stage which uses predefined criteria or models to classify the road user proposals as vehicles, pedestrians, or bikes. Classifiers can be simple, such

as a speed measurement, or can be complex appearance models obtained from machine learning techniques.

Intersection activity analysis is conducted on the extracted trajectories of road users for object-specific understanding. Camera calibration is required to convert trajectories in image coordinates to the 3D world coordinates. The same trajectories can be used for different types of analysis as shown in the bottom right of Figure 3.1. The following sections highlight major parts of the tracking system for front-end processing which provides trajectories for further analysis (e.g., safety analysis).

3.1 Detection by Feature Extraction

A cascaded system is proposed for reliable vehicle/pedestrian detection at intersections presented in Figure 3.2. The main advantage of this system is the use of both motion and appearance feature cues in a contextually meaningful manner for accurate detection. The addition of appearance-based detection to the traditional surveillance processing pipeline is motivated as the following:

1. Although motion is used reliably on highways, it is not consistent at intersections since traffic signals force participants to stop temporarily.

2. Pedestrians tend to be more stationary and their detection is more challenging than vehicles since they are small in size, and they have a non rigid body.

3. Partial occlusion (i.e., blob merging) might occur between pedestrians crossing intersection and recently stopped vehicles. Thus, pedestrian and vehicle detection by motion results in a large single object which causes false recognition and trajectory. In addition, only one object is detected for pedestrians who walk together in groups in crosswalks.

3.1.1 Motion-Based Object Detection

Standard adaptive background estimation methods such as time averaging of the frames perform poorly on scenes with a large amount of clutter and slow moving objects [123]. A Gaussian mixture model (GMM) [124] is used to create an adaptive background from the sequence of frames. Each pixel is modeled by K Gaussian mixtures to address lighting changes and slow moving objects.

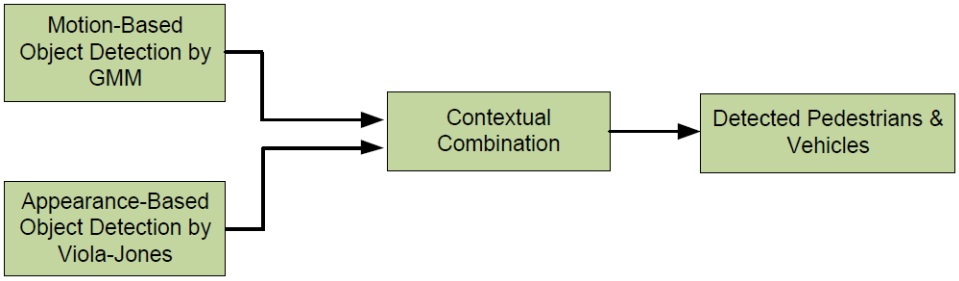


Figure 3.2: Vehicle & pedestrian detection system. The group of motion and appearance features provide detections and contextual combination chooses the best candidates based on overlapping criterion.

Moving objects such as pedestrians and vehicles are detected as pixels which do not fit any of the background Gaussian models.

3.1.2 Appearance-Based Object Detection

Haar-like features are rectangular based features which are popular for object detection due to fast computation (e.g. image integrals), and high detection performance. Haar-like features are utilized to construct weak classifiers (i.e. small efficient boosted classifiers) which are cascaded to detect almost all objects of interest while rejecting a certain fraction of the non-object patterns in a computationally efficient manner [28].

3.1.3 Contextual Combination

The key role in the detection system is the contextual combination or pooling of several positive detection responses. Contextual combination provides fusion at the decision level to combine the outputs from the GMM and Haar detections in special mix areas (Figure 3.11) where both detectors are active. In this way, appearance detection is run on the smaller processing regions for speed and reliability.

The contextual combination is defined to:

1. Reject many false appearance-based pedestrian/vehicle detections outside mix areas to reduce a total amount of false positives.

2. Select the most reliable detection from either GMM or Haar in each detection pool cluster that have overlapping bounding boxes.

The full contextual combination process operates in mix areas where both the GMM and Haar detectors are active. The detections from each are pooled into detection clusters based on bounding box overlap. All bounding boxes with more than 50% overlap are considered part of a same cluster. The bigger bounding box of each cluster is removed at each step until there are no more overlapping of higher than 50% between each two bounding boxes of the cluster.

Examples of contextual combination is shown in Figure 3.3. Clusters of interest are noted by black arrows and GMM detections are in blue, Haar vehicles in brown, and Haar pedestrians in green. Figure 3.3a shows three detections for two recently stopped vehicles behind the stop bar. From (a) to (b), the GMM bounding box (i.e., the bigger bounding box) is abandoned in favor of the correct detection from multiple Haar boxes. Figure 3.3c shows two detections of a wrongly detected vehicle as a pedestrian and a walking pedestrian. The correct detection of the pedestrian is retained which can not be accurately indicated by GMM.

3.2 Tracking Group of Features

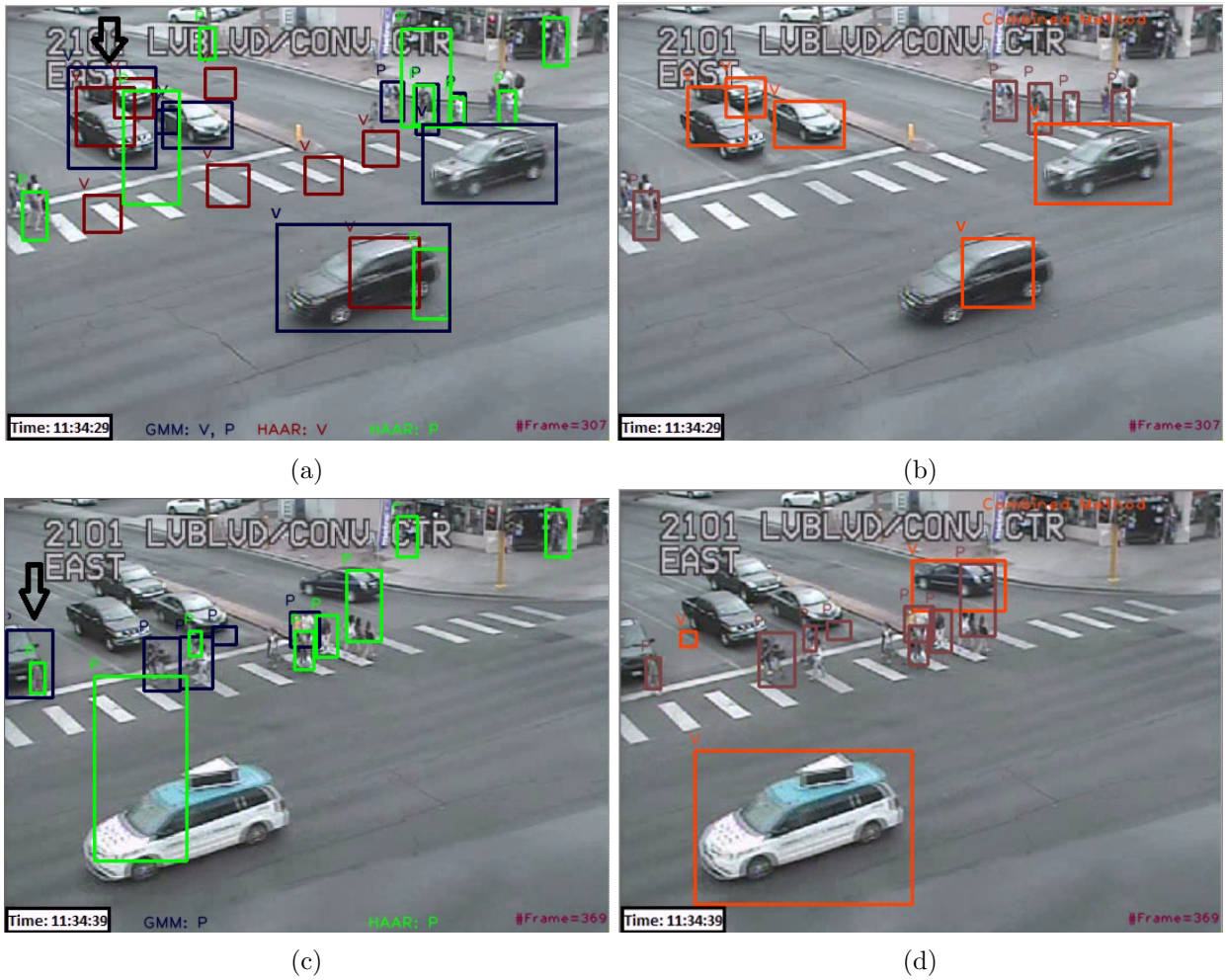
3.2.1 Tracking System

The tracking system utilizes the detection system along with enhanced optical flow in order to provide a robust tracking of vehicles and pedestrians. As it is shown in Section 6.1.1, detection results of appearance-based classifiers for vehicles are not satisfactory and GMM is a dominant detector on mix areas. The optical flow undergoes some improvements to keep track of vehicles for a long time period.

Figure 3.4 shows the tracking system which benefits cooperation with the bipartite graph to handle temporarily undetected pedestrians. Final detected pedestrians (i.e., through contextual fusion in mix area) and vehicles (i.e., through contextual fusion in motion area) feed the tracking system which uses bipartite graph at first to initialize the tracks. The initialized tracks use enhanced optical flow which utilizes different feature types inside tracks including corners and textures. This idea provides different advantages in comparison with basic optical flows as follows:

1. Detected standing pedestrians at intersections are successfully tracked by optical flow in spite of not having an initial motion.

2. Basic optical flow methods rely on updating and grouping features based on direction and displacement of motion vectors. Grouping features for pedestrians and vehicles move together



(a) (b)

(c) (d)

Figure 3.3: Contextual combination: (a), (c) Vehicle & pedestrian detection by GMM and Haar, V and P correspond to vehicle and pedestrian respectively, (b), (d) Detection results using contextual combination, vehicles are shown with the orange color while pedestrians are shown with the brown color. For multiple detections shown by black arrow in (a) and (c), two stopped vehicles and a single pedestrian are successfully recognized by the detection system shown in (b) and (d) respectively.

with a same speed and direction is a challenging problem. The initialized tracks provided by bipartite graph tracker helps to group features and solve the problem.

The initialized tracks use optical flow as long as the tracking process is successful. The likelihood of successful tracking is determined based on the quality of the detected feature matches and the estimated bounding box around the features. If these values are less than predefined thresholds, which means the optical flow tracker failed, detected pedestrians are utilized by the bipartite graph tracker to handle the problem. This is due to running appearance-based classifiers on each frame in parallel with the optical flow tracker. If contextual fusion does not provide any detection for the

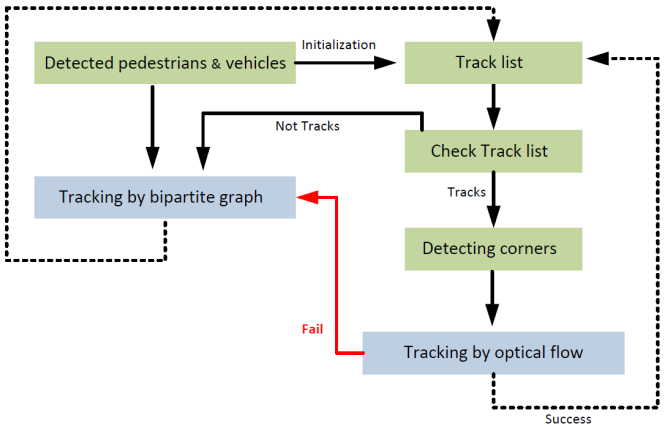


Figure 3.4: Pedestrian and vehicle tracking system. Bipartite graph initializes tracks and tracking proceeds by optical flow. When optical flow fails bipartite graph keeps a track by finding its closest detection from contextual combination.

bipartite tracker at the time of optical flow failure, the track of the pedestrian is finally lost.

Vehicles lose their track if optical flow fails any time outside the motion area which is defined as an area with a large distance from the stop bar where vehicles do not usually stop. This tracking scheme provides a robust tracking of vehicles against blob merging and occlusion since detection is only performed over the small area (i.e., motion area). The motion areas for vehicle is shown in Figure 3.11b. The appearance-based classifiers are not used for vehicles since they do not show the significant improvement in comparison with motion-based techniques [14]. In addition, vehicles have large and rigid bodies which reduce the failure ratio of tracking by optical flow.

Enhanced Optical Flow

Optical flow is a primary tracker since it is robust against partial occlusion, and its high performance is shown for pedestrian activity analysis in [41, 100]. The optical flow is enhanced in this work through the three steps of enriching feature points, filtering features, and bounding box estimation. The improvements are performed in order to tackle tracking drifts and provide tracks of waiting vehicles and pedestrians.

Enriching feature points

Small pedestrians provide a very small number of high quality corners (e.g. one or two corner features for a pedestrian with width of less than 15 pixels) and this number might be reduced during tracking. The problem worsens for stopped or slow moving vehicles and pedestrians. The idea is to sample each vehicle and pedestrian with more feature points, called enriching feature points, to prevent the mentioned problem. Although Harris corner features have shown good performance for high quality images, their performance lessens for small and low quality image samples. As a result, Speeded-Up Robust Features (SURF) [125] are accumulated as representative of texture based features.

Filtering features

Filtering features removes falsely detected features inside each track. This is a crucial step in order to avoid tracking drift since all detected features inside a bounding box are not located on a vehicle or a pedestrian image sample. Since texture based features are also utilized, there is a higher chance of detecting static features around the moving features regarding vehicles and pedestrians. These noisy features are the major source of false bounding box estimation which might lead to a tracking drift. The filtering process is performed by determining the state of the track (i.e. waiting, moving) and removing the opposing features. For example, when a vehicle or pedestrian is in state of waiting, the static features are predominant and moving features are removed. The way of determining the state of the track is based on the average of the displacement vector.

Bounding box estimation

Bounding box estimation is a crucial part of the tracking process since it helps to keep the track of stopped vehicles in a queue. When there is a motion cue, a bounding box can be efficiently estimated based on spatial distribution of matched features obtained by optical flow. However, the bounding box estimation of stopped vehicles and pedestrians is a challenging task which leads to a tracking drift. As a result, the fixed bounding box around the fixed position is leveraged when the waiting state is determined. This helps significantly to maintain a track since feature points are reduced by optical flow when an object is stationary.

Bipartite Graph

Bipartite graph uses a greedy approach to find a nearest detection for a track [12]. Tracks and frame's detections are nodes of the graph and a cost between each two nodes is the difference between appearance measurements as calculated in (3.1),

$$A(v_1, v_2) = A_{pos}(p_1, p_2)A_{size}(s_1, s_2)A_{appr}(h_1, h_2) \tag{3.1}$$

where A_{pos}, A_{size} and A_{appr} are affinities based on position, size and appearance defined as follows.

$$A_{pos}(p_1, p_2) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left(-\frac{(x_1 - x_2)^2}{2\sigma_x^2}\right) exp\left(-\frac{(y_1 - y_2)^2}{2\sigma_y^2}\right) \tag{3.2}$$

$$A_{size}(s_1, s_2) = \frac{1}{\sqrt{2\pi}\sigma_s} exp\left(-\frac{(s_1 - s_2)^2}{2\sigma_s^2}\right) \tag{3.3}$$

$$A_{appr}(h_1, h_2) = \sum_{i=1}^{m} \sqrt{h_1^u h_2^u} \tag{3.4}$$

Difference of distance and size plugs into Gaussian kernel to compute A_{pos} and A_{size}. σ_x, σ_y and σ_s are standard deviations calculated for sequence of locations (i.e., x, y) and sizes separately. A_{appr} is a Bhattachayya distance measurement calculated separately for two histograms (i.e., h_1, h_2) of colors and edges.

3.3 Road Users Classification

The classification of road users to vehicles and pedestrians can be performed at either detection or tracking steps. At detection step appearance-based detectors are utilized by training specific features (i.e., Wavelets, Histogram of Oriented Gradients (HOG)) of positive and negative samples and get a model which can be later used to distinguish vehicles from pedestrians. In this dissertation, we have used two different methods of classification [14, 42] at appearance and tracking levels. However, we finally decided to use the tracking level classification (i.e., Bayesian method using speed & aspect Ratio) since appearance-based detectors need more time to provide the raster scan over entire frame to detect objects.

3.3.1 HOG features with Support Vector Machine

In order to classify road users as a pedestrian or a vehicle, a moving object can be classified based on its appearance. As a result, different appearance-based techniques can be applied to classify object as a vehicle or a pedestrian such as Histogram of Oriented Gradients (HOG) [29] and local binary patterns (LBP) [126]. The major difficulty of appearance-based classification is due to speed, and collecting effective training samples. For example, the segmentation of foreground objects might be noisy which leads to failure in extracting features and then classification fails when appearance-based classifier is used at verification step. In addition, the accuracy of the classifier degrades for a new scene since samples have not been introduced for the classifier. As a result, an active learning framework is required which enhance the classifier performance through collecting training data for a new scene [127].

Object recognition using HOG features with Support Vector Machine (SVM) is quite popular for vehicles and pedestrians [29,128]. This features count occurrences of gradient orientation computed on a dense grid of uniformly spaced cells. They use overlapping local contrast normalization for improved accuracy. HOG Features with SVM Classifier is used as third stage of the cascaded system for verifying detected objects as vehicles or pedestrians. Using HOG with SVM classifier as verification step has some benefits like reducing false positives and leading to more speed up [129, 130].

Positive and negative samples from the first stage are used at this stage but there is still possibility that one object is recognized as both pedestrian and vehicle by two different classifiers. So, positive samples of each classifier, are added to negative samples of another. For example, positive samples of vehicle detection classifier are added into negative samples for pedestrian classifier. If there is still the same positive predication response for both pedestrian and vehicle, width and height of detected object is used to distinguish them. Vehicles more often have higher width and pedestrians have higher height.

After extracting HOG features of positive and negative samples, feature vector of each image is put into training file . For training with LIBSVM [131], the linear kernel for reaching to high speed with same setting in [29] is used. After preparing test data set by using 75% of training set and 25% of new object samples, that have manually labeled, the ROC curve for vehicle and pedestrian classifiers are obtained. Figure 3.5 shows that vehicle classifier has little bit better performance than pedestrian classifier.

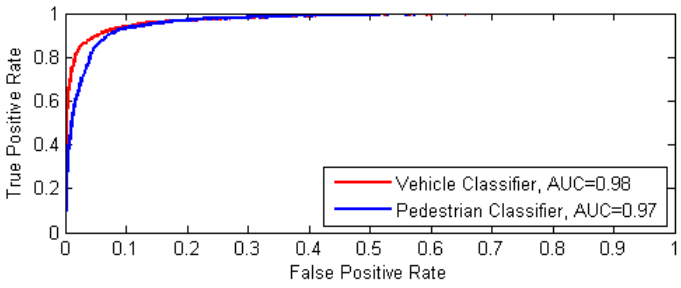


Figure 3.5: ROC curve for Vehicle & Pedestrian Classifiers

3.3.2 Bayesian Method Using Speed & Aspect Ratio

Practical vision-based systems use speed as a criteria for classification [42, 44, 45] for intersection monitoring. However, classification of slow moving vehicles become challenging at intersections. In this paper, three different parameters are used to solve a classification problem which are an aspect ratio, speed and path similarity. Suppose there is a $k = \{ratio, speed, path\}$ where $ratio = width/height$ of a track and, $path$ shows trajectories of the track. Typical models are known by $m_j = \{ratio_j, speed_j, path_j\}$, where $j = \{1, 2\}$ which indicates a vehicle or a pedestrian. Suppose two classes of vehicle and pedestrians C_1 and C_2 are defined and $p(C_j|ratio, speed, path)$ is the posterior probability,

$$p(C_j|k_1, k_2, k_3) \propto p(C_j) \times p(k_1|C_j) \times p(k_2|C_j) \times p(k_3|C_j) \tag{3.5}$$

where k_1 , k_2, k_3 correspond to on observed $ratio$, $speed$ and $path$ of the track. The similarity between an observed parameter and each model (i.e. $j = 1, 2$) are obtained for $i = 1, 2$ using normal kernel as below.

$$p(k_i|C_j) = \frac{1}{\sqrt{2\pi}\sigma_{k_i}} exp\left(-\frac{(k_i - m_j)^2}{2\sigma_{k_i}^2}\right) \tag{3.6}$$

The similarity score of typical paths for k_3 is calculated using LCSS method and N is the total number of typical paths for each class of vehcile (C_1) and pedestrian (C_2).

$$p(k_3|C_j) = \max(D_{LCSS}(T, F_m)) \; where \;\; 0 < m < N \tag{3.7}$$

Moving object is classified to a vehicle or a pedestrian if the calculated posterior probability is

Dataset	Number of samples
Caltech [132]	946
Graz [133]	127
MIT [134]	143
Tripod [135]	2162
UIUC [136]	550
VOC [137]	645
UNLV	16035
Total	20608

(a)

(b)

Figure 3.6: Positive image samples for training appearance based vehicle detection a) Number of collected samples from each dataset 1 b) Typical samples, last row refers to UNLV dataset

higher. $p(C_j)$ is a prior probability which is initialized to 1 but it is replaced by posterior probability for next frame prediction. As a result the prediction over time frame is recursively updated by a Bayesian method for better estimation. For example, when a track has been classified as a pedestrian for 3 frames one noisy misclassification can not change the final classification output.

3.4 Datasets

3.4.1 Vehicles

Large number of image samples were collected from the available public datasets for training vehicle detection classifier. Image samples from several recorded videos of Las Vegas highways and intersections with different camera views were also collected to create new dataset called UNLV. Figure 3.6 shows number of samples collected from each dataset along with UNLV dataset. The reasons for introducing UNLV data set is as follow:

1. The available vehicle datasets are few and the number of samples are also small. They are mostly high quality and large samples that have poor performance for highly cluttered and noisy videos.

2. Vehicles at intersections have different orientations and sizes. Therefore, different vehicle image scales and orientations is needed for training vehicle detection classifier.

3.4.2 Pedestrians

Appearance-based classifiers require large datasets with varieties of poses to be effectively trained. Subsequently, available public datasets were organized and prepared to train an appropriate appearance-based detector for video surveillance based on accuracy and speed (see Figure 3.7).

Dataset	Samples
Daimler [138]	14401
ETH [139]	1243
INRIA [29]	3542
MIT [140]	924
NICTA [141]	37344
TUD Brussel [142]	3272
Caltech [143]	2014
UNLV	3000
Total	65740

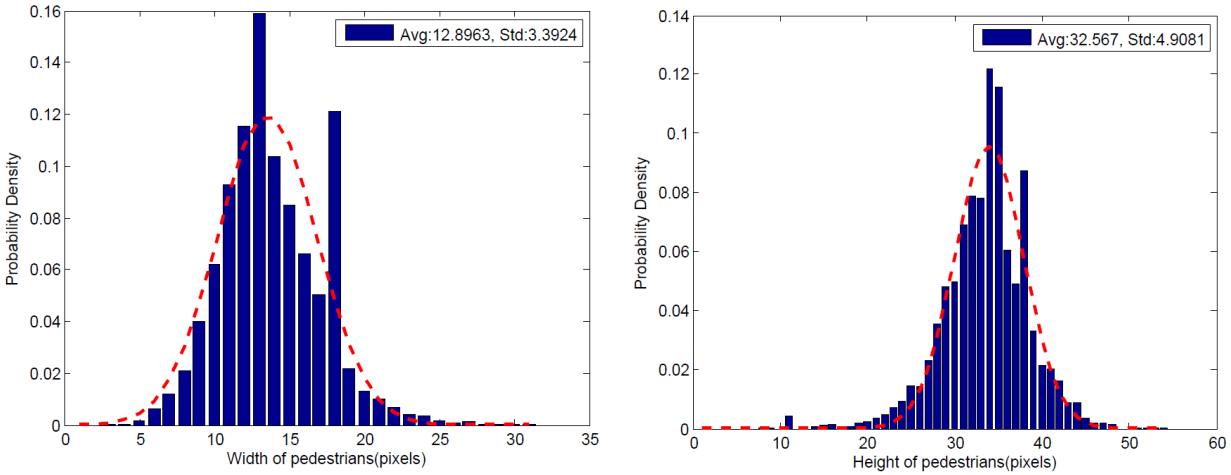
(a)

(b)

Figure 3.7: Positive image samples used for training pedestrian detectors a) Number of collected samples from each pedestrian dataset b) Typical samples

The manual annotated pedestrians help to provide samples' statistics to adjust parameters required to train pedestrian classifiers. Since pedestrians' width and height are in same range for a typical video, width and height distribution should follow the normal distribution. These two parameters are important to be adjusted during the training process and they also show the distance of the camera from the scene which indicates difficulty level of the tracking process.

Figure 3.8 shows the example of the width and height distribution with fitted log normal curves for first intersection from the annotated pedestrians of 1000 frames. There are some peaks that violate normal curve which is due to erroneous labeling of occluded pedestrians. The pedestrian statistic (i.e. (Avg, Std)) was $(12.89, 3.39), (17.06, 5.13), (14.64, 3.87)$ pixels for width and $(32.56, 4.90), (41.35, 10.68), (33.59, 4.07)$ pixels for height using three testing intersection videos (i.e.,



(a) Width distribution (pixel)

(b) Height distribution (pixel)

Figure 3.8: Distribution of pedestrian width and height for INT1

INT 1, INT 2, INT 3). As a result, the minimum width and height are chosen to resize all samples before the training process (i.e., width: 12, height: 32).

Dollar et al. [143] grouped pedestrians' distance to vehicle into three classes: close, medium and far. The height of less than 30 pixels is considered 'far' which indicates a difficult level of detection in Advanced Driver Assistant Systems (ADAS). The intersection sample statistics imply the difficult level of pedestrian detection since the distribution of heights are in the range of 30-40 pixels, and infrastructure-based cameras have more perspective distortion. As a result, the large public available datasets are trained and evaluated in this work to check their efficiency for detecting small pedestrians in video surveillance. Since most available datasets are captured through the vehicle-based cameras (i.e. Caltech dataset), they are not appropriate for intersection monitoring purposes. As a result, Caltech dataset is pruned, and UNLV surveillance styled dataset is prepared.

Caltech Dataset

The Caltech pedestrian dataset consists of approximately 10 hours of video taken from a vehicle driving through regular traffic in an urban environment. About 250,000 frames with a total of 350,000 bounding boxes were annotated which includes labeling for pedestrians, people and occlusion scenarios [143]. The large dataset consists of low quality and fully occluded samples that make

(a) Caltech (b) Caltech (c) UNLV (d) UNLV

Figure 3.9: Example of positive samples, (a) Group of pedestrians occluded by leading vehicles, (b) Low quality, partially and fully occluded pedestrians, (c) Crossing pedestrian samples, (d) Waiting pedestrians and crossing pedestrians from the side and front views

it challenging to prepare an efficient detector based on speed and accuracy. The low performance of appearance-based classifier was observed when all the samples are used to train the pedestrian detectors by Haar and LBP features. The reason is due to lots of difficult samples that can not be easily distinguished from the individuals since low quality and fully occluded pedestrians should be counted as negative samples. Figures 3.9a and 3.9b show some typical examples of difficult samples annotated in Caltech dataset.

The Caltech dataset should be pruned in order to provide an effective pedestrian detector for video surveillance. As a result, the difficult examples from the whole training samples were manually removed, and 2014 samples remained after the pruning process. The improvement in accuracy and speed is shown in Figure 3.10 obtained after the pruning process which is noted by "after the treatment". Although there is a small reduction in accuracy performance for LBP detector, the applicability of the Caltech dataset is shown for video surveillance due to high speed-up (i.e., Haar: 60, LBP: 19) obtained after the treatment.

UNLV Dataset

UNLV dataset [144] is specifically provided for video surveillance by leveraging available traffic cameras in the Las Vegas area. There are two major reasons for developing a new dataset. First, some of the datasets like Caltech and NICTA have not been proposed for video surveillance purposes, and they do not perform well in spite of having a large number of samples (i.e., NICTA). Dataset samples have been collected from vehicle-based cameras to provide a robust detector for ADAS. As a result, they are mostly large and high quality samples at a ground-level viewing angle which make them inefficient for video surveillance and traffic monitoring applications. Secondly,

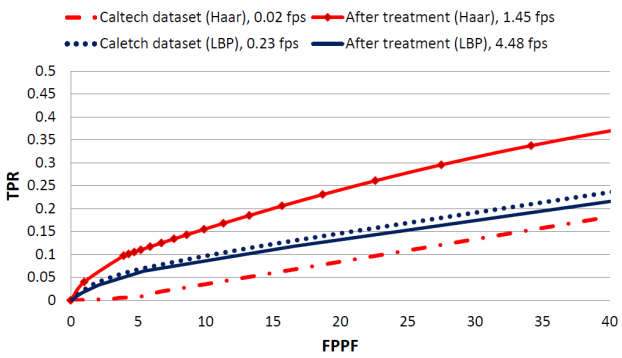


Figure 3.10: Caltech dataset performance before and after removing complicated samples and pedestrian detection speed (frame per second: fps)

there is a direct relation between the context and quality of collected samples with the test videos. Similar to the idea of active learning [127], if samples are collected from available cameras with varieties of poses, the optimized classifier can be generated for the testing videos.

Table 3.1 shows different cameras along with collected samples at intersections. Locations and samples were selected based on the availability of pedestrians. Subsequently, large number of samples were collected from the Las Vegas boulevard which is a famous street for tourism purposes. In order to provide an efficient classifier, medium number of samples (i.e., 3000 samples) with varieties of poses and scales were collected. For example, pedestrian crossing samples are different from a time of waiting. We collected samples from both moving and waiting pedestrians shown in Figures 3.9c and 3.9d.

3.4.3 Negative Samples

Collecting negative samples is easier since it can be any picture except vehicles or pedestrians for each corresponding classifier. 23985 and 89798 negative sample were collected for vehicle and pedestrian detection classifiers.

3.5 Scene Preparation

3.5.1 Mix Area

GMM is used for entire scene since all objects might have motion in all areas. However, some areas like behind the stop bars , around signals (box area) and crosswalks are considered as mix areas.

Table 3.1: Intersection names [1] and corresponding number of samples collected for UNLV dataset

Location	# Samples	Location	# Samples
WELCOME LV SIGN	65	CASINO/ BRUCE WO	35
FLAMING/ MARLAND	108	LV BLVD/ TRES ISL	256
LV BLVD/ CIRCUS S	112	CASINO/ COL BELL S	161
LV BLVD/ CATHEDRAL	80	LV BLVD/ HARRAHS	95
LV BLVD/ CONV CTR	40	LV BLVD/ CAESARS	80
LV BLVD/ WELCOME N	80	LV BLVD/ SLS	32
FLAMING/ KOVAL NE	70	LV BLVD/ VENETN	25
LV BLVD/ MONTE CRLO S	313	LV BLVD/CONV CTR E	161
LV BLVD/ MONTE CRLO N*	994	FLAMING/ KOVAL SE	293

Mix areas are shown in Figure 3.11 for vehicles and pedestrians with red and purple colors. Since mix area will be only used for pedestrians due to evaluation results (see Section 6.1.1), Figure 3.12 shows it for only pedestrians along with typical paths for three different intersections. Although there is motion in crosswalk, it is considered as mix area because of some reasons as below:

1. Pedestrians mostly move together around the crosswalk and detection by GMM only give one object which is group of people. Detecting one large pedestrian instead of separated pedestrians leading to poor pedestrian detection as occlusion result.

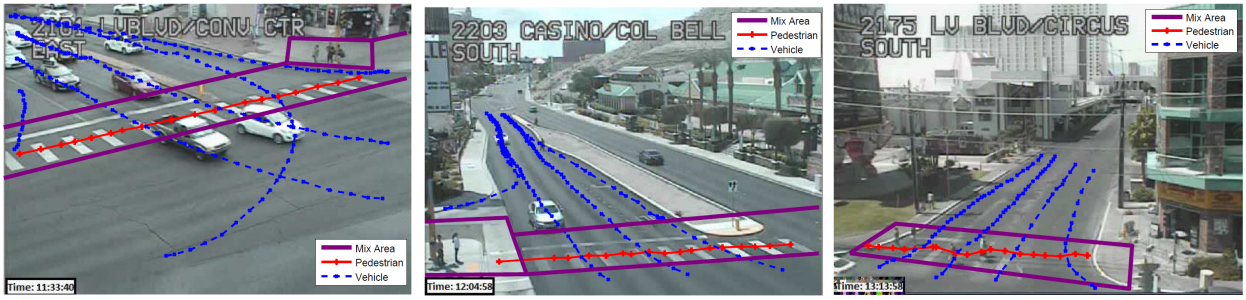
2. When vehicle have recently been stopped and pedestrians are moving in front of them, detection by GMM gives one bounding box for both pedestrian and vehicle as result of occlusion. Since whole object including vehicle and pedestrian is given to classifier, classifier works poorly in pedestrian recognition.

3.5.2 Typical Paths

Typical paths are defined to provide crossing counts and remove false tracks. Since crossing behavior is an important parameter for intersection design, the tracking system recognizes pedestrians while crossing using temporal alignment techniques between an observed trajectory and the stored typical crossing path. The longest common subsequence (LCSS) is used as a similarity measurement for comparison since it has shown a good performance for variable length trajectories [12]. Moreover, a

(a) (b)

Figure 3.11: Mix areas definition in order to run appearance-based plus motion-based detectors. Areas with red and magenta colors are used for vehicles and pedestrians respectively. The motion area is defined in (b) for vehicles.



(a) (b) (c)

Figure 3.12: Typical paths (i.e. pedestrians (blue), vehicles (green)) and mix areas defined for pedestrians : a) INT1, b) INT2, c) INT3

single "strong" bakground frame is collected and saved for later use in removing false positives. The "strong" background is currently manually selected from a video as a frame with no pedestrians. Future work will look to develop an adaptive method for its creation.

3.5.3 Camera Calibration

The main purpose of camera calibration is to establish a set of camera parameters in order to find a relationship between the image plane coordinates and world coordinates. This step is essential for translation of an image measurement into meaningful world units. For example, the speed of a vehicle is calculated in an image as pixels per second and should be converted to another metric like miles per hour. The process of converting image coordinates to world coordinates is usually performed for transportation monitoring through ground plane homography normalization [145].

A 2D point (x, y) in an image can be represented as a homogeneous 3D vector $x_i = (x_1, x_2, x_3)$

where $x = x_1/x_3$ and $y = x_2/x_3$. A homography is an invertible mapping of points and lines on a projective plane P_2 between two images. In order to calculate the homography which maps each x_i to its corresponding x_j in a second image, it is sufficient to compute the 3×3 homography matrix H which only has 8 degrees of freedom.

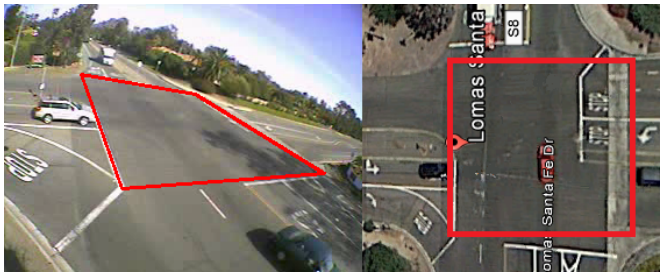


Figure 3.13: Four-points correspondence between camera image plane and map-aligned satellite image to estimate the homography (H) matrix and convert image locations to world latitude and longitude (INT 5).

Typically, homographies are estimated between images by finding feature correspondences in those images. The basic method is Direct Linear Transform (DLT) that can be applied to obtain the homography matrix H given a sufficient set of point correspondences. Since we are working in homogeneous coordinates, the relationship between two corresponding points x_i and x_j can be re-written as:

$$cx_j = Hx_i \tag{3.8}$$

$$c\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = H\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{3.9}$$

where c is any non-zero scale constant.

Dividing the first row of equation 3.9 by the third row and the second row by the third row results in the following two equations:

$$-h_1x - h_2y - h_3 + (h_7x + h_8y + h_9)u = 0 \tag{3.10}$$

$$-h_4x - h_5y - h_6 + (h_7x + h_8y + h_9)v = 0 \tag{3.11}$$

Equations 3.10 and 3.11 can be written in matrix form as

$$A_i h = 0 \tag{3.12}$$

with

$$A_i = \begin{pmatrix} -x & -y & -1 & 0 & 0 & 0 & ux & uy & u \\ 0 & 0 & 0 & -x & -y & -1 & vx & vy & 1 \end{pmatrix} \tag{3.13}$$

$$h = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & h_7 & h_8 & h_9 \end{pmatrix}^T. \tag{3.14}$$

Since each point correspondence provides 2 equations, 4 correspondences are sufficient to solve for the 8 degrees of freedom of H. Figure 3.13 shows an example of defining four-point correspondences for conversion between image and world GPS coordinates.

Chapter 4

Behavior Analysis

This section provides different methods for activity analysis of vehicles and pedestrians in order to provide measurements of capacity, and delay. These measurements are used for signal design and intersection planning. Traffic cameras are utilized through vision-based measurements of intersection counts (e.g., turning movements count, crossing count), speed, waiting time, and queue analysis.

4.1 Vehicles

Turning movement recognition, count and speed estimation are three important steps toward better understanding turning characteristics, signal and intersection design. In order to streamline intersection analysis, fully automated systems that can accurately estimate turning movements' count, speed and waiting time is highly desired.

4.1.1 Turning Movement Count

Traditionally, the Turning Movement (TM) counts were obtained by technicians who manually count the number of vehicles. Technicians must physically observe a location, typically during A.M. and P.M. peak hours, and record the number of vehicles as they pass either on paper or using a dedicated device. In addition, larger or more complicated intersections with high volume may require multiple observers. Therefore, robust automated counting methods are needed because manual counting requires significant cost in general and becomes almost impossible for large projects [146, 147].

TM count systems should provide high accuracy as well as long time operation to provide

average daily traffic patterns for road design [12]. Most TM counting systems use the same basic zone definition framework where an intersection is divided into legs, and they count a turn based on the zones traversed [31, 148, 149]. For example, if zones of south and east legs are traversed by a vehicle, TM count of south to east is increased. The SCOCA system [31] performs hybrid region and feature-based tracking using background subtraction which is still a common method for moving object detection in video surveillance systems. TM counts are collected based on enter and exit zones of the trajectories. Zone techniques are simple but are ineffective when a vehicle track is lost, resulting in poor zone localization and low accuracy in TM counting.

Recognizing TMs has been addressed by some non vision based monitoring systems [23, 52, 54] that record trajectories by ego-vehicle equipped with some sensors such as global positioning system (GPS) and laser scanner. Turning trajectories are learned through different methods such as regression models, dynamic time warping (DTW) and longest common subsequence (LCSS). For example, turnings are learned in [23] using Gaussian regression models from collected trajectories. Kafer et al. [52] shows a way of predicting future motion trajectories which are probabilities of turning left, right and straight. Prediction values are extracted using trained paths and finding the best match by the Quaternion-based Rotationally invariant Longest Common Substring (QRLCS) method. In [54], the ego-vehicle predicts observed vehicle behavior, like stopping or turning, and then trajectories are saved into database by on-line update. While these systems recognize turning movements, speed estimation of each turning movement is missed. In following sections, we describe our method of turning recognition, and turning movement count which benefits high accuracy in comparison with traditional count methods.

Turning Recognition & Count Method

Turning movements are recognized and counted through cooperation of two different modules named zone and trajectory comparison modules. The typical zone comparison module counts based on predefined image regions traversed by a vehicle. Unfortunately, this simple method fails in the following situations [12]:

1. When a track is incomplete as might occur when two vehicles move from a stop bar in unison. When they get separated, of the two resulting trajectories, only one is complete (the second does not begin before the stop bar).

2. When the position of the tracked vehicle falls into an undefined region due to various noise

$$LCSS(F_i, F_j) = \begin{cases} 0 & T_i = 0 | T_j = 0 \\ 1 + LCSS(F_i^{T_i - 1}, F_j^{T_j - 1}) & d_E(f_{T_i}, f_{T_j}) < \epsilon \ \& \ |T_i - T_j| < \delta \\ \max(LCSS(F_i^{T_i - 1}, F_j^{T_j}), LCSS(F_i^{T}, F_j^{T_j - 1})) & \text{otherwise} \end{cases} \tag{4.2}$$

sources like poor background subtraction.

3. When two objects get merged and a track gets broken due to occlusion or blob merging [150].

The trajectory comparison module complements the zone module by addressing these situations by finding the most probable path based on similar distance between the observed partial trajectory and the typical paths of the intersection.

Path labeling is the process of path recognition for a track using zone and longest common subsequence (LCSS) techniques. While zones work based on vehicle traversal over predefined intersection areas, LCSS is also utilized as a path similarity measurement due to its robustness to noise, outliers and good performance for two unequal paths [12]. Instead of a one-to-one mapping between all points in two trajectories, some points with no good match can be ignored. The LCSS distance between a trajectory and path (parametrized by a prototype trajectory) is computed in (4.1),

$$D_{LCSS}(F_i^{T_i}, F_j^{T_j}) = 1 - \frac{LCSS(F_i^{T_i}, F_j^{T_j})}{min(T_i, T_j)} \tag{4.1}$$

where T_i is the length of trajectory F_i. The LCSS computes number of matching points between two trajectories and it is defined in (4.2). $F^t = \{f_1, ..., f_t\}$ denotes the trajectory points (vehicle centroids) up to time t. The compared points of two trajectories should be within a small Euclidean distance ϵ and points should not be separated by more than δ samples to ensure the lengths are comparable.

Zone Comparison Module

The four cardinal directions {north, south, east, west} along with center of the intersection specify the predefined regions, called zones, of an intersection image. The zones are contextually defined

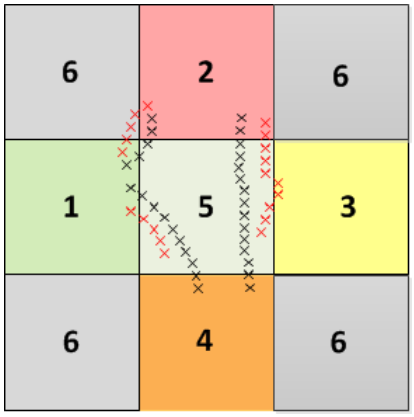


Figure 4.1: Complete tracks (black) versus incomplete or undefined tracks (red). The red trajectories cannot be accurately counted using simple zone comparison

based on intersection legs and center. A prototype intersection is shown in Figure 4.1.

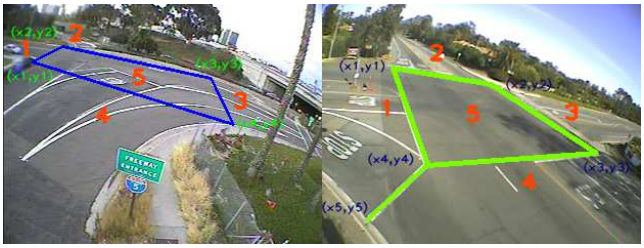


Figure 4.2: Contextual definition of zone areas for INT 4 and INT 5

The zones are used to define a Regular Sequence (RS) set, which is, the set of acceptable zone traversals. For example, {1,5,4} indicates a westbound right turn while {1,5,3} a westbound through. Traversed regions by vehicles are recorded, which show the transitions between zones, to build the track zone sequence. Upon completion of the track, its associated zone sequence is compared against the regular sequence set. If the resulting zone sequence exists in the regular sequence set of the intersection, a counter for the associated TM count is incremented. There is a zone flag that gets set to indicate counting success by zone module which is used as a signal for the trajectory module.

Figure 4.2 depicts two examples of defined zones (shown in orange colored numbers) for two intersections. Five regions are defined using four coordinates for both intersections. The second

Table 4.1: Regular Sequence Set for Turning Movement Directions

	WBL	WBT	WBR	NBL	NBT	NBR	EBL	EBT	EBR	SBL	SBT	SBR
INT 4		{1,5,3}	{1,5,4} {1,4}	{2,5,3}	{2,5,4}							{4,5,3} {4,3}
INT 5	{1,5,2}	{1,5,3}	{1,5,4} {1,4}	{2,5,3}	{2,5,4}	{2,5,1} {2,1}	{3,5,4}	{3,5,1}	{3,5,2} {3,2}	{4,5,1}	{4,5,2}	{4,5,3} {4,3}

intersection (INT 5) requires an extra line to distinguish the right-turn in the lower left of the image. The complete RS set is shown for two intersections in Table 4.1. Notice not all TMs have a sequence and there is some with two zone sequences (those not required to go through center zone) due to the intersection configuration.

Trajectory Comparison Module

Although counting by zone comparison is simple, it fails when the track zone sequence is not a member of the RS. Figure 4.1 depicts some scenarios when zone comparison fails. The black trajectories are complete and can be counted using zones. The red trajectories have issues and can not be counted by the zone module. The sequence of top left trajectory, $\{1, 6, 2\}$, has an undefined region 6. The bottom left sequence $\{1, 5\}$ stops in zone 5 and never goes to an exit zone. Finally, the rightmost track has obtained sequence $\{5, 3, 5, 2\}$, due to occlusion and noisy measurements.

If the zone flag is not set and RS has not been used by zone comparison module, the trajectory comparison module is utilized for counting. This module examines the entire trajectory and compares it with typical paths in the scene. The typical paths could be learned in different ways such as through clustering observations [151] or even predefined by hand drawing, similar to zone definition. In this work, the first complete trajectory that traversed each of the entries in the RS set are defined as the TM path for simplicity.

When the zone module fails, the trajectory comparison module is triggered and it compares the trajectory with all the stored typical paths of the intersection. The path with the smallest D_{LCSS} value is considered the best match and used to increment the related TM counter. Examples of the intersection paths are shown in Figure 4.3. The paths are color-coded on their approaching zone. So, all path starting in a particular zone (heading direction) have the same color.

4.1.2 Turning Speed

Although speed is a common way of analyzing pedestrians' crossing and walking behavior [81], it can be also used for vehicles as well to understand the safety effects of behavior such as speeding,

(a) (b) (c)

Figure 4.3: Typical paths: a) INT 4, b) INT 5, c) INT 6

stopping and slowing. For instance, high vehicle acceleration behavior during left turns is unsafe for pedestrian crossing.

Most vision-based systems rely on motion for vehicle detection and tracking, e.g. optical flow [80] or background subtraction [123, 152]. However, they are not appropriate for vehicle detection at intersections since they stop temporarily due to traffic signal or yielding the way [14]. A path reconstruction technique is introduced to address the common problem of stopped objects being learned into the background.

Speed measurements are calculated using trajectories and a model of speed patterns is constructed for each typical path. The speed profile, which has already been proposed for highways in [151], is modified for intersections as presented in (4.3),

$$speed\ \ profile = \begin{cases} Stopped & V \leqslant \alpha_1 V_t \\ Slow & \alpha_1 V_t < V \leq \alpha_2 V_t \\ Normal & \alpha_2 V_t < V \leq 1.1 V_t \\ Speeding & V > 1.1 V_t \end{cases} \tag{4.3}$$

where V_t is an average speed and $0 < \alpha_1 < \alpha_2 < 1$. The α parameters are chosen empirically based on intersection settings since it varies based on the number of vehicles which stop at each leg behind red signals leading to a huge plunge in average speed.

The frequency of speed belonging to each category $\{i.e.\ Stop, Slow, Normal, Speeding\}$ is computed for each track. Since turning movements are recognized during the run time, speed behavior is estimated implying whether characteristics of turning movements are abnormal or not.

Path Reconstruction

Path reconstruction is a technique used to solve the disappearance problem of vehicles that come to a stop behind the stop bars at intersection. When a vehicle stops, its track disappears and when it starts moving it appears as a new track. The idea is to recognize those tracks that belongs to a same vehicle and then connect them. Path reconstruction first finds incomplete tracks by context and then it finds a match based on spatial, temporal and appearance proximity for signature coherence. Since one object might move and stop several times, a multi-level approach is needed. These steps are described with more details in following subsections.

Finding Incomplete Tracks at Stop Areas

First step, the path reconstruction module determines those tracks that are suspected to be incomplete by using stop areas as clue. Stop areas can be easily distinguished since zone areas have already been contextually determined and mapped by numbers for turning recognition and counting purposes (see Figure 4.2). Incomplete tracks, which end up at stop areas, are determined if their track zone sequence is not a member of RS. For example, a track with one zone digit indicates the disappearance of a vehicle at a stop area since it starts and ends up in the same area.

Spatial Proximity

After finding track candidates (incomplete tracks), the last and first points of each track are examined. Two tracks T_i and T_j are spatially connected if their Euclidean distance is less than a threshold,

$$T_m = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2)} < \delta_1 \tag{4.4}$$

where (x_i, y_i) is a last location point of T_i and (x_j, y_j) is a first point location of T_j and δ_1 is the threshold value.

Temporal Proximity

Since tracking is performed in small time windows, candidates of incomplete tracks should be evaluated based on their disappearance and appearance time. In addition to each track's location, its appearance and disappearance frame number is also recorded. Two tracks can be connected

over time if the difference between frame numbers F_d of disappearance (first track) and appearance (second track) is less than a threshold,

$$0 < F_d = (\#F_f(T_i) - \#F_l(T_j)) < \delta_2 \tag{4.5}$$

where $\#F_i(T_i)$ and $\#F_l(T_j)$ are two frame numbers indicating disappearance and appearance time of two tracks (i.e. T_i, T_j).

Appearance Proximity

The appearance of an object is an important feature that can be used as a similarity cue. Since color [153] and Histogram of Oriented Gradient [29] are two important features used for object detection and recognition in computer vision, the similarity of two objects are determined by both methods. In addition to frame number, color distribution and HOG features of tracks are estimated for appearance and disappearance times. Suppose $S_c = p(match|color)$ and $S_{HOG} = p(match|edge)$; the similarity of two tracks are evaluated based on the appearance score defined in (4.6),

$$S_a = P(match|edge, color) = S_c \times S_{HOG} \tag{4.6}$$

where S_c is the color score and S_{HOG} is the HOG score.

Histogram of Oriented Gradient feature: HOG is a popular edge-based feature used for object recognition in computer vision. The HOG feature [29] proposed for pedestrian detection which is normally a difficult task since pedestrians vary in size and pose. First, the gradients are calculated in blocks and then the histograms of gradient magnitudes are accumulated into orientation bins.

Since all image samples are resized to 32×32, 9 blocks with 50 percent overlap will include 4 cells of 8×8 pixels. The final feature vectors, $F = \{f_1^u\}$ and $Y = \{f_2^u\}$ are obtained for $u = 1, 2, ..n$ ($n = 3 \times 3 \times 9 \times 4 = 324$). S_{HOG} is calculated between appearance and disappearance of two tracks by plugging the Euclidean distance of HOG feature vectors into the Gaussian kernel as shown in (4.7),

$$S_{HOG} = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{\sqrt{\sum_{i=1}^{n}(f_1^i - f_2^i)^2}}{2\sigma^2}\right). \tag{4.7}$$

Color feature: Color is another important feature used for object recognition for similar shapes and it is most effective when combined as an attribute with other features in a fusion technique [154]. The color distribution is calculated to improve appearance match accuracy of HOG since vehicles might look similar based on gradient information only.

At first, each track bounding box is extracted from an image and then for each R, G and B plane, the color histogram of an object is computed in 64 bins using the foreground detection mask. The foreground mask is used as map to distinguish the color distribution of an object from the background in an image aligned bounding box. A feature vector including 192 values is created for each track and is normalized to estimate the color probability density function. If $P = \{p^u\}$ and $Q = \{q^u\}$ are first and last histograms of two tracks (i.e. T_i, T_j) with $u = 1, 2, ..m$ bins, their similarity is evaluated using Bhattacharyya distance defined in (4.8),

$$d_c = \sqrt{1 - \rho\left[P, Q\right]} \tag{4.8}$$

where

$$\rho[P, Q] = \sum_{i=1}^{m} \sqrt{p^u q^u}. \tag{4.9}$$

Larger ρ indicates greater similarity of color distributions. $\rho = 1$ for a perfect match between two identical histograms. The color score is also defined in (4.10),

$$S_c = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{d_c^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{1 - \rho[P, Q]}{2\sigma^2}\right). \tag{4.10}$$

4.1.3 Waiting Time

Waiting time is an important characteristic of vehicles implying the state of being inactive due to the red signal phase at intersections. Vehicle waiting time is usually used for some reasons as below:

1. Vehicle waiting time could be used for safety through signal design at unsignalized intersections. Its optimization maximizes the throughput of traffic by designing signals in a way that minimizes pedestrians' and vehicles' idle time.

2. It can contribute to intersection danger due to driver's becoming impatient in a queue to cross.

Existing methods mostly rely on manually extracting pedestrians' waiting time and they do not address vehicles while taking turns.Since turning movements have been recognized, the total number of frames that vehicles are stopped based on current speed (i.e. $V \leqslant \alpha_1 V_t$) during each turning movement can be estimated by counting frames in different speed profiles (4.3) during tracking. Path reconstruction plays a key role for waiting time estimation since it finds two incomplete tracks and the number of disappearance frames between them during a stop.

4.1.4 Waiting Time Factors of Right Turns

The contributing factors for vehicle waiting time were evaluated using turning movement count information. In particular, right turns were examined since they can be taken during red phase when available gap is safe. The waiting time of vehicles that take a right turn were evaluated against different opposing TMs to determine which contributed significantly using a linear regression model.

Waiting time (WT) is defined as function of turning movement counts (TMC) in Eq. (4.11),

$$WT_\omega(TMC) = \omega^T \times TMC \tag{4.11}$$

where $TMC = \begin{bmatrix} 1 \\ TMC_1 \\ . \\ . \\ . \\ TMC_n \end{bmatrix}$, and $\omega = \begin{bmatrix} \omega_0 \\ \omega_1 \\ . \\ . \\ . \\ \omega_n \end{bmatrix}$.

TMC and ω are two $(n+1) \times 1$ vectors used in the linear regression model to predict the waiting time and ω is the weight vector indicating coefficients. The maximum number of n is 12 and weight vector is estimated by minimizing a cost function over training data set using the Gradient Descent algorithm. The cost function is defined in (4.12),

$$j(\omega) = \frac{1}{2m} \sum_{i=1}^{m} (WT_\omega^i(TMC) - y^i)^2 \tag{4.12}$$

where m is the total number of elements in training data set and y is the extracted waiting time. i refers to ith row of the data set and m refers to number of rows in dataset. In order to build the dataset, one zone area of an unsignalized (INT 5) and signalized (INT 6) intersections were observed and waiting time of vehicle tracks along with TM counts during their life time was recorded.

4.1.5 Queue Analysis

Vehicular traffic data such as flow, speed and density is an important criterion used to design intersections and boost safety. Another important parameter is the vehicle queue analysis used in control models to improve the passing capacity. Moreover, queue length estimation and associated delay are useful for devising traffic management strategies that would help to optimize traffic signals and improve the performance of a traffic network.

Data collection is the first step towards queue length estimation, which is usually performed by the loop detectors [155] and video cameras [156, 157] at junctions. Manual methods require human observation to collect data which is difficult for long time evaluation. Vision-based methods are the one of the most preferred automatic methods due to their low cost and high support of different real time applications (e.g. vehicle count, speed and classification). Long time data collection of traffic flow with high speed is an essential requirement in Intelligent Transportation Systems (ITS).

Vision-based queue analysis can be performed by two major groups of tracking and non-tracking methods. Non-tracking methods determine the existence of vehicles based on different introduced features on the road. Local Binary Pattern (LBP) [156], spatial edges [158], Fast Fourier Transform (FFT) [158], and image gradients [159] are the important features used to detect stopped vehicles in the literature. As an example of other features, the entropy method is proposed [160] to detect stopped vehicles and Harris corner features [157] are useful in detecting stopped vehicles when they build a queue.

Although queue analysis by non-tracking methods is simple, its application is limited to queue length estimation. However, tracking methods can provide other important vehicular traffic data such as speed, count, waiting time and time headway. A tracking method is appropriate for queue analysis if it addresses the two below problems.

1. Stopped vehicle detection is difficult since motion is mostly used as a cue in video surveillance [14]. Moreover, detection by motion leads to occlusion for slow moving vehicles in the line.

2. Tracking of stopped vehicles is a difficult task since motion-based tracking methods like optical flow work poorly for stationary objects.

Tracking methods rarely address queue analysis. For example, tracking of vehicles on the head and tail position of the line is estimated in [161] and other techniques aim to detect and track stopped vehicles [162, 163] like parked vehicles [164]. Tracking methods based on optical flow are presented in [165] and [32] to provide robust vehicle tracking at intersections. However, there is no effort towards estimating queue length and waiting time of separate vehicles in a line.

This section presents a method of queue analysis by estimating queue length, vehicle numbers in the queue and their associated delays (i.e., waiting times). The proposed method solves formerly mentioned problems since detection by motion is only performed over the small areas which initializes tracks (see motion area defined in Figure 3.11b). Then, optical flow tracking handles the partial occlusion between slow moving vehicles in the line. When the waiting state of a vehicle is determined, an appropriate bounding box estimation is performed to handle the possible failure of tracking by optical flow due to reduction of feature points.

Queue analysis includes estimating the number of vehicles in the queue, their waiting time and queue length estimation. Since a tracking method is used, vehicles' waiting time are readily estimated. Three steps are needed to perform, explained below.

Path recognition

Vehicles paths are recognized using temporal alignment techniques for similarity measures of trajectories with typical paths which have been collected for each lane (see Figure 3.12b). Longest common subsequence (LCSS) distance is a popular technique for comparing unequal length trajectories [12, 166].

Waiting state detection

Tracks are labeled regarding each lane and their moving or waiting state is determined at each frame based on the state diagram shown in Figure 4.4. When the absolute waiting state of vehicles is determined, they become candidates for queue length estimation if there is spatial proximity between them. The track candidates are saved into separate lists regarding each lane for queue length estimation.

The moving/waiting state detection [166] is shown in Figure 4.4. Two displacement thresholds

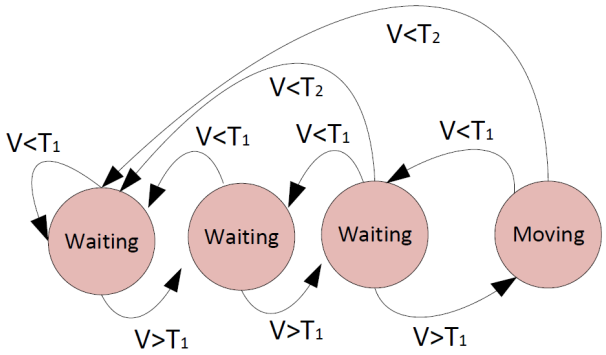


Figure 4.4: The diagram of waiting/moving state estimation

T_1 and T_2 are considered as soft and hard thresholds to determine waiting state of a track ($T_1 > T_2 > 0$). When the sequence of "waiting" state happens for a track, its state doesn't change to "moving" by one displacement measurement (e.g. $V > T_1$), which could be noisy. The very small value of the displacement vector ($V < T_2$) implies the waiting state of the vehicle that only can accept transition to moving state by sequence of displacement vectors higher than the soft threshold ($V > T_1$).

Queue length estimation

Queue length is gauged for waiting vehicles of each lane using feature points of the tracks. Texture and corner feature points of stopped vehicles are used to estimate the queue length. Since vehicles' queue lines might have different orientations, the line is projected into each lane by selecting (x, y) coordinates of feature points according to highest and least y values. Two selected feature points find the nearest neighbor coordinate from typical paths and the Euclidean distance is used to measure the distance.

Figure 4.5 shows an example of the queue length estimation using feature points. The estimated lines (red line) are shown for two queues with different orientations. As shown in Figure 4.5, one vehicle has missed the track since it has not come into the motion area to get initialized from the beginning. Note that, the bounding box of the vehicles closer to the stop bar (second and third lanes) are smaller since there is possibility of occlusion with crossing pedestrians, as discussed in Section 3.2.1.

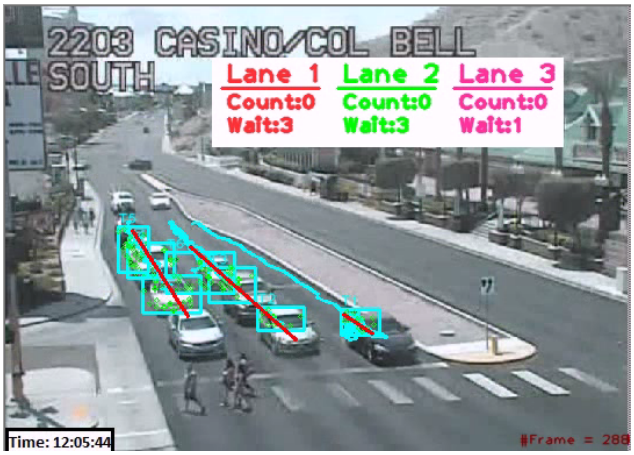


Figure 4.5: The queue length estimation (red line), detected stopped vehicles (aqua bounding box) and feature points (green points)

4.2 Pedestrians

4.2.1 Crossing Count

Automated count of non-vehicular participants include pedestrians and bicycles for design of crosswalks and bicycle routes or trails. Pedestrian counts are also required for designing facilities at crossing points, such as signals and timers. For this purpose, pedestrian counts will be carried out over some length on either side of the proposed crossing point (e.g. along a crosswalk). Similar to a TM count, typical trajectories are modeled along the crosswalk. Unlike for vehicles, non-vehicular traffic is typically bidirectional (e.g., right-to-left and left-to-right) which means that typical paths must be defined in both directions of a crosswalk in order to keep track of direction of travel. The same LCSS trajectory comparison technique used above for vehicles can be implemented for non-vehicle trajectories.

Generally pedestrian counts are more difficult to obtain than vehicle counts since detection and tracking is more difficult with traffic cameras. Pedestrians are small in size and they tend to walk in group causing heavy occlusion which degrades tracking performance and undercounts the pedestrians in a group. Some efforts have addressed these shortcomings by avoiding tracking all together and instead learning to count groups using dynamic textures. For instance, the foreground object is extracted using background subtraction and set of features such as segment, edge and texture are measured and learned according to number of pedestrians using learning algorithms [167] and regression models [5]. Figure 4.6 shows an example of the pedestrian counting system [5].

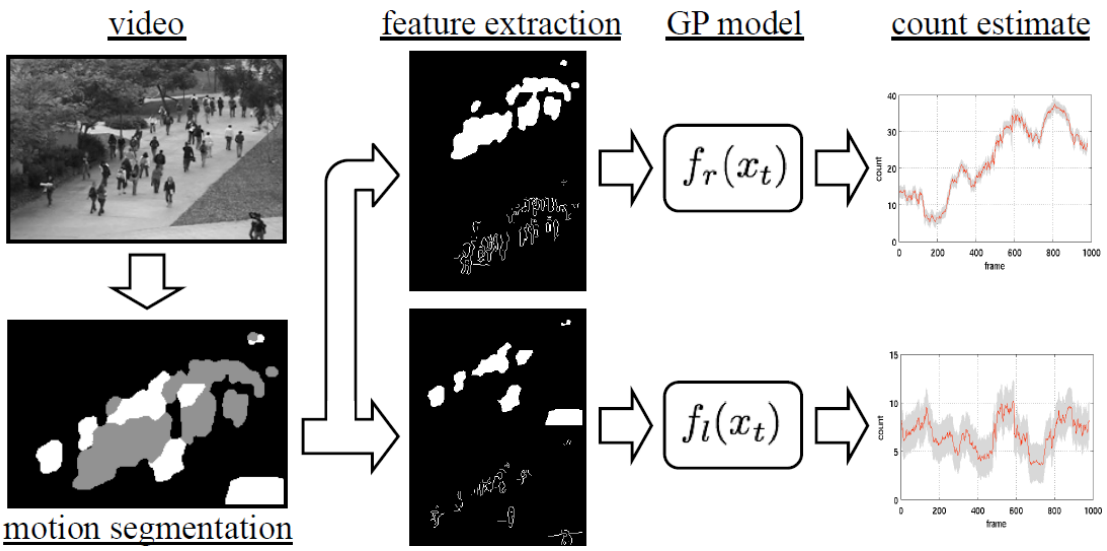


Figure 4.6: Crowd counting system [5]

In this dissertation, pedestrians are counted using the LCSS method. In order to match a typical path, a fraction of data points must match. As a result, broken trajectories could be counted if they still have enough matching points with a typical path. Crossing count results are compared with ground truth to determine the counting accuracy rate as defined in Eq. (6.4). The typical paths for pedestrians are shown in Figure 3.12 for three different intersections. We improve detection and tracking of pedestrians through contextual fusion of motion and appearance as it was discussed in Section 3.1.3. Different classifiers are prepared and an appropriate one is selected through its comparison with other detectors based on speed and accuracy. Finally, the cooperation between optical flow and bipartite graph provides robust tracking presented in Section 3.2.1.

4.2.2 Crossing Speed

In addition to the counts, non-vehicular crossings are analyzed further for crossing characteristics. The trajectories are further analyzed to extract information such as crossing speed since this directly relates to the functionality of an intersection. For instance, a crosswalk countdown timer may be extended if pedestrians are not reliably able to cross in time. In one of the typical tracking based system (i.e., PedTrack [80]), moving objects are detected using background subtraction technique and potential pedestrians are determined by their characteristics, such as the size and width over

height proportion. An inherent cost function is adopted to track subsequent potential objects based on their attributes of size, height, width, and gray scale color distribution, and occlusion reasoning is performed based on reasoning through splitting and merging events. The waiting zones are defined over the scene to register entered pedestrians and estimate their arrival time on the other side. The goal of the system is to track pedestrians in complete crossing events (waiting and completing the crossing from one registration line to another). In this dissertation crossing speed is obtained by investigating trajectories of the pedestrian tracks and converting that to world coordinates using Homography matrix.

4.2.3 Waiting Time

There are quite a few works that study pedestrian behaviors at intersection by extracting their waiting times since it has been shown that a long waiting time results in impatience to cross [74]. In a non-vision-based work, pedestrian waiting time is modeled by using risk function [74] since they take a risk to cross the intersection. Malinovski et al. [80] uses background subtraction to automatically detect pedestrians and calculate their waiting time and arrival time by examining the pedestrians tracks. Hamed et al. [74] found in 2001 that long waiting time resulted in impatience to cross which could lead to undesirable behavior.

The waiting time is a fundamental measurement for signal design and it has been shown as a criteria which affects safety as well. Given a pedestrian trajectory, the waiting time is calculated as the total time that the pedestrian has a speed value lower than some predefined speed threshold. At each frame, the instantaneous speed is used to characterize the waiting or walking state of a pedestrian. This is important when calculating the average pedestrian walking speed since waiting frames should be excluded from the calculation.

Fully automated waiting time estimation is difficult since waiting pedestrians lack the necessary motion required for successful detection and tracking in traffic video. With major advances in pedestrian detection [143], the use of both appearance and motion cues has become a promising methodology even in lower resolution video. In this dissertation, pedestrians trajectories are investigated and their crossing speed is obtained. They are marked as waiting if their speed is lower than a predefined threshold. The count of waiting state over time frames indicates waiting time.

Chapter 5

Safety Analysis

Traditionally, accident statistics were used to address a range of safety-related concerns, such as identifying hazardous locations, and evaluating driver behaviors or safety programs. Although accident data can be proven to be a useful metric for safety evaluation, there are several serious limitations in its usage [98]:

1. An occurrence of an accident is an outcome of a complex process of interactions (e.g., driver interaction with his vehicle and the road environment) which make safety analysis a difficult task such as finding the main causes of accidents just from accident counts alone.

2. Although number of deaths from road accidents is high, their frequencies segregated by locations, time and type are generally low. Given this low rate of occurrence and the statistical nature of the problem, the task of statistical inference by merely examining accident counts is a difficult task.

3. Accidents are not always uniformly reported and this can hamper good comparative analyses. Indeed many accidents, especially those not involving any injury may not be reported at all. As a result, accident reports can sometimes provide biased conclusions.

In order to overcome the mentioned shortcomings, many ways of employing non-accident data have been suggested. One of the more recently used forms of non-accident information is traffic conflicts, which are defined as critical incidents not necessarily involving collisions. A conflict is defined as an observable situation in which two or more road users approach each other in time and space for such an extent that there is risk of collision if their movements remain unchanged [99].

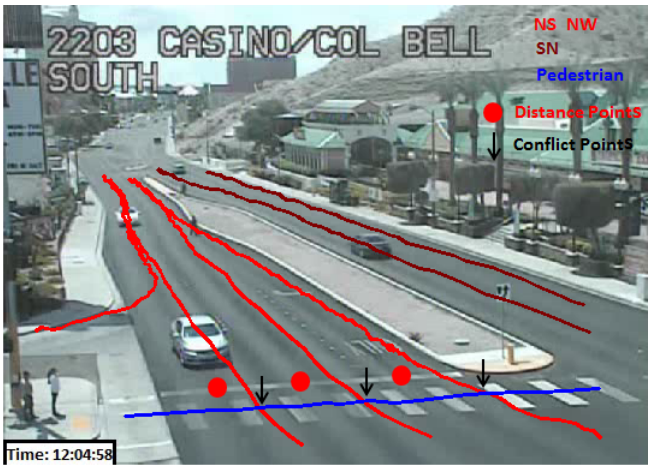


Figure 5.1: Distance points regarding each lane are shown with red cirles on the stop bar.

This method has a long history of development including research on recommended data collection methods, and still continuing to be used to rank locations with respect to safety for construction upgrades.

Although surrogate safety measurements have been studied by transportation engineers, there is a lack of vision-based framework to provide comprehensive analysis. Most techniques rely on estimating TTC and other measurements have not been studied thorough computer vision techniques. Since a vison-based tracking system is developed to provide long term trajectory of participants (i.e., waiting and moving vehicles), some other important measurements such as DTI and TTI are investigated as well.

5.1 Distance and Time to Intersection

TTI and DTI are two important parameters at intersections that are calculated based on vehicle distance to stop-bars [168]. These two parameters provided the needed information to perform a safe maneuver when a vehicle is approaching a dilemma zone and estimating the appropriate gap [25]. The first step is robust tracking of vehicles and then path recognition provide a lane number for each vehicle since different distance points are considered regarding DTI, TTI calculation. The stop-bar distance points are shown in Figure 5.1 by red circles.

DTI and TTI are used in different applications regarding traffic safety. They are the criteria to analyze behavior of approaching vehicles form the major legs of an intersection which are usually

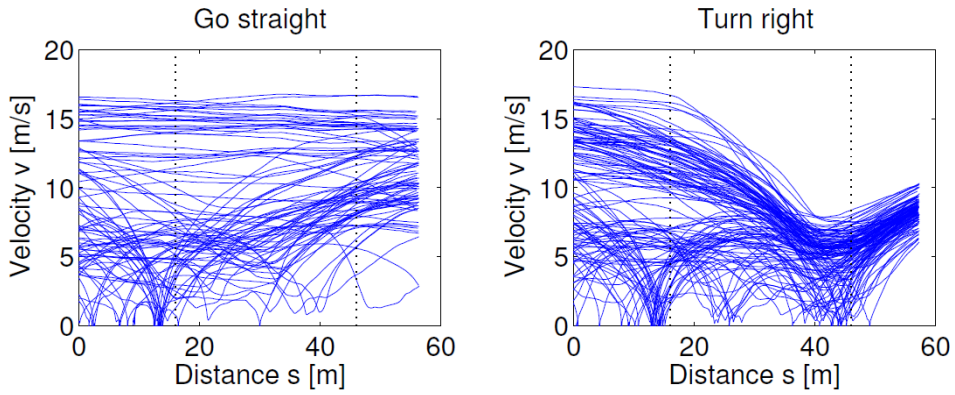


Figure 5.2: Velocity profiles of vehicles for going straight and turning right. The position of the stop line and that of the pedestrian crossing are indicated by dotted lines [6, 7].

tracked through intersection decision support systems. Intersection decision support systems help driver to take the safe decision for crossing the intersections. Another important usage of DTI is to obtain models for approaching vehicles which can help for their future trajectory prediction. Vehicles approaching an intersection follow a certain model that helps to determine their abnormal behavior shown in Figure 5.2.

5.2 Time to Collision (TTC)

TTC is calculated based on the predicted arrival time and it is defined as the time for two objects to collide if they continue with their present speed on their paths. Each partially observed trajectory of vehicles is compared against typical paths to find its most probable path and its associated conflict point. The time to conflict point for a vehicle is compared to those pedestrians that are moving toward it and the minimum TTC value is counted if both timings are in a same window [168].

5.2.1 Trajectory prediction

If a real-time system can accurately predicts accidents in advance, a warning signal can be generated on time, and many traffic accidents could be avoided. The vision-based system is able to quantify safety if it can robustly track road users, and provide reliable predictions about the future trajectories. The future predictions help to determine their probability of being involved in a collision.

It is simple to predict future trajectories of a typical object using its dynamics but prediction will probably be noisy due to light, shadow and occlusion effects. Predictions could be more accurate if there is a secondary available data regarding future locations. Since there is no other sensor available to infer the secondary data in videos, the history of the trajectories can be used to correct future predictions. As suggested by Hu et al. [109], history of motion patterns can help to find the probability that an observed trajectory of a track belongs to each pattern. For example, by finding a vehicle position $A(x_1, x_2)$ and direction of motion $(\delta x_1, \delta x_2)$, four lines around the vehicle could be predicted for the vehicle. if any two surrounding lines of vehicles A and B intersect in future, the accident is detected.

In this paper, another way of prediction is presented using vehicle dynamics and learned paths. The Kalman filter uses current velocity of an object for future predictions, and the learned typical paths are used as measurements to correct the predictions. An observed trajectory find its typical trajectories match using LCSS method. Suppose there are N typical paths for vehicles and pedestrians. The proposed method using object dynamic for prediction and nearest neighbor trajectory of found model for correction as shown in Algorithm 1. T is an observed trajectory of a track and F_{index1} will be a found match from typical path models. Then, the nearest pair $(F_{index_1,index_2}, F_{index_1,index_2+1})$ of the model (i.e. F_{index1}) to the predicted location is found to set observation matrix Z_k and correct measurements using Kalman filter.

5.2.2 Conflict point inference

When future trajectories of vehicles and pedestrians are predicted using Algorithm 1 as sequence of (x, y) points, the conflict point of predicted trajectories should are inferred investigating each two points on lines that are closed enough in distance. Finally, their average is used as an estimate of conflict point. After inferring the conflict point, the time to conflict of each participant is calculated for each road user. The TTC is indicated when absolute difference of time to conflict points are less than a second.

Algorithm 1 The proposed prediction method of future trajectories using Kalman filter

function PREDICTION
  $dist_1 = D_{LCSS}(T, F_1)$;
  $index = 1$; ▷ $index_1$ will refer to highest probable path with the track trajectory.
  for $i = 2$ to N **do**
    if $(D_{LCSS}(T, F_i) < dist_1)$ **then**
      $dist_1 = D_{LCSS}(T, F_i)$;
      $index_1 = i$;
    end if
  end for
  $A = [1\ 0\ \Delta t\ 0; 0\ 1\ 0\ \Delta t; 0\ 0\ 1\ 0; 0\ 0\ 0\ 1]$, $H = [1\ 0\ 0\ 0; 0\ 1\ 0\ 0]$;
  $X_K = Kalman.predict(A, X_{K-1})$;
  $\delta = \sqrt{(X_{k,1} - F_{index_1,1})^2 + (X_{k,2} - F_{index_1,2})^2}$;
  $index_2 = 1$;
  $M = Length(F_{index_1})$; ▷ $(index_2, index_2 + 1)$ will be the closest pair to track location X_k.
  for $j = 3$ to M **do**
    $dist_2 = \sqrt{(X_{k,1} - F_{index_1,j})^2 + (X_{k,2} - F_{index_1,j+1})^2}$;
    if $dist_2 < \delta$ **then**
      $\delta = dist_2$;
      $index_2 = j$;
    end if
  end for
  $Z_k = [F_{index_1,index_2}, F_{index_1,index_2+1}]$; $X_K = [X_{k,1},\ X_{k,2},\ \dot{X}_{k,1},\ \dot{X}_{k,2}]$;
  $X_K = Kalman.estimate(Z_K, X_K)$
end function

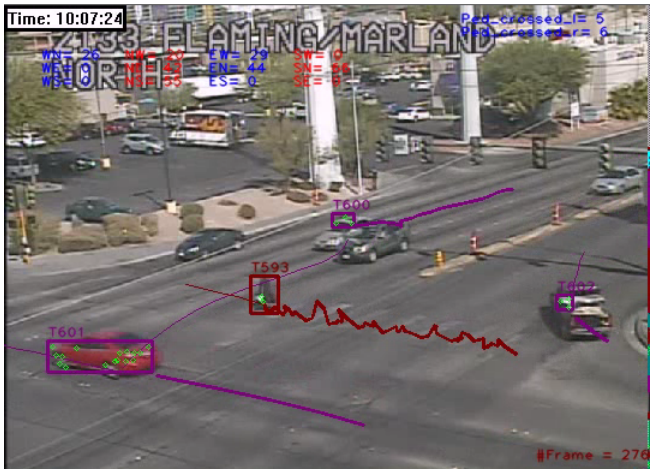


Figure 5.3: The snapshot from the output of video, tracks and prediction trajectories are shown with thick and thin lines.

Figure 5.3 shows an output of the tracking system including trajectories of tracks (i.e. thick

lines), prediction paths (i.e. thin lines) and conflict point inference. Vehicle and pedestrian tracks are shown with magenta and red colors respectively. The trajectory of a pedestrian is highly noisy which causes noisy prediction for its future trajectories. However, the system has efficiently predicted the future trajectory (i.e. thin red line) of the pedestrian movement using the proposed method.

5.3 Post Encroachment Time (PET)

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. Based on the definition, future trajectory prediction is not required to estimate PET and reliable trajectories of road users, which were in conflict, can greatly help to estimate PET. As a result, the proposed system utilizes the observed trajectories to estimate PET after the conflict event happens.

5.3.1 Trajectory observation

In order to estimate PET, the assessment of reliable trajectories can help to estimate the amount of time required for second road user to reach to the conflict point. Typical paths have already been used by the proposed system to provide different applications (e.g., turning movement count, pedestrian crossing count), and counted pedestrian and vehicles indicate the reliable trajectories. The system assess the counted trajectories of pedestrians and vehicles and then it reconstruct the conflict event based on trajectories and time information.

5.3.2 Temporal and spatial conflict

The PET inference is based on finding road users that have temporal and spatial conflict. The system evaluate the trajectories of counted pedestrians and vehicles that have time overlap. For example, a vehicle track starts at frame 100 and finish by frame 500 has timing overlap with a pedestrian starts at frame 300. The start frame of a road user needs to be checked against the ending track frame of other users to find the tracks that have temporal conflicts.

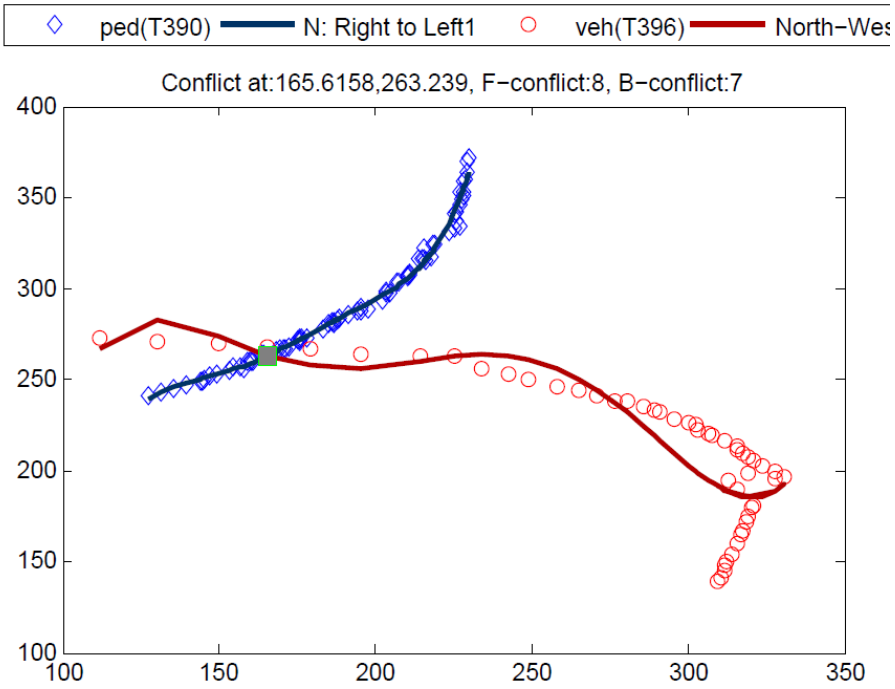


Figure 5.4: The regression lines passing vehicle and pedestrian trajectories show their spatial conflict.

The second step is finding the trajectories that have spatial conflict. The trajectories are evaluated using regression models and a smooth curve passing them is estimated. The intercept of two regression lines show the spatial conflict and then time frame of each road user at conflict point can be estimated. Figure 5.4 shows two vehicle and pedestrian with temporal conflict have spatial conflict as well since their estimated regression lines intercept each other. The vehicle is taking right turn and a pedestrian crossing the intersection.

Since the time frame information for each trajectory point is recorded, the closest trajectory points to intercept point and time of conflict is recovered. Finally, the time for the second road user which reach to the conflict point can be obtained. The proposed method reconstructs the conflict event after they occur using their real trajectories. The method is better suited than a prediction method for estimating PET.

Chapter 6

Results & Discussion

6.1 System Evaluation

The system is developed and evaluated to work on both pedestrians and vehicles. However, the system is developed and dedicated for pedestrian behavior analysis at second phase by merely addressing pedestrians to provide higher speed and reduce possible classification errors. Experimental results include four steps, Firstly, the performance of contextual fusion is evaluated in order to check the applicability of using appearance-based classifiers on mix areas for detecting vehicles and pedestrians. Tracking performance of the proposed system is evaluated and then trajectories are depicted using heat-maps. Finally, behavior analysis of participants (i.e., turning movement count, crossing speed) and surrogate safety measurements (i.e., DTI, TTC) are estimated using the proposed system.

Table 6.1 shows different intersections which were chosen with variety of objectives in this dissertation. The objectives are different applications such as crossing count, turning movement count and TTC estimation regrading vehicles (V) and pedestrians (P). The vision-based tracking system was implemented by C++ using OpenCV 2.3, and it was run on Intel i7 quad core with 6 GB RAM.

6.1.1 Detection System Evaluation

The performance of each detection method is evaluated for two Las Vegas intersections: INT 1 and INT 2. Positions of pedestrians and vehicles were manually marked for 1000 frames of each intersection video. GMM, Haar-like features and combined methods are separately used at the

Table 6.1: Intersection names, applications and their monitoring time (V: vehicles, P: pedestrains)

INT	Camera No	Location	Application	Monitoring time/frame	Dissertation section
INT 1	2101	LVBLVD/CONV CTR	Appearance-based classifiers evaluation (P)	166 frames	Section 6.1.1
INT 1	2101	LVBLVD/CONV CTR	System detection evaluation (V, P)	1000 frames	Section 6.1.1
INT 1	2101	LVBLVD/CONV CTR	System detection evaluation (P)	1000 frames	Section 6.1.1
INT 1	2101	LVBLVD/CONV CTR	System tracking evaluation (P)	1000 frames	Section 6.1.2
INT 1	2101	LVBLVD/CONV CTR	False track removal (P)	3000 frames	Section 6.1.3
INT 1	2101	LVBLVD/CONV CTR	Intersection usage (P)	3000 frames	Section 6.2
INT 1	2101	LVBLVD/CONV CTR	Crossing count, speed, waiting time (P)	11:33-11:49 am	Sections 6.8,6.9, 6.10
INT 2	2203	CASINO/COL BELL S	Appearance-based classifiers evaluation (P)	166 frames	Section 6.1.1
INT 2	2203	CASINO/COL BELL S	System detection evaluation (V, P)	1000 frames	Section 6.1.1
INT 2	2203	CASINO/COL BELL S	System detection evaluation (P)	1000 frames	Section 6.1.1
INT 2	2203	CASINO/COL BELL S	System tracking evaluation (V, P)	3000 frames	Section 6.1.2
INT 2	2203	CASINO/COL BELL S	System tracking evaluation (P)	1000 frames	Section 6.1.2
INT 2	2203	CASINO/COL BELL S	False positive removal (P)	1000 frames	Section 6.1.3
INT 2	2203	CASINO/COL BELL S	Intersection usage (P)	3000 frames	Section 6.2
INT 2	2203	CASINO/COL BELL S	Queue length evaluation (V)	800 frames	Section 6.7
INT 2	2203	CASINO/COL BELL S	Queue length estimation (V)	2400 frames	Section 6.7
INT 2	2203	CASINO/COL BELL S	Safety evaluation: DTI, TTI (V)	12:04 p.m- 12:56 p.m	Section 6.11
INT 2	2203	CASINO/COL BELL S	Safety evaluation: TTC (V, P)	12:04 p.m- 12:56 p.m	Section 6.12
INT 2	2203	CASINO/COL BELL S	Crossing count, speed, waiting time (P)	12:04-12:54 pm	Sections 6.8,6.9, 6.10
INT 3	2175	LV BLVD/CIRCUS S	Appearance-based classifiers evaluation (P)	166 frames	Section 6.1.1
INT 3	2175	LV BLVD/CIRCUS S	System detection evaluation (P)	1000 frames	Section 6.1.1
INT 3	2175	LV BLVD/CIRCUS S	System tracking evaluation (P)	1000 frames	Section 6.1.2
INT 3	2175	LV BLVD/CIRCUS S	Crossing count, speed, waiting time (P)	1:13: 1:39 pm	Sections 6.8,6.9, 6.10
INT 4	-	BRANT/HAWTHORN	Turning movement count (V)	4:00-6:00 pm	Section 6.3
INT 5	-	LOMAS SANTAFE/HIGHLAND	Turning movement count, waiting time (V)	4:00-6:00 pm	Section 6.3
INT 5	-	LOMAS SANTAFE/HIGHLAND	Waiting time factors of right turns (V)	4:00-5:00 pm	Section 6.6
INT 5	-	LOMAS SANTAFE/HIGHLAND	Turning Speed (V)	4:00-5:00 pm	Section 6.4
INT 6	2200	CASINO/RIVER NORTH	Waiting time (V)	8:00 am-8:00 pm	Section 6.5
INT 6	2200	CASINO/RIVER NORTH	Turning movement count (V)	5 days, 8:00 am-8:00 pm	Section 6.3.1
INT 6	2200	CASINO/RIVER NORTH	Waiting time of each turning movements (V)	1 day, 8:00 am-8:00 pm	Section 6.5
INT 7	2310	LV BLVD/CATHEDRAL S	Detection evaluation for vertical movements (P)	300 frames	Section 6.1.1
INT 8	2134	FLAMINGO/PARADISE WEST	Safety evaluation: TTC, PET (V, P)	10:00-11:00 am	Section 6.12, 6.13

detection step and each method performance is evaluated by comparing detection results with manually annotated files.

Detection results are generally evaluated using true positive rate (TPR), false positive per frame ($FPPF$) and $Jaccard$ coefficient shown in Eq. (6.1). False positive per frame is calculated based on total number of false positives divided by total number of frames indicating an average of false positive per frame. $Jaccard$ coefficient is a way of accounting both false positive (FP) and false negative (FN) values with one indicator. $Jaccard$ value is always less or equal to TPR and its value is increased when the wrongly detected objects (FP) and missed detected objects (FN) are reduced.

$$TPR = \frac{TP}{TP + FN}, \quad Jaccard = \frac{TP}{TP + FP + FN} \tag{6.1}$$

Table 6.4 shows each method's performance for different traffic signal phases. The total column shows the average for entire frames regardless of traffic signals. Motion by GMM shows better performance than appearance-based techniques for vehicles. The combined method provides higher true positive rate (TPR) than GMM with slightly more false positive per frame ($FPPF$). Although

Table 6.2: Vehicle-pedestrian detection performance during traffic phases (INT 1, INT 2)

Object	Classifier	Green		Red		Total	
		TPR	FPPF	TPR	FPPF	TPR	FPPF
Vehicle	GMM	0.6	1.95	0.33	2.15	0.46	1.75
	Haar	0.34	8.34	0.23	7.39	0.28	7.82
	Combined	0.64	3.06	0.42	2.70	0.52	2.69
Pedestrian	GMM	0.12	0.74	0.13	0.74	0.13	0.65
	Haar	0.20	8.51	0.21	8.72	0.20	8.76
	Combined	0.24	1.13	0.23	1.25	0.24	1.15

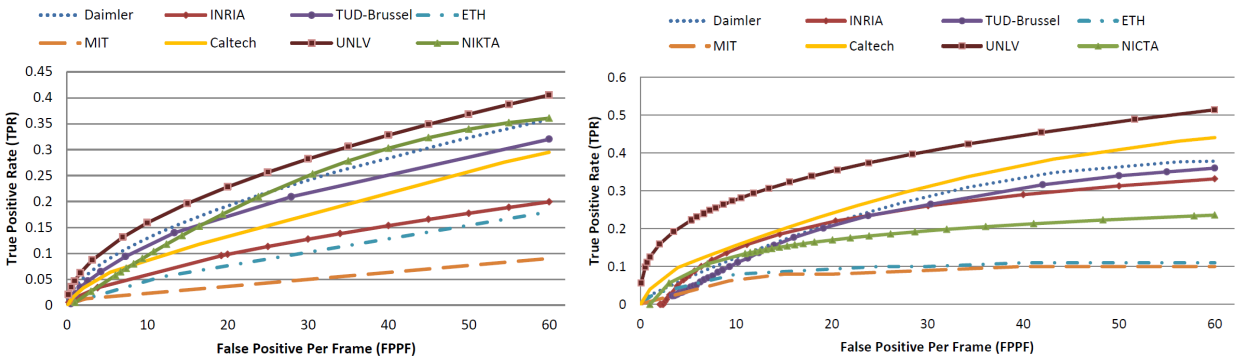
combined method has better performance than other methods, it is not recommended for mix areas since detections of GMM lead to occluded and merged vehicles when vehicles stop in a queue line. This is a reason that the optical flow tracker is enhanced and detection by combined method is only performed on the motion area for vehicles.

Haar outperforms GMM for pedestrian detection since they are very small and stand at sidewalks. The outstanding performance of combined method shows higher TPR and large reduction in $FPPF$ than Haar. Since crosswalk is also considered for appearance-based detection, both motion and appearance are utilized during red and green phase signals leading to less difference in improvement. Finally, combined method introduces less than 1 $FPPF$ that can be eliminated during tracking. As a result, combined method is chosen as a pedestrian detector on mix areas.

Pedestrian Detection Evaluation

In order to analyze pedestrian behaviors at intersection, we concentrated on evaluation of our detection system specifically on pedestrians regardless vehicles. The major motivation was to evaluate the performance of the system when it only addresses the pedestrians by providing the optimized pedestrian classifiers based on accuracy and speed for Haar and LBP features.

In order to choose an appropriate classifier, we annotated and prepared available public datasets. Publicly available datasets were evaluated on three recorded intersection videos for 500 frames in total. Fast detection is a critical parameter [143,169] used in many pedestrian detection applications such as automotive safety and surveillance systems. It provides an inference about the real-time and long time monitoring capability of the system for ADAS and surveillance applications. As a result, we conclude by jointly considering both accuracy and speed as criteria to determine the best



(a) LBP (b) Haar

Figure 6.1: Pedestrian detector performance for video surveillance using different publicly available training datasets.

Table 6.3: Pedestrian Detection Speed for Different Dataset Classifiers (fps: frame per second)

Dataset	LBP (frame)	LBP (roi)	Haar (frame)	Haar (roi)
Daimler	2.64	9.52	1.70	7.46
ETH	12.5	21.73	12.19	21.27
INRIA	6.07	17.24	2.14	8.33
MIT	14.08	22.22	12.82	21.27
NICTA	0.39	3.21	0.17	1.90
TUD Brussel	4.73	14.49	1.90	9.70
Caltech	4.48	13.69	1.45	7.63
UNLV	4.85	13.88	0.98	3.76

pedestrian detector for surveillance application.

The OpenCV implementation of the Haar and LBP cascaded classifiers were used to learn weak classifiers which are combined to form a strong classifier. A pedestrian classifier was learned for each of the available pedestrian datasets and tested on the UNLV Pedestrian dataset. The performance is compared using the receiver operating characteristic (ROC) curves shown in Figure 6.1. Caltech, UNLV and Daimler outperform others as was expected for both Haar and LBP classifiers due to dataset size and difficulty. Although NICTA shows a good performance for its LBP detector, its detection rate is very low (see Table 6.3). The Daimler dataset has strong performance since pedestrian collected samples match pedestrian samples in testing videos based on quality and size. As a result, these three datasets are the best candidates for the system.

Besides accuracy, another important operational parameter is the detection runtime. Sliding window evaluation at different scales is a time consuming process and too much complexity can

Table 6.4: Pedestrian detection performance during traffic phases

INT	Method	Dataset	Green			Red			Total		
			TPR	FPPF	Jaccard	TPR	FPPF	Jaccard	TPR	FPPF	Jaccard
INT 1	GMM	-	0.24	0.78	0.21	0.14	0.37	0.13	0.21	0.66	0.19
	LBP	UNLV	0.23	1.44	0.18	0.22	1.42	0.17	0.23	1.44	0.18
	Haar	Caltech	0.35	17.01	0.08	0.46	12.58	0.12	0.38	15.68	0.09
	GMM+LBP	UNLV	0.22	0.32	0.21	0.20	0.27	0.19	0.21	0.31	0.20
	GMM+Haar [12]	Caltech	0.31	0.78	0.27	0.37	1.18	0.30	0.33	0.90	0.28
INT 2	GMM	-	0.13	0.69	0.11	0.03	0.15	0.03	0.08	0.37	0.07
	LBP	UNLV	0.16	1.52	0.11	0.11	1.72	0.06	0.13	1.64	0.09
	Haar	Caltech	0.40	13.93	0.08	0.40	14.00	0.06	0.40	13.97	0.07
	GMM+LBP	UNLV	0.22	0.94	0.17	0.07	0.45	0.06	0.15	0.65	0.12
	GMM+Haar [12]	Caltech	0.34	1.29	0.25	0.19	0.62	0.16	0.26	0.90	0.20
INT 3	GMM	-	0.49	0.32	0.43	0.11	0.07	0.11	0.45	0.29	0.39
	LBP	UNLV	0.29	1.87	0.15	0.17	2.09	0.09	0.27	1.89	0.14
	Haar	Caltech	0.34	10.26	0.05	0.42	10.76	0.07	0.35	10.31	0.06
	GMM+LBP	UNLV	0.56	0.39	0.47	0.24	0.09	0.23	0.52	0.35	0.44
	GMM+Haar [12]	Caltech	0.57	0.42	0.47	0.48	0.18	0.44	0.56	0.39	0.47

limit real-time implementation. In the proposed method, the raster scan can be performed over the mix areas or the region of interest (roi) to improve the speed performance. Table 6.3 shows the pedestrian detection speed of different classifiers in video surveillance. There is a large speed up when the detector is only performed on mix areas as part of the contextual fusion process. Since UNLV has higher detection speed than Caltech, it is chosen as a final candidate for LBP detector. However, the UNLV Haar detector has lower detection speed and, as a result, the Caltech dataset is chosen for the Haar-based detector.

Finally, the detection system (i.e., contextual fusion (CF)) is compared against motion and appearance-based detection methods. The GMM is implemented as a representative of motion-based pedestrian detection methods, and it is fused with appearance-based methods. The evaluation results use true positive rate (TPR) and Jaccard coefficient.

Table 6.4 shows the performance of the detection methods for 1000 frames of three intersections. Motion-based methods have higher J values during the green phase signal since pedestrians cross the intersection and motion-based and appearance-based are actively dealing with moving pedestrians. However, motion-based and also appearance based detectors might introduces some false positive at some time (e.g., INT 3: Haar and GMM+Haar) which reduces the J value in comparison with

(a)

(b)

Method	TPR	Jaccard
GMM	0.08	0.08
LBP	0.12	0.11
Haar	0.16	0.15
GMM+LBP	0.16	0.15
GMM+Haar	0.21	0.2

(c)

Figure 6.2: Experiments for an intersection (INT 7) with vertical movement of pedestrians: a) Detection results, b) Contextual fusion, c) Quantitative assessment.

the red signal phase Contextual fusion methods (i.e., GMM+LBP, GMM+Haar) show the best performance since they have the highest J values during all light phases (i.e., red, green, total). The contextual fusion methods are evaluated further inside the tracking system to better discern performance.

In order to assess the performance of the contextual fusion for different perspective settings, A new intersection (INT 7) was selected with the pedestrian movement in vertical direction. The complex scenario included larger groups of pedestrians and was annotated for 300 frames. Figure 6.2 shows the vertical experiment results along with a snapshot of the tracking system results.

Different detectors are shown with different colors (i.e, GMM: blue, HAAR: green and LBP: brown) in Figure 6.2a and the contextual mix area is in purple. The final detections using contextual fusion are shown in Figure 6.2b which shows improved pedestrian localization by pooling of motion and appearance inside mix area. Figure 6.2c shows the detection results which again demonstrate higher J values through contextual fusion. The lower J value in this scene is comparable to INT 2 and results from the complexity of pedestrian grouping and very small pedestrians passing the top of contextual mix area which are very difficult to reliably detect.

6.1.2 Tracking System Evaluation

Five criteria are defined to evaluate the performance of the tracking system quantitatively (see Figure 6.3):

1. Number of mostly tracked (MT) trajectories: more than 80% of the trajectory is tracked.

2. Number of mostly lost (ML) trajectories: more than 80% of the trajectory is lost.

3. Number of fragments (FG) of trajectories: the generated trajectory is between 80% and 20%

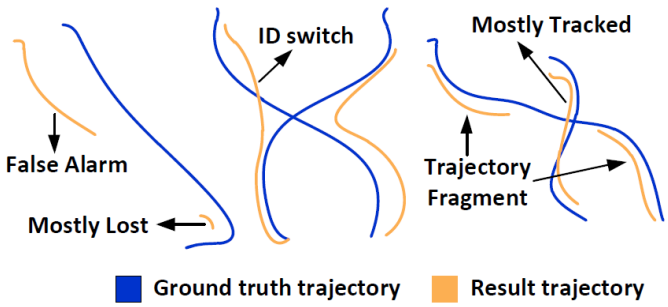


Figure 6.3: Tracking evaluation criteria [8]

Table 6.5: Comparison of the optical flow with the proposed tracking methods (INT 2)

Tracker	Tracking Method	GT	MT	FG	ML	FT	IS	SR
Vehicle	Optical flow [100]	64	26	20	18	0	8	71%
	Proposed	64	42	12	10	0	3	84%
Pedestrian	Optical flow [41]	23	3	8	12	2	0	48%
	Proposed	23	13	8	2	3	0	91%

of the ground truth.

4. Number of false trajectories (FT): trajectories corresponding to no real object

5. Frequency of identity switches (IS): identify exchanges between a pair of result trajectories.

The tracking system is compared against our implementation of the pure optical flow used in [41, 100]. The pure optical flow detects moving objects by clustering the features using motion direction and magnitude. Each moving object is initialized by features which find their match using optical flow. The main difference in comparison with the proposed method is the lack of three introduced steps (i.e., enriching features, filtering features, bounding box estimation) by enhanced optical flow. Moreover, when pure optical flow fails due to high reduction in the number of matched features, an object loses its track since detections from contextual fusion are not utilized by the optical flow tracker.

Table 6.5 shows the comparison of the pure optical flow with proposed tracking method for 3000 frames of INT 2. Since an object might takes a new track couple of times, its maximum trajectory length is classified as an either MT, FG, or ML to simplify the evaluation process. The

success rate, $SR = (MT + FG)/GT$ is defined as a evaluation criteria which indicates the higher rate of complete tracking of an object during its lifetime. The SR value of vehicles are higher than pedestrians for optical flow as it was expected. For vehicles, optical flow show higher IS which occurs for those waiting in a queue behind the red signal. Since bounding box estimation process is not performed for waiting vehicles in the pure optical flow method, grouping features of nearby vehicles become challenging and they might falsely be grouped for another vehicle. As a result, a tracking drift happens which could lead to IS as well.

Since appearance-based classifiers are performed for pedestrians on the mix areas, there might be some falsely detected pedestrians on the background objects such as trees or bollard. As a result, there is a higher value of FT than a pure optical flow for pedestrians. The higher success rate of tracking for pedestrians than vehicles indicates the capability of the system to retain the track of small pedestrians in video surveillance.

Pedestrian Tracking Evaluation

Similar to Section 6.1.1, the proposed system which only tracks pedestrians were also evaluated at tracking step for pedestrians. The six introduced criteria, MT, FG, ML, FT, IS, SR were used to evaluate the system.

Table 6.6 shows the comparison of the pure optical flow with proposed tracking method which fuses appearance and motion at the detection step. Since ML, FG could happen several times for a pedestrian, a track with maximum length is selected to be classified as a ML, FG or MT to simplify the evaluation process. The proposed tracking method using two contextual fusion techniques which showed better performance in Table 6.4 compared against the pure optical flow tracking method.

Table 6.6: Comparison of Optical Flow with Proposed Tracking Fusion Methods

Intersection	Tracking Method	GT	MT	FG	ML	FT	IS	SR
INT 1	Optical flow [170–172]	25	4	12	9	4	1	0.64
	Optical flow (GMM+LBP)	25	8	13	4	20	1	0.84
	Optical flow (GMM+Haar)	25	15	7	3	25	2	0.88
INT 2	Optical flow [170–172]	18	2	6	10	7	0	0.44
	Optical flow (GMM+LBP)	18	5	7	6	12	0	0.66
	Optical flow (GMM+Haar)	18	11	6	1	14	0	0.94
INT 3	Optical flow [170–172]	17	3	6	8	3	0	0.53
	Optical flow (GMM+LBP)	17	11	4	2	7	2	0.88
	Optical flow (GMM+Haar)	17	13	3	1	7	2	0.94

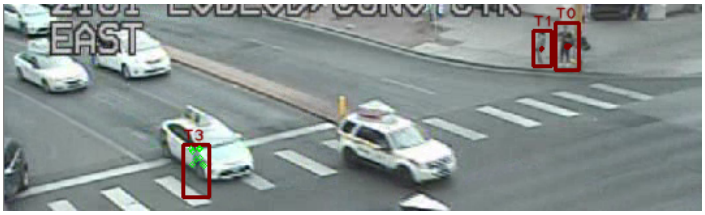
Optical flow(GMM+Haar) has the lowest ML and highest SR value. However, there is a higher number of FT than optical flow(GMM+LBP) which is due to higher number of false positives by Haar classifier. Since most FT could be removed at filtering step mentioned in Section 6.1.3, intersections are monitored using Haar+GMM inside the proposed tracking system.

6.1.3 Filtering False Positives and False Tracks

A common problem with appearance-based classifiers in low resolution settings is mistaking other similar looking objects as pedestrians. Since contextual fusion limits the detection area for appearance, the number of false positives are significantly reduced but there might be some objects on the scene that are detected as a pedestrian like bollards, trees and parts of cars (e.g., wheels) inside mix area. The falsely detected pedestrian leads to a false track if it consistently occurs around the same location. Avoiding this situation is is crucial for pedestrian behavior analysis since false tracks will contribute into final analysis which affect behaviors such as waiting time and crossing speed. Figure 6.4a shows an example of falsely detected pedestrians which led to the false tracks.

There are different methods to deal with false positives such as cascading another classifier as hypothesis verification [173] which limits the speed. Another efficient method learns false positives through the scene and imports scene samples to retrain a classifier through the active learning

(a)



(b)

Figure 6.4: Typical examples of false positives on background objects and moving vehicles. (a) Background objects: The curb (T0) and bollard (T3) (b) Part of a vehicle (T3).

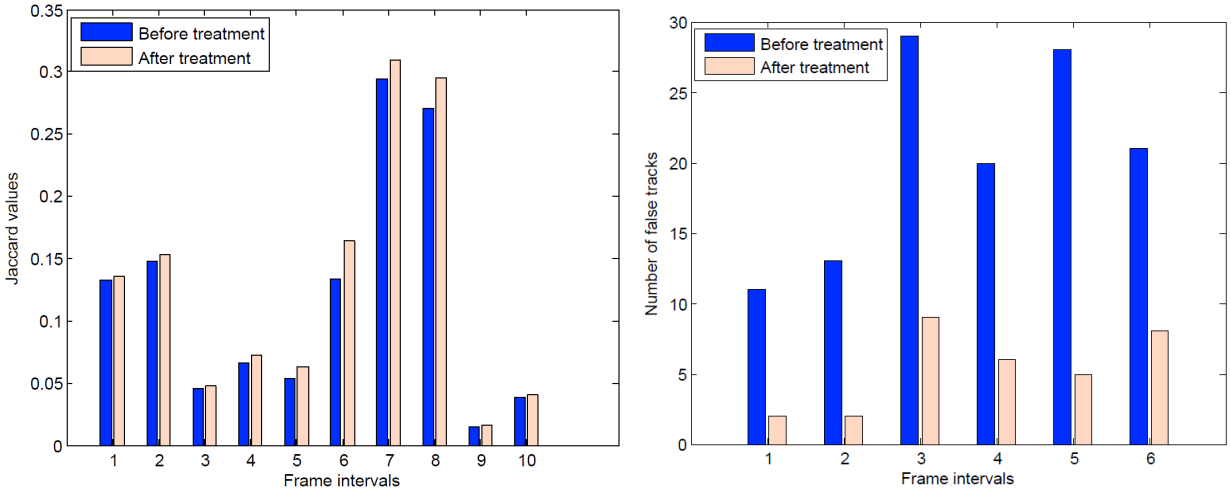
process [127]. Efficient false positive removal requires learning and understanding the scene.

Suppressing false positives are performed at two stages of detection and tracking levels. If there is a background image of the scene with no pedestrians, the falsely detected pedestrians can be efficiently removed using color similarities. The similarity score is evaluated based on template matching of detection on a frame with the background image shown in Eq. (6.2).

$$S(x,y) = \sum_{i=0}^{D_{rows}} \sum_{j=0}^{D_{cols}} D(i,j).B(x+i,y+j) \tag{6.2}$$

where D is a detected pedestrian region on a frame I and B is the background image. Supposedly, (x_{center}, y_{center}) is the center location of D with respect to frame I, and $S(x,y)$ is computed for the whole background image to obtain the highest score location (i.e. $(x_{high-score}, y_{high-score}) = argmax\ S[x][y]$). The detected pedestrian is removed if highest score on the background image exactly matches the same location on the frame shown in Eq. (6.3). For example, if a tree is falsely detected as a pedestrian on a frame, highest scores of the same locations are obtained for frame and background images.

$$\sqrt{(x_{center} - x_{high-score})^2 + (y_{center} - y_{high-score})^2} < \epsilon \tag{6.3}$$



(a) (b)

Figure 6.5: Improvements after removing static false positives and moving false tracks using proposed methods (a) 100 frame intervals (INT2) (b) 500 frame intervals (INT1)

Figure 6.5a shows the improvement on average of Jaccard value (i.e., each 100 frames) for 1000 frames in total. Treatment refers to the process of false positive removal and Jaccard value is utilized (see Eq. (6.1)). The improvement is seen for most bins and the largest difference is shown for the sixth interval (around 3%).

Trajectories are investigated and filtered to remove the false tracks regarding moving vehicles. Typical paths of vehicles can be used to find a false track since pedestrians do not usually walk on vehicles' paths (see Figure 3.12 for vehicles' typical paths). When a track is saved, the average of speed and its similarity with vehicles' typical paths are calculated using LCSS method. Figure 6.5b shows the obtained improvement based on reduction of false tracks per 500 frames. Most false tracks are efficiently removed using the proposed technique and remaining tracks will be filtered by speed criteria before behavior analysis.

6.2 Intersection Usage

The spatial distribution of extracted trajectories is depicted by the heat-maps for 3000 frames of the intersection videos. Heat-maps show the frequency of the intersection locations (e.g., sidewalk, crosswalk) that have been used by participants.

Figure 6.6 shows the pedestrian trajectories of the pure optical flow and the proposed tracking method. As it was expected, more complete trajectories are observed for the proposed method

(a) Optical flow (b) Proposed method

Figure 6.6: Heat-maps of pedestrians' trajectories (INT 1). The proposed method better depicts the crosswalk and sidewalk usage.

which allows us to better analyze different locations used by pedestrians. The sidewalk around the signal clearly shows the success of the system for tracking the waiting pedestrians. The pedestrians usually wait at two distinguished spots around the signal before crossing.

Figure 6.7 shows the trajectories of waiting vehicles using the pure optical flow and proposed tracking method for INT 2. The optical flow has higher rate of success for tracking waiting vehicles since it perfectly shows the waiting locations on the right lane. However, the distribution has higher deviations which indicates tracking drifts of waiting vehicles in a queue. As mentioned earlier, the problem is solved in the proposed method by bounding box estimation. The proposed method depicts frequency of waiting trajectories on the middle and left lanes as well. The heat-map also shows the higher flow of vehicles on the right and middle lane than a left lane.

6.3 Turning Movement Count

Experimental evaluation of turning movement counts was performed in four parts. Counting accuracy was evaluated on two different intersections, INT 4 and INT 5 for two hours and over a long observation period on a third intersection (INT 6) during daylight hours for 5 days. Speed profiling and waiting time was evaluated on one hour of INT 5 data. Finally, a study was conducted to examine the causal relationships for right-turn movements.

Each intersection was set up by defining zones and paths before examining the TM counts. TM

(a) Optical flow (b) Proposed method

Figure 6.7: Heat-maps of waiting vehicles' trajectories (INT 2). The proposed method depicts better tracking of waiting vehicles during red phases.

counts were accumulated for 15 minute intervals as it is typical in transportation studies and it was compared with ground truth to determine the counting accuracy rate. The accuracy rate [12, 146] is defined in (6.4),

$$Ar = 1 - \frac{|M - C|}{M} \tag{6.4}$$

where C refers to automatic count and M specifies the manual count.

Typical paths were learned by recording the first full trajectory for each TM directions. These full trajectories were selected based on a successful count by zone comparison and they were kept for later LCSS path comparison. Typical paths of INT 4-6 are shown in Figure 4.3. The typical paths, starting from North, West, South, and East zones, have red, green, brown, and yellow colors respectively. Notice there is a high degree of overlap between WE and NE paths in INT 4 that make counting difficult by either zone or trajectory techniques.

Quantitative Two Hour Evaluation

The system was evaluated for eight intervals of 13500 frames (15 minute intervals and 2 hours in total) and compared against the ground truth to verify the obtained accuracy.

Table 6.7 shows the TM counting and its accuracy of the zone comparison module at INT 4. The westbound through (WBT) (green) direction has the most traffic. The errors in this direction

Table 6.7: Manual Counting \ Automatic Counting by Zone for INT 4 (4:00-6:00 p.m) [63% Accuracy]

Typical Path	WBT	WBR	NBL	NBT	SBR
4:00-4:15 p.m	103\74	0\0	1\0	10\1	21\20
4:15-4:30 p.m	102\89	1\0	1\0	12\1	21\19
4:30-4:45 p.m	111\79	1\1	0\0	12\0	21\21
4:45-5:00 p.m	129\110	4\3	1\1	14\2	23\19
5:00-5:15 p.m	135\84	1\1	1\6	14\2	18\18
5:15-5:30 p.m	147\74	2\3	2\6	20\7	25\23
5:30-5:45 p.m	142\88	2\1	4\4	23\4	23\20
5:45-6:00 p.m	135\107	4\3	2\0	9\0	19\20
Total	1004\706	15\12	12\17	114\17	171\160
|Difference|	298	3	5	97	11
Accuracy rate	70%	80%	58%	15%	94%

Table 6.8: Manual Counting \ Automatic Counting by Zone+ LCSS for INT 4 (4:00-6:00 p.m) [84% Accuracy]

Typical Path	WBT	WBR	NBL	NBT	SBR
4:00-4:15 p.m	103\95	0\0	1\0	10\9	21\22
4:15-4:30 p.m	102\103	1\0	1\0	12\12	21\20
4:30-4:45 p.m	111\101	1\0	0\0	12\10	21\23
4:45-5:00 p.m	129\120	4\5	1\1	14\13	23\21
5:00-5:15 p.m	135\115	1\1	1\6	14\14	18\19
5:15-5:30 p.m	147\116	2\3	2\6	20\14	25\28
5:30-5:45 p.m	142\119	2\1	4\4	23\18	23\23
5:45-6:00 p.m	135\126	4\3	2\0	9\8	19\21
Total	1004\895	15\14	12\17	114\98	171\177
|Difference|	109	1	5	16	6
Accuracy rate	89%	93%	58%	86%	96%

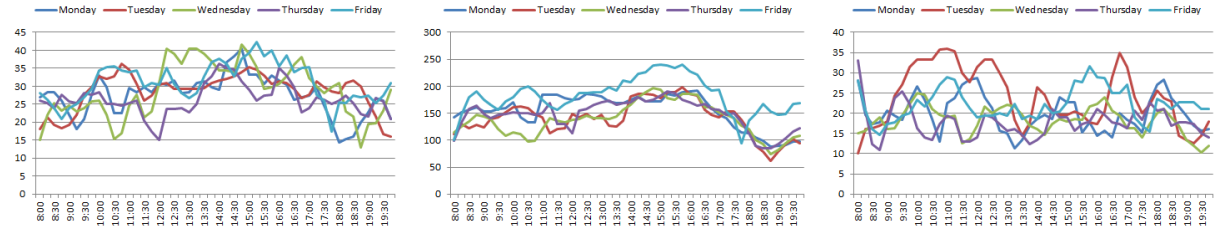
are due to merging or occlusion with northbound vehicles. The northbound directions (red) suffer the lowest accuracy by far because of their distance away from the camera leading to poor tracking due to instability of the blob appearance in the north area.

Table 6.8 shows the results of counting by cooperative zone and trajectory modules. Overall, the accuracy gets improved by LCSS trajectory measure from 63% to 84%. Moreover, there is significant improvement in the NBT and WBT directions indicating the effectiveness of the proposed method to resolve the tracking issues.

Table 6.8 presents the TM count results for INT 5. INT 5 is busier than INT 4 but the camera is placed closer to the intersection resulting in higher resolution and less distortion and

Figure 6.8: Manual & Automatic Counting Results for INT 5 (4:00-6:00 p.m)

Typical Path	WBL	WBT	WBR	NBL	NBT	NBR	EBL	EBT	EBR	SBL	SBT	SBR
Manual Counting	20	78	160	54	518	38	108	145	77	218	547	71
Zone	24	68	164	53	404	36	92	101	73	129	357	62
Zone + LCSS	24	73	172	58	513	43	99	126	77	218	503	82
Accuracy rate (Zone)	80%	87%	98%	98%	78%	95%	85%	70%	95%	59%	65%	87%
Accuracy rate (Zone+LCSS)	80%	94%	93%	93%	99%	87%	92%	87%	100%	100%	92%	85%



(a) (b) (c)

Figure 6.9: Turning Movement Counts for INT 6, a) left turn b) go straight c) right turn

overlapping between different paths. There is a significantly enhanced accuracy, like with INT 4, when using both the zones and LCSS trajectory comparison. In particular, SBL, SBT, and NBT all have accuracy improvements higher than 20%. Since these are the longest traffic paths, they are prone to occlusion but the trajectory comparison module is able to handle the negative effects on TM counting. However, the accuracy in some of the other directions decreased by using LCSS. This happens with noisy tracks that are severely broken leading to mis-matches using LCSS or to multiple smaller trajectories (up-counts). Since the zone comparison module successfully count short trajectories of right turn vehicles (i.e., WBR, NBR, SBR), erroneous counting of other trajectories by LCSS leads to up-counts. These false up-counts using the trajectories for WBR, NBR, and SBR directions marginally affected these directions. The average accuracy rate increased from 83% to 92% by cooperative modules highlighting the effectiveness of the proposed system.

6.3.1 Long-Term Evaluation

In the second part of counting evaluation, the cooperative TM system was run for longer times to depict the operational utility of the proposed system. A third intersection (INT 6) was examined for five working days (8:00 to 20:00) with 15 minute intervals. Moving average filtering was performed over the count data to handle inconsistent measurements due to network connection failures and frame drops. Figure 6.9 shows the TM counts for the 48 intervals in the 12 hour daily observation period separated by day.

Some interesting results are inferred from the counting patterns presented in Figure 6.9. Figure 6.9 (a) shows that the average number of left turns for all five days are in the range of 27 to 31. Wednesday has more fluctuation (higher standard deviation) indicating the burst-like nature of left turns. This is an important metric for designing traffic signal timing to optimize traffic throughput. The busiest time of Wednesday is shown by peaks at certain times (12:15-12:30, 13:00-13:30 and 14:45-15:00).

Vehicle counts for going straight (see Figure 6.9 (b)) shows a fairly consistent pattern from Tuesday through Thursday. Friday and Tuesday have higher counts for going straight and turning right (see Figure 6.9 (c)). These subtle count variations cannot be obtained through traditional TM data collection since counts are only sporadically obtained and only during peak travel hours.

6.4 Turning Speed

Speed behavior was estimated for each turning movement of INT 5 for 52000 frames (around 1 hour). The frequency of speed for each category $\{Stop, Slow, Normal, Speeding\}$ is computed and then probability density function of speed for turning right, left and going straight is estimated (See Figure 6.10). The average of speed (V_t) was 27.86 MPH and α_1, α_2 were set to 0.1 and 0.5 respectively using Eq. 4.3.

Experimental results verify some interesting facts. The going straight path usually has longer travel time and it lets vehicles increase their speed while moving in a straight direction. This is shown by 43% value of 'speeding' bin for vehicles that go to straight shown in Figure 6.10. Left turns and go straights show a similar pattern from 'stopped' to 'normal' bins.

Right turns show different behavior by their high probability for 'stopped' and 'slow' bins. Right turns have short travel times and they can be even taken during the red signal phase. Drivers are cautious during taking right turns since they consider crossing pedestrians and opposing vehicles entering the same zone.

6.5 Waiting Time of Turning Vehicles

The vehicles' waiting time is extracted by examining the speed of each track after performing path reconstruction for five different areas shown in Figure 6.11. Vehicle's disappearance time should also be added to the calculated waiting times which is noted by 'After treatment' in Figures 6.12-6.13. The waiting time becomes closer to ground truth by path reconstruction which incorporates

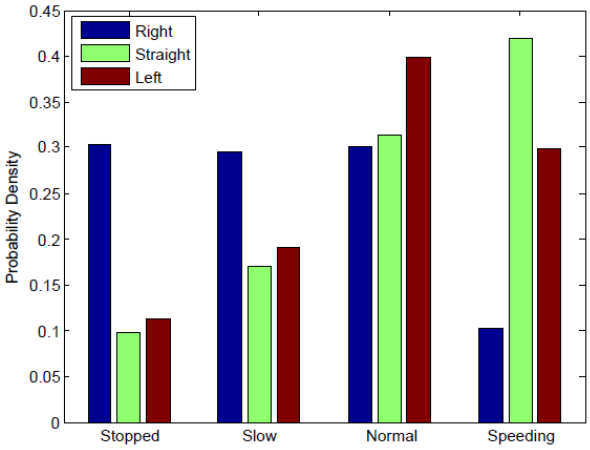


Figure 6.10: Vehicles' speed profile for left turn, right turn and going straight (INT 5, 4:00 pm-5:00 pm)

vehicle disappearance time.

$$E(Ae) = \sum_{i=1}^{T} \frac{\left| H_i - \hat{H}_i \right|}{T} \tag{6.5}$$

where H_i is the manually extracted waiting time, $\hat{H}_i$ is the estimated waiting time and T refers to total track numbers. The average of absolute error reduced from 1.35 to 0.30 second for reappeared tracks. This reduction is the result of adding disappearance time of reappeared tracks (tracks that are learned into the background when stop and re-appear upon re-start). As shown in Figure 6.13, the cumulative distribution of waiting times after applying the proposed method becomes close to the ground truth.

Figure 6.11: Five stop areas used for waiting time analysis (INT 5)

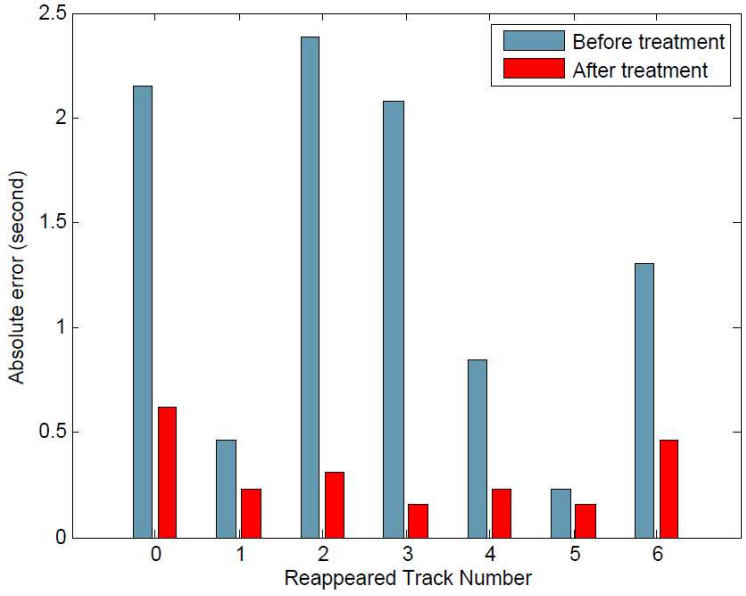


Figure 6.12: Absolute error of reappeared tracks for west to south (WS) direction (INT 5, 4:00 pm-5:00 pm), RS={{1,5,4}, {1,4} . The average of absolute error is used in our evaluations and it is defined in (6.5)

Table 6.9 shows average waiting time of four stop areas in INT 5 by 15 minutes intervals using the proposed method. Vehicles in North and South areas have more waiting times since there is a larger number of vehicles entering these two areas. This can also be verified in NBT and SBT columns of Table 6.8. High flow entering these areas cause more vehicles to stop behind the stop

Table 6.9: Average Waiting Time (Seconds) of Four Approaches

Time	West	North	East	South
4:00-4:15 pm	1.03	1.48	1.15	2.59
4:15-4:30 pm	1.31	1.88	1.31	2.44
4:30-4:45 pm	1.42	1.77	1.20	2.10
4:45-5:00 pm	0.44	1.41	1.12	1.85

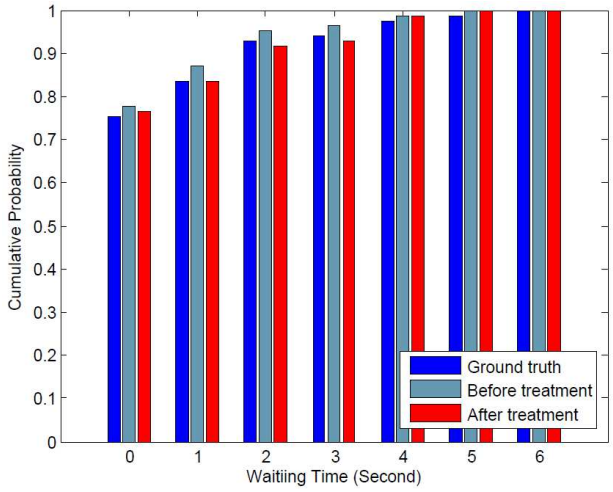


Figure 6.13: Cumulative distribution function of vehicles' waiting time moved in WS direction (INT 5, 4:00 pm- 5:00 pm) , RS={{1,5,4}, {1,4}}

bar creating queue lines and subsequently more waiting time is expected.

Waiting time of vehicles that entered from the east area of INT 6 was examined to estimate waiting time of right turn, left turn and going straight vehicles for one day. INT 6 is a signalized intersection and less waiting time is expected for right turning vehicles than other ones that might face with the red signals.

Figure 6.14 shows cumulative distribution functions of each turning movements and demonstrates the shorter waiting time of right turns than the other TMs since it reaches to 90% of distribution sooner than other movements. Although turning left and going straight follow relatively similar patterns, there are more vehicles that do not stop when going straight since they reached the intersection at a green phase of the signal. However, some vehicles that go straight had waiting time around 20 seconds which increased the cumulative probability.

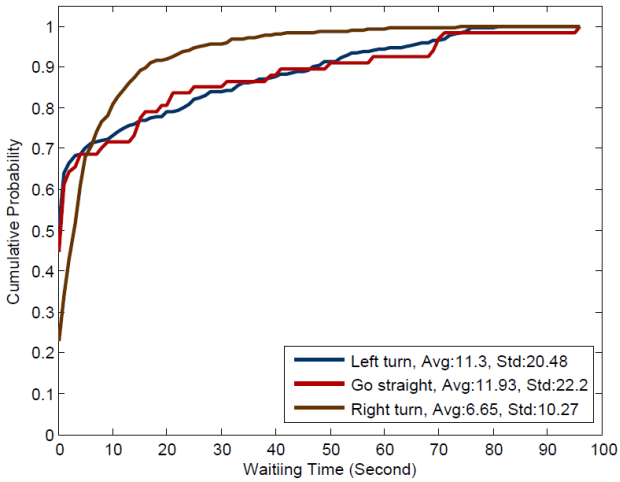


Figure 6.14: Cumulative distribution function of waiting time for turning left, right and going straight (INT 6, Monday: 8:00 am- 8:00 pm)

Figure 6.15: Regression coefficients of each movement counts for waiting time of right turns shown by *

Location	Time	WBL	WBT	WBR	NBL	NBT	NBR	EBL	EBT	EBR	SBL	SBT	SBR
INT2	4:00-5:00 pm	0	0	*	0	1.73	0.32	0.49	0	0.07	−0.39	0.38	−0.38
INT3	8:00-9:00 am	9	0	0	1	−0.02	0.05	0	0	*	−11.85	1.52	NA

6.6 Waiting time factors of right turns

Two right turns taken by different vehicles, which were examined during their tracking, are shown by solid arrows in Figure 6.16. Table 6.15 shows the estimated w coefficients using linear regression method (see section 4.1.4). In each intersection type, there are two competing TMs which contributed most to right turn waiting time, those that compete in the final zone. High coefficient values were computed for NBT, EBL of INT 5 and WBL, SBT of INT 6. This implies the attention of drivers to opposing traffic flows for right turns as an integral part of their gap estimation process.

6.7 Queue Analysis

The proposed system was utilized to estimate queue length, and number of waiting vehicles in the queue. The proposed system was evaluated for a highly cluttered video from one of the Las Vegas junction: INT 2. The experiments included evaluation of the system for queue length estimation,

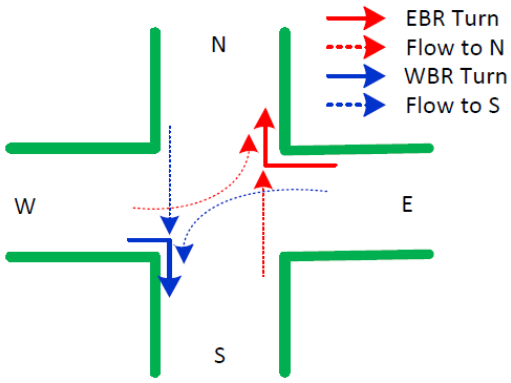


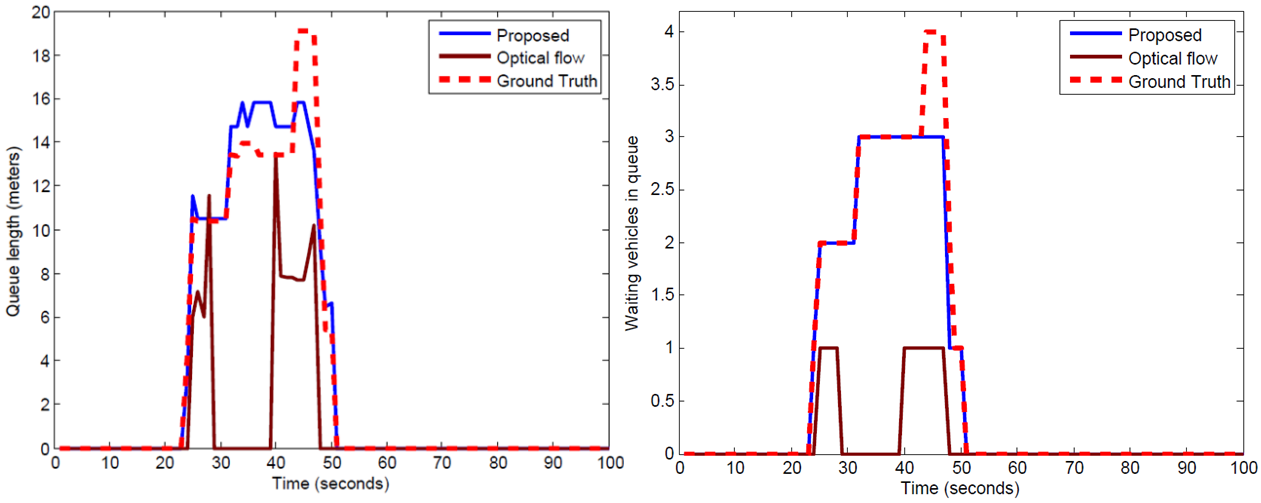
Figure 6.16: Two examined right turns (solid arrows) and opposing flows (dashed arrows) that affect their waiting times: The waiting time increases when the two opposing flows increase.

count of waiting vehicles in queue and waiting time estimation of vehicles in regards to different lanes.

The queue length and number of waiting vehicles are estimated and compared against the ground truth for 800 frames (i.e. 8 frames per second) of the junction video (See Figures 6.17a, 6.17b). The ground truth is created based on manual observation of each frame. Figure 6.17a shows that the estimated queue length follows the ground truth with small offset due to projection error. The average of absolute error (Ae) shown in Eq. (6.5) is used in our evaluations where H_i is the manually extracted queue length, $\hat{H}_i$ is the estimated queue length and N refers to total time (seconds). The average of absolute error was 0.42 meter for $N = 100$ which indicates the effectiveness of the proposed method for estimating queue length. The only offset shown in Figure 6.17b, manifested as big gap in Figure 6.17a (seconds 42-48) due to losing the track of the last vehicle.

Figure 6.18a shows the queue length estimation and Figure 6.18b depicts the corresponding number of waiting vehicles regarding two different lanes. These figures show a similar pattern and traffic signal phases can be inferred from them. Different queue lengths for the same number of waiting vehicles are shown (e.g. 24-50 seconds) due to tracking error and different sizes of waiting vehicles in lane 1.

Figure 6.19a shows the waiting time distribution of the vehicles regarding two different lanes. There is a higher average of waiting time for lane 1 since it is a more congested path. As shown in Figure 6.19a, there are some vehicles with waiting times of more than 40 seconds since they reach to junction at the beginning of the red phase signal. Moreover, more vehicles lead to higher queue



(a) Queue length estimation versus the ground truth (b) Waiting vehicles in the queue

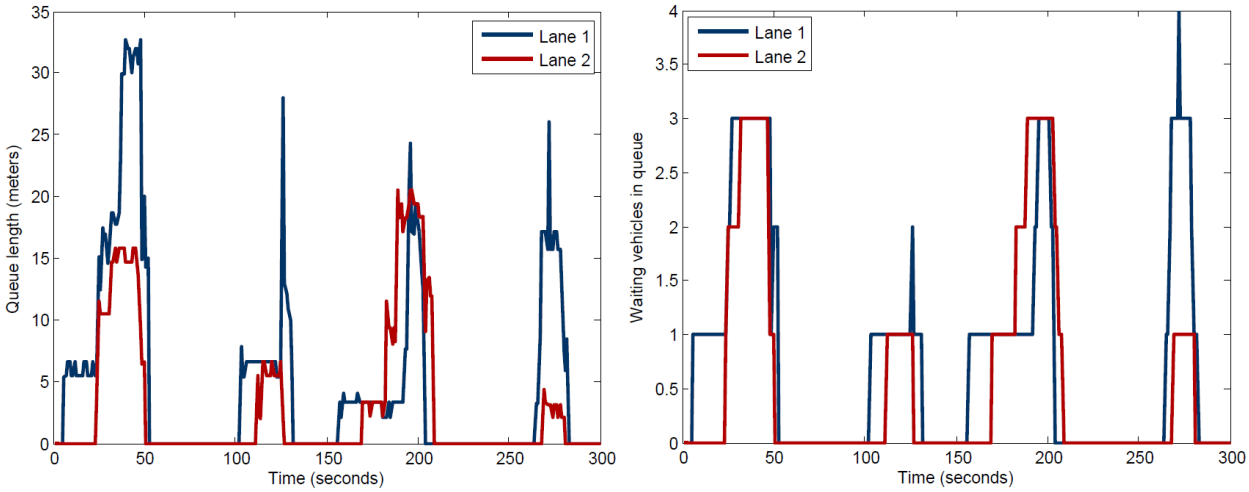
Figure 6.17: Evaluation of queue estimate (INT 2, lane 2)

length which results in more waiting time of the last vehicles.

6.8 Crossing Count Results

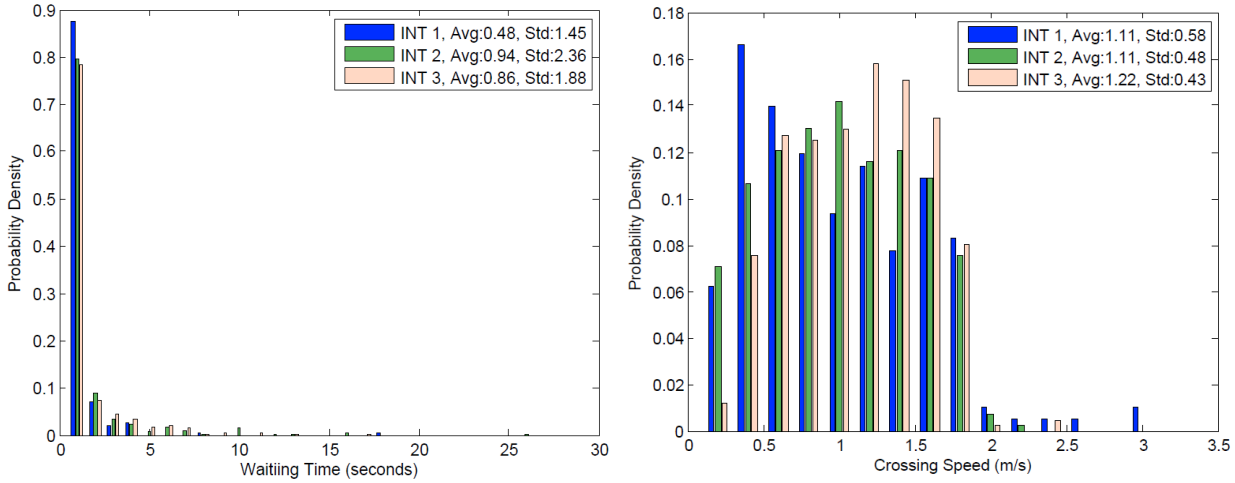
Pedestrians are counted using the LCSS method. In order to match a typical path, a fraction of data points must match. As a result, broken trajectories could be counted if they still have enough matching points with a typical path. Crossing count results are compared with ground truth to determine the counting accuracy rate. The accuracy rate [12] is calculated based on the similar approach shown in Eq. (6.4).

Table 6.10 shows the pedestrian crossing count results of two different directions for different time intervals. INT 1 is the more crowded intersection since it has the high crossing counts during its monitoring time (i.e. 11:43- 11:48). Besides tracking performance that directly affects crossing count, groups of two pedestrians affect the counting performance since they usually walk side by side and the perspective view of the camera makes it difficult to distinguish the individual pedestrians.



(a) Queue length (b) Waiting vehicles in queue

Figure 6.18: Queue length and waiting vehicles estimation (INT 2)



(a) Waiting time (b) Crossing speed

Figure 6.19: Waiting time and crossing speed distribution of pedestrians for three different intersections

6.9 Crossing Speed

Behavior analysis includes waiting time and crossing speed estimation used for design and planning of intersections and signals. The waiting time is obtained by checking the total time during the track life time that a pedestrian has speed value lower than a predefined threshold. Since pedestrians usually have two states of waiting or walking at intersections, the time of waiting is excluded from the entire track to estimate the average speed.

Table 6.10: Pedestrian Crossing Counts (LR: Left to Right, RL: Right to Left, AR: Accuracy Rate)

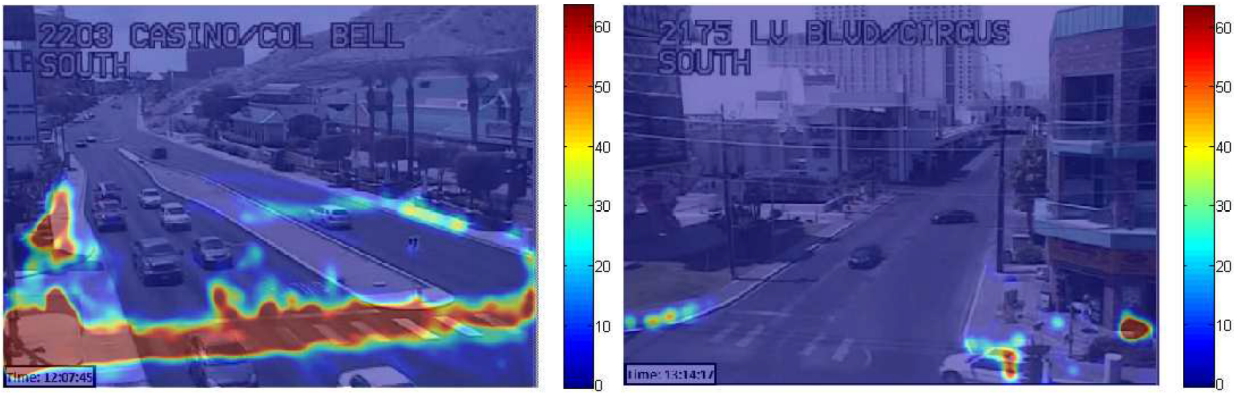
Intersection	Time	LR	AR	RL	AR	Time	LR	AR	RL	AR
INT 1	11:33-11:38	13	100%	23	95%	11:38-11:43	20	70%	10	90%
	11:43-11:48	16	75%	25	72%	11:48-11:49	5	80%	5	80%
INT 2	12:04-12:09	13	93%	4	75%	12:09-12:14	5	80%	0	-
	12:14-12:19	0	-	5	80%	12:19-12:24	2	100%	4	75%
	12:24-12:29	4	75%	5	80%	12:29-12:34	8	87%	0	-
	12:34-12:39	4	75%	4	75%	12:39-12:44	4	75%	0	-
	12:44-12:49	4	75%	0	-	12:49-12:54	4	75%	4	75%
INT 3	13:13-13:18	5	80%	9	88%	13:18-13:23	11	91%	10	70%
	13:23-13:28	14	85%	14	71%	13:28-13:33	13	92%	17	82%
	13:33-13:38	7	71%	10	80%	13:38-13:39	0	-	7	85%

Figure 6.19a shows the estimated waiting distribution of pedestrians during the monitoring time. There are a large number of tracks with waiting time of less than a second for all three intersections. Pedestrians at the second intersection wait more which indicates more time of red traffic signal. The results show the success of tracking system to keep track of small size waiting pedestrians up to 7 seconds since corrosponding values are observed on the figure. Since pedestrian tracks might break during their wait, there is a high portion of waiting time with value of less than a second.

6.10 Watitng Time of Pedestrians

Figure 6.19b shows the higher crossing speed of INT 3 in comparison with others. This value is even higher than the previously reported walking speed of 1.2 m/s at intersections [81]. This shows the high noisy tracks of this video. Occluded pedestrians crossing together cause variation in the size of a bounding box during tracking which adds noise to the trajectory. The small noise variations could have significant impact on crossing speed estimation since noisy measurements would map to large distances in real world coordinates.

Figure 6.20 shows the heat maps of moving and waiting pedestrians. There is a high density region mark out the crosswalk usage patterns for moving pedestrians (Figure 6.20a). Interestingly, jaywalkers path is clearly observed on this map with the aqua color. There is also a high density on two sidewalks in INT 2 which indicates the successful GMM motion-based detection of pedestrians outside of the mix area (i.e., sidewalk). Figure 6.20b shows a high frequency of pedestrians who wait behind curbs to find a chance to cross. Surprisingly, a high density region of waiting pedestrians is



(a) INT 2 (b) INT 3

Figure 6.20: Heat maps of moving and waiting pedestrians

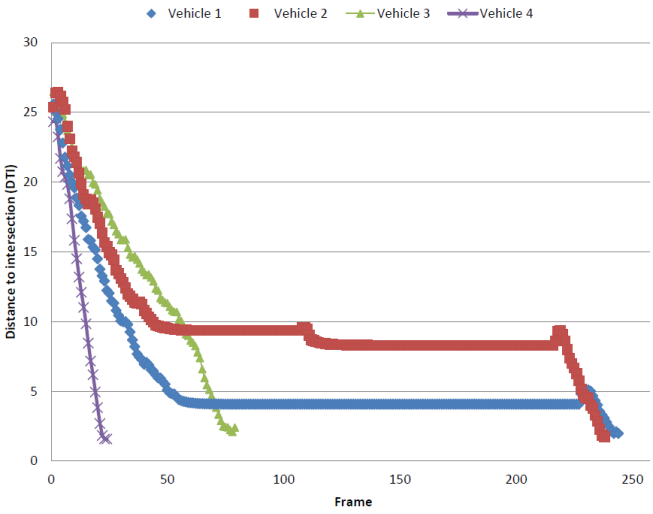


Figure 6.21: DTI of four typical vehicles. Vehicles 1 and 2 face with red signal and wait in a queue.

manifested around the store market which indicated a waiting pedestrian talking with his cellphone for a long time period.

6.11 Distance & Time to Intersection

Surrogate safety measurements such as DTI and TTI are estimated as two typical examples of safety applications that can be addressed by the proposed system. These surrogate safety measurements were estimated for 18318 frames (i.e., 12:04 p.m- 12:56 p.m) of INT2. TTI and DTI are naturally

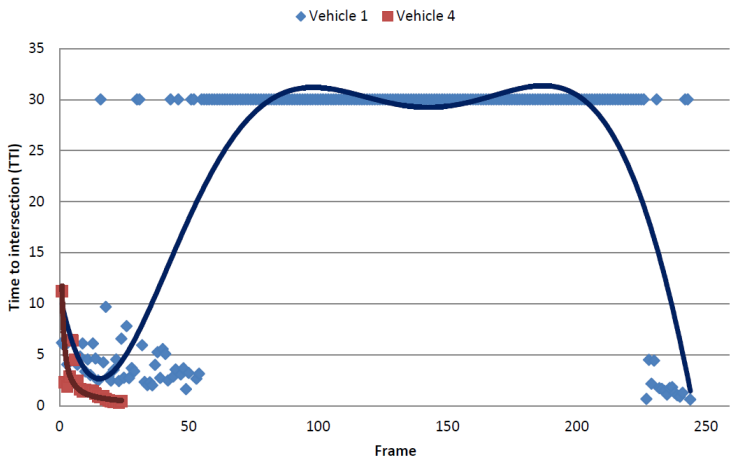


Figure 6.22: TTI of two corresponding vehicles. TTI of vehicle 1 decreases since it faces green signal. TTI of vehicle 1 reaches infinity (i.e. replaced by 30) since it waits behind a red signal phase.

conducted by drivers before taking turns to assess the level of threat posed by opposing traffic. These important safety measurements are used in safety systems including advanced driver assistant systems (ADAS) and intersection decision support systems (IDS). DTI is calculated based on the distance of the vehicle from the stop bar.

Figure 6.21 shows the DTI plots of 4 vehicles. D2I value has a decreasing trend until a vehicle stops which causes DTI remains unchanged (i.e., vehicles 1 and 2). DTI values of vehicles 1 and 2 indicate different positions of waiting vehicles in a queue behind the red signal.Vehicles 3 and 4 never face with the red signal and they continue their moving during their tracking. Vehicle 3 brakes some time since its slope is reduced but vehicle 4 moves with the same speed during its travel time. DTI plots can not be easily utilized with traditional vision-based frameworks since the trajectories of stopped vehicles are lost after a while.

The TTI value is obtained by dividing the DTI on speed at an instant. Figure 6.22 shows the TTI plots of two vehicles. TTI plots have linear decreasing trends similar to D2I for moving vehicles. Since estimated velocity might be noisy, the trend line by polynomial regression (with degree of five) is utilized to better understand the approaching behavior (black curve). High TTI including noisy or infinity values due to zero velocity are replaced by 30 to get an observable trend line. Vehicle 4 has a decreasing trend and it never stops at the intersection since it doesn't face with the red signal during its travel time. Vehicle 1, which stops for a long time before turning, shows a different behavior. First, its TTI value decreases during its approach to the intersection.

(a) (b)

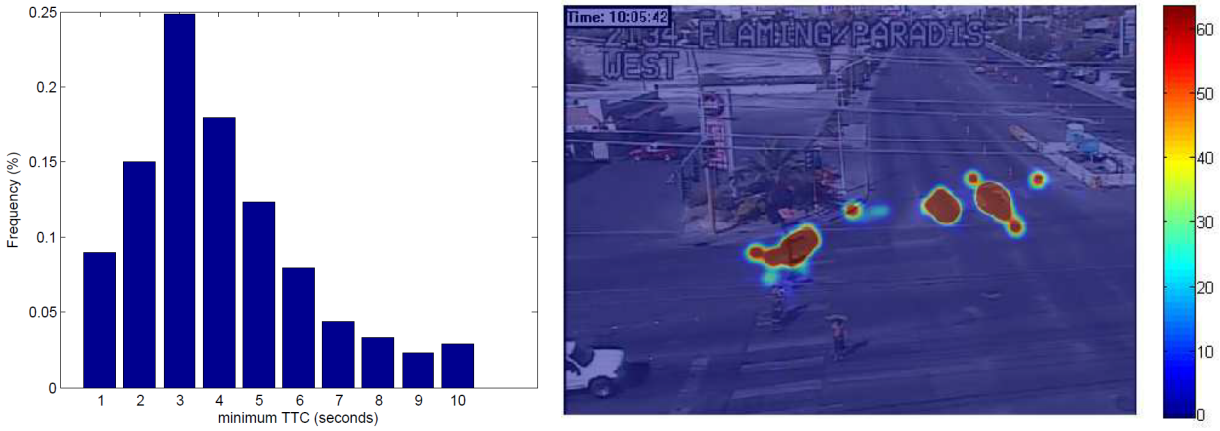
Figure 6.23: TTC evaluation (a) Probability density of TTC values (b) Vehicle-pedestrian conflict heat-map frequency

Since it encounters a red signal, its speed decreases and the TTI value caps at 30 when it stops. After it starts moving again, the typical decreasing trend line is reestablished.

6.12 Time to Collision

TTC was estimated during the monitoring of INT 2 (i.e., 12:04 p.m- 12:56 p.m) and INT 8 (i.e., 10:00 a.m- 11:00 a.m). TTC is calculated based on the predicted arrival time and it is defined as the time for two objects to collide if they continue with their present speed on their paths. Each partially observed trajectory of vehicles is compared against typical paths to find its most probable path and its associated conflict point. The time to conflict point for a vehicle is compared to those pedestrians that are moving toward it and the minimum TTC value is counted if both timings are in a same window.

Figure 6.24a shows the probability density of TTC values for vehicle-pedestrian conflicts (INT 2). The intersection has a peak at 3 (seconds) which classifies intersection as a medium hazardous level since TTC values of less than 2 seconds indicate the high severity of conflicts. Figure 6.25b shows the high frequency of conflict on the crosswalk between pedestrians and vehicles on the right lane. The major reason is the higher flow of vehicles on this lane and higher number of pedestrians crossing from the right to left.



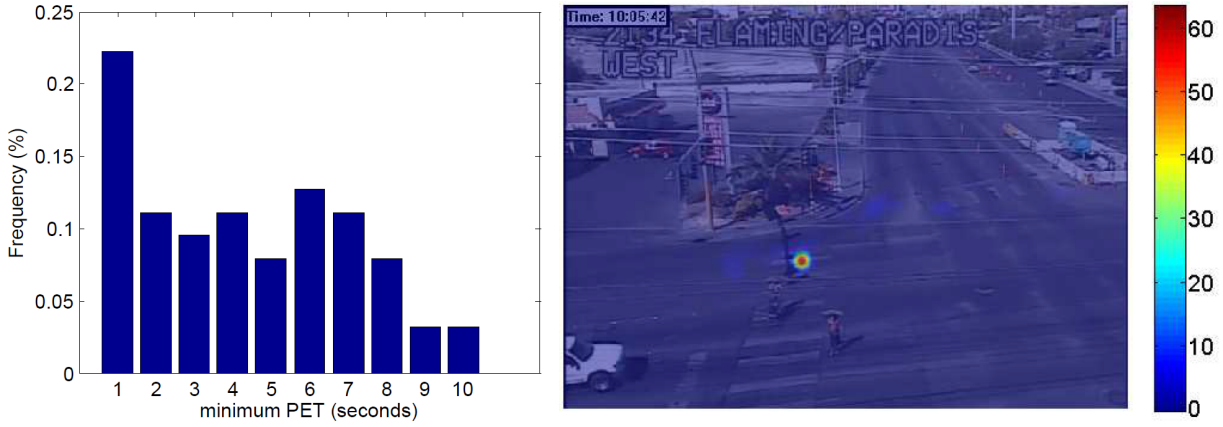
(a) (b)

Figure 6.24: TTC evaluation (a) Probability density of TTC values (b) Vehicle-pedestrian TTC conflict heat-map frequency

TTC results of INT 8 is shown in Figure 6.24. The probability density function of TTC values shows the peak at 3 second similar to INT 2. However, it has less amount of TTC values less than 3 second in comparison with INT 2. The INT 2 looks slightly more dangerous based on their TTC comparison. This shows that two intersection have similar level of hazard level. The heat-map shows high volume of conflicts for pedestrians crossing with approaching vehicles. The areas are depicted for vehicles entering from the north area and then either going straight or turning right. Turning lefts show small amount of conflict.

6.13 Post Encroachment Time

The PET is estimated through analyzing trajectories of counted vehicles and pedestrians. The counted vehicles and pedestrians are selected since they handle the miss classification problem between vehicles and pedestrians. The classification performance is improved during the time since a Bayesian method is utilized based on observed aspect ratio and speed of road users. The PET was estimated for INT 8 during its monitoring time (i.e., 10:00 a.m- 11:00 a.m).



(a) (b)

Figure 6.25: TTC evaluation (a) Probability density of PET values (b) Vehicle-pedestrian PET conflict heat-map frequency

Figure 6.25 shows the probability density of PET values less than 10 seconds for vehicle-pedestrian conflicts (INT 8). The peak on 1 seconds indicates that vehicles usually wait for crossing pedestrian and then quickly pass the conflict point. The heat-map shows the high volume of PET conflict occur between turning right vehicles and crossing pedestrians.

Chapter 7

Conclusion & Future Perspective

In this dissertation, a vision-based tracking system was presented in order to provide behavior analysis (e.g., waiting time) and safety analysis (e.g., DTI) of intersection participants. The tracking system uses contextual fusion at mix and motion areas for robust detection of pedestrians and vehicles, and it undergoes some improvements to become robust against partial occlusion and tracking drifts.

The proposed system estimated safety measurements such as waiting time, DTI, and TTI can be further applied in other fields such as health-care, retail, and intersection decision support systems. For instance, customers can be monitored at stores and their activity analysis including waiting time can help to investigate profits for business intelligence. Future work will address the speed of appearance-based detectors using parallel processing and graphic processing units.

The dissertation focuses on existing infrastructure-based traffic cameras to provide a solution for intersection safety. The overall improvement of safety at intersections is an intricate problem due to complex behaviors and interactions between a range of participants. There are many open issues and further research that should be conducted to fully realize the promise of intersection analysis. The following highlights challenges and possible solutions for complete intersection safety.

7.1 Cooperating Sensing Modalities

Each sensing technology encounters challenges for intersection use. Radars trade-off field of view for range and camera-based detection of pedestrians is difficult in highly occluded scenes and during night. Newer types of non-visible light sensors, such as thermal infrared and LIDAR, show promising results and their cost are decreasing with mass production. Sensor fusion and integration

of sensing modalities with vision can be important for performance enhancement. A video provides a data stream that can be easily understood by humans while complementary sensors can simplify detection algorithms. Further, infrastructure-based sensors can complement vehicle-based sensing to develop a complete intersection description without blind spots. Practical safety systems, such as CICAS or forward warning systems, demonstrate the effectiveness of sensor cooperation.

7.2 Wide FOV and Small Participants

Video-based detection and tracking can be a challenging problem in traffic videos where the FOV is wide to cover an intersection completely resulting in small objects of interest. Detection algorithms must handle pedestrians that are very small in size and scene participants that can become stationary for some time. Motion can be reliable for small objects while appearance can be used to detect stopped objects. Contextual fusion of appearance-based classifiers and motion-based detectors improves intersection detection and tracking robustness [14]. Strategies, such as optical flow, can be utilized to handle partial occlusions and further improve performance.

7.3 Long-Time Monitoring

Long-time monitoring of vehicles' and pedestrians' behavior is a major requirement for intersection analysis (e.g., behavior modeling or scenario understanding). As a result, monitoring system should have real-time capability to generate measurements as they occur over lengthy observation periods. Embedded systems connected to surveillance cameras is a good solution for processing with communication systems in place to share measurements (e.g. with a turning movement count). With new networking technology it may be possible to stream video, with high quality compression, and make use of cloud-based services for computation. With the cloud, less processing equipment is required and software maintenance and improvements are easier.

7.4 Enhancing Behavior Inference with Topic Modeling

Although tracking methods provide microscopic behavior of participants, it can be strengthened through unsupervised learning of motion patterns which provides installation flexibility since the do not require manual calibration. Machine learning approaches are used to extract meaningful patterns, in aggregate, to describe scene and identify behaviors through clustering and topic modeling [174]. Topic models, such as latent Dirichlet allocation (LDA) [175] and hierarchical

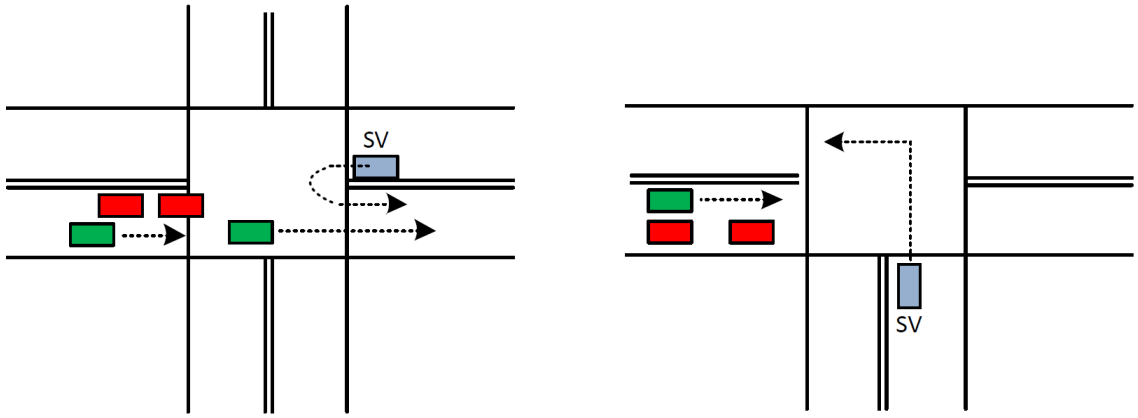
Dirichlet process (HDP) [176], have become quite popular for video surveillance due to their success with natural language processing. The basic idea is to cluster motion vectors based on their co-occurrence in a video clip to describe behavior/movement patterns without any requirement of highly accurate tracking or complex detection methods. The Bayesian modeling approach makes it possible to adjust a model to address different applications (e.g., abnormal behaviors or signal phase discovery).

7.5 Enhanced Abnormal Behavior Detection

Accidents often occur when there are some abnormal behaviors conducted by scene participants. These abnormalities are easier to detect through the traffic monitoring cameras since they provide wide view and scene context. When these abnormalities occur, intersection participants should be warned. Recognizing crosswalks and traffic signal timings are crucial for the detection of intersection misuse, such as jaywalkers or red-light running. Vision-based crosswalk detection can be performed by corner feature detection techniques (e.g., Harris corner detection [177]). The crosswalk corners can be recognized as static features points whose position does not change and form a rectangular shape. Tracking information and cumulative waiting time distribution can be used to infer traffic signal timings [178]. In addition, that queue length estimation helps to infer the traffic phase. Topic modeling is another way to estimate signal phase and detect abnormal behaviors without explicit tracking.

7.6 Human Features and Characteristics

A variety of features are required to characterize humans' behavior and estimate more accurate models. As it was mentioned earlier, some of these data can not be automatically collected from traditional intersection video (including pedestrians' and drivers' age and gender). Personal connected vehicles and smart phones can be good gateways for providing human characteristics to the infrastructure system. ADAS systems (in car or on a smart device), can learn customized driving models and patterns which can populate individual safety databases with different characteristic fields. A personalized system provides opt-in options for characteristics such as age which would not be possible to obtain otherwise and allows for the system to learn specifics based on data (e.g. route choice or average walking speed).



(a) (b)

Figure 7.1: Two obstructed line of sights scenarios during the gap estimation process (a) Driver (SV) to make u-turn on a flashing yellow, but waiting vehicles (red) block the driver view of a moving vehicle (green). (b) Parked cars (red) block the driver view (SV) when attempting to make a left turn at a junction.

7.7 Networked Traffic Monitoring System

The technology of cooperative driving with inter-vehicle communication [179] is a potential solution to suppress traffic jams and prevent collisions. Operational trajectory planning should be solved considering potential conflicts. However, it is very difficult to infer the intention of another driver using only external sensors which increases the conflict search space. Vehicles need to send an intention signal (e.g., turn signal) to the monitoring system. Turning intentions, which are labels, along with other features would be collected by infrastructures for on-line updates and training models. Since traffic cameras have wider FOV, they can, in turn, provide more contextual data for planning and safety ADAS. Cooperation between vehicles and infrastructure can be used to augment the effective sensor coverage and provide preferential views for solving problems such as occlusion, line of sight and situational awareness. Figure 7.1 shows two blocked line of sights which can be addressed through the communication of dashboard mounted camera which faces the direction of travel and infrastructure cameras that provide a complementary view.

7.8 Enhancing Pedestrian Protection Systems

Pedestrians are the most vulnerable intersection participants and require the most protection. However, common pedestrian safety approaches are within vehicle systems that only activate if a pedestrian is detected. Future research can address pedestrian safety on both sides through cooperating sensing (e.g., with cellphones). GPS and navigation data from pedestrian cellphones may be used to improve ADAS detection. The challenges are data management for real-time operation through data windowing only on relevant signals and online behavior model updates. Consumer GPS will be noisy and will require filtering and estimation techniques to have tight localization. As an example, pedestrian localization has been improved by the use of WIFI signals [180].

7.9 Intersection Safety Map

A potential output of the safety quantification process at intersections is a safety map. Surrogate safety measures obtained through tracking or topic modeling methods could be utilized to highlight dangerous intersections on a navigation map. The danger level of a particular intersection could be used by ADAS to adjust control parameters or more simply as a warning for drivers and pedestrians to be more cautious when they are in a hazardous area.

7.10 Joint Warning Infrastructures

While smart devices can provide localization data, they may not be sufficient for alerts due to battery and attention issues. For more reliable notifications, cooperative warning infrastructures can be installed at known hazardous locations (as obtained from the intersection safety map). When an unsafe event is predicted through ADAS or infrastructure-based camera system, warnings (lights or sounds) can be generated; for example, a pedestrian crossing light that turns on automatically.

Bibliography

[1] "Rtc/fast live freeway traffic cams," http://www.rtcsnv.com/fast-cam/.

[2] Y. Liu and R. Tao, "Empirical observations of dynamic dilemma zones at signalized intersections," in *Proceeding Transportation Research Board 86nd Annual Meeting*, Washington, D.C., January 2006.

[3] L. Leden, "Pedestrian risk decrease with pedestrian flow. a case study based on data from signalized intersections in hamilton, ontario," *Accident Analysis & Prevention*, vol. 34, no. 4, pp. 457–464, July 2002.

[4] A. Svensson and C. Hyden, "Estimating the severity of safety related behaviour," *Accident Analysis & Prevention*, vol. 38, no. 10, pp. 379–385, October 2006.

[5] A. Chan, Z.-S. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2008, pp. 1–7.

[6] M. Liebner, M. Baumann, F. Klanner, and C. Stiller, "Driver intent inference at urban intersections using the intelligent driver model," in *Proceeding IEEE Intelligent Vehicles Symposium*, Alcala de Henares, Spain, June 2012, pp. 1162–1167.

[7] M. Akamatsu, Y. Sakaguchi, and M. Okuwa, "Modeling of driving behavior when approaching an intersection based on measured behavioral data on an actual road," *Proceeding in Human Factors and Ergonomics Society Annual Meeting*, vol. 47, no. 16, pp. 1895–1899, January 2003.

[8] B. Wu and R. Nevatia, "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, 2007. [Online]. Available: http://dx.doi.org/10.1007/s11263-006-0027-7

[9] F. D. Salim, S. W. Loke, A. Rakotonirainy, B. Srinivasan, and S. Krishnaswamy, "Collision pattern modeling and real-time collision detection at road intersections," in *Proceeding IEEE Intelligent Transportation Systems Conference*, Seattle, Washington, October 2007, pp. 161–166.

[10] M. S. Shirazi and B. T. Morris, "Vision-based turning movement monitoring:count, speed and waiting time estimation," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 1, pp. 23–34, Spring 2016.

[11] M. Shirazi and B. Morris, "Observing behaviors at intersections: A review of recent studies and developments," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*, June 2015, pp. 1258–1263.

[12] M. S. Shirazi and B. Morris, "Vision-based turning movement counting at intersections by cooperating zone and trajectory comparison modules," in *Proceeding 17th International IEEE Conference on Intelligent Transportation Systems*, Qingdao, China, October 2014, pp. 3100–3105.

[13] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1114 – 1127, June 2008.

[14] M. S. Shirazi and B. Morris, "Contextual combination of appearance and motion for intersection videos with vehicles and pedestrians," in *10th International Symposium, ISVC 2014*, Las Vegas, USA, December 2014, pp. 708–717.

[15] "Inrersection collision warning system," Furner-Fairbank Highway Research Center, Rep. FHWA-RD-99-103, Tech. Rep.

[16] Y. Liu, U. Ozguner, and E. Ekici, "Performance evaluation of intersection warning system using a vehicle traffic and wireless simulator," in *Proceeding IEEE Intelligent Vehicles Symposium*, June 2005, pp. 171–176.

[17] T. Gandhi and M. M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 413–430, September 2007.

[18] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773–1795, July 2013.

[19] Laugier, Paromtchik, Perrollaz, Yong, Yoder, Tay, Mekhnacha, and Negre, "Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety," *IEEE Transaction on Intelligent Transportation Systems Magazine*, vol. 3, no. 4, pp. 889–901, October 2011.

[20] K. Streib, U. Ozguner, J. Martin, Y. Mochizuki, and K. Ishikawa, "A sensor based assessment of imminent collisions at right angle intersections," in *Proceeding IEEE International Conference on Vehicular Electronics and Safety*, Columbus, Ohio, September 2008, pp. 210–215.

[21] K. C. Fuerstenberg, "A new european approach for intersection safety the ec-project intersafe," in *Proceeding IEEE Intelligent Transportation Systems*, Vienna, Austria, September 2005, pp. 432–436.

[22] Q. Tran and J. Firl, "Modeling of traffic situations at urban intersections with probabilistic non-parametric regression," in *Proceeding IEEE Intelligent Vehicles Symposium*, Gold Coast, Australia, June 2013, pp. 334–339.

[23] Q. Tran and J. Firll, "Online maneuver recognition and multimodal trajectory prediction-for intersection assistance using non-parametric regression," in *Proceeding IEEE Intelligent Vehicles Symposium*, Dearborn, Michigan, June 2014, pp. 918–923.

[24] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Behavior classification algorithms at intersections and validation using naturalistic data," in *Proceeding IEEE Intelligent Vehicles Symposium (IV)*, Baden-Baden, Germany, June 2011, pp. 601–606.

[25] C.-Y. Chan, "Characterization of driving behaviors based on field observation of intersection left-turn across-path scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 3, pp. 322–331, September 2006.

[26] C.-Y. Chan and D. Marco, "Traffic monitoring at signal-controlled intersections and data mining for safety applications," in *Proceeding The 7th International IEEE Conference on Intelligent Transportation Systems*, Washington, D.C., October 2004, pp. 355–360.

[27] C.-Y. Chan and B. Bougler, "Evaluation of cooperative roadside and vehicle-based data collection for assessing intersection conflicts," in *Proceeding IEEE Intelligent Vehicles Symposium*, June 2005, pp. 165–170.

[28] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[29] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[30] A. Barth and U. Franke, "Tracking oncoming and turning vehicles at intersections," in *Proceeding 13th International IEEE Conference on Intelligent Transportation Systems*, Madeira Island, Portugal, September 2010, pp. 861–868.

[31] S. Messelodi, C. M. Modena, and M. Zanin, "A computer vision system for the detection and classification of vehicles at urban road intersections," *Pattern Analysis and Applications*, vol. 8, no. 1-2, pp. 17–31, September 2005.

[32] N. Saunier and T. Sayed, "A feature-based tracking algorithm for vehicles in intersections," in *Proceeding 3rd Canadian Conference on Computer and Robot Vision*, Quebec, Canada, June 2006, p. 59.

[33] H. Veeraraghavan, O. Masoud, and N. P. Papanikolopoulos, "Vision-based monitoring of intersections," in *Proceeding 5th International IEEE Conference on Intelligent Transportation Systems*, Singapore, September 2002, pp. 7–12.

[34] H. Veeraraghavan and N. Papanikolopoulos, "Combining multiple tracking modalities for vehicle tracking at traffic intersections," in *Proceeding IEEE International Conference on Robotics and Automation*, New Orleans, LA, April 2004, pp. 2303–2308.

[35] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Surveys*, vol. 38, no. 4, pp. 1–45, December 2006.

[36] P. Prez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *7th European Conference on Computer Vision*, Copenhagen, Denmark, May 2002, pp. 661–675.

[37] M. Yokoyama and T. Poggio, "A contour-based moving object detection and tracking," in *Proceeding of 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, October 2005, pp. 271–276.

[38] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2.

[39] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *International Joint Conference on Artificial Intelligence*.

[40] B. Horn and B. Schunk, "Determining optical flow," *Artificial Intelligence*.

[41] M. H. Zaki, S. Tarek, A. Tageldin, and M. Hussein, "Application of computer vision to diagnosis of pedestrian safety issues," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2393, pp. 75–84, December 2013.

[42] T. Sayed, K. Ismail, M. H. Zaki, and J. Autey, "Feasibility of computer vision-based safety evaluations: Case study of a signalized right-turn safety treatment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2280, pp. 18–27, December 2012.

[43] N. Saunier and T. Sayed, "Clustering vehicle trajectories with hidden markov models application to automated traffic safety analysis," in *Proceeding International Joint Conference on Neural Network*, Vancouver, Cananda, July 2006, pp. 4132–4138.

[44] K. Ismail, T. Sayed, N. Saunier, and C. Lim, "Automated analysis of pedestrian-vehicle conflicts using video data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2140, pp. 44–54, December 2009.

[45] K. Ismail, T. Sayed, and N. Saunier, "Automated analysis of pedestrian-vehicle: Conflicts context for before-and-after studies," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2198, pp. 52–64, December 2010.

[46] M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, August 1998.

[47] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic monitoring and accident detection at intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 2, pp. 108–118, June 2000.

[48] M. Park, Y. Liu, and R. Collins, "Efficient mean shift belief propagation for vision tracking," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[49] A. Barth, D. Pfeiffer, and U. Franke, "Vehicle tracking at urban intersections using dense stereo," in *Proceeding 3rd Workshop on Behaviour Monitoring and Interpretation*, Ghent, Belgium.

[50] F. Viti, S. P. Hoogendoorn, H. J. van Zuylen, I. R. Wilmink, and B. van Arem, "Speed and acceleration distributions at a traffic signal analyzed from microscopic real and simulated data," in *Proceedings 11th International IEEE Conference on Intelligent Transportation Systems*, Beijing, China, October 2008, pp. 651–656.

[51] P. Kumar, S. Ranganath, H. Weimin, and K. Sangupta, "Framework for real-time behavior interpretation from traffic video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 43–53, March 2005.

[52] E. Kafer, C. Hermes, C. Wohler, H. Ritter, and F. Kummert, "Recognition of situation classes at road intersections," in *Proceeding International IEEE Conference on Robotics and Automation*, Anchorage, Alaska, May 2010, pp. 3960–3965.

[53] T. Hulnhagen, Dengler, I. A. Tamke, T. Dang, and G. Breuel, "Maneuver recognition using probabilistic finite-state machines and fuzzy logic," in *Proceeding of IEEE Intelligent Vehicles Symposium*, San Diego, USA, June 2010, pp. 65–70.

[54] R. Graf, H. Deusch, F. Seeliger, M. Fritzsche, and K. Dietmayer, "A learning concept for behaviour prediction at intersections," in *Proceeding IEEE Intelligent Vehicles Symposium*, Dearborn, Michigan, June 2014, pp. 939–945.

[55] G. S. Aoude and J. P. How, "Using support vector machines and bayesian filtering for classifying agent intentions at road intersections," Massachusetts Institute of Technology, Tech. Rep., September 2009.

[56] S. Lefevre, C. Laugier, and J. Ibanez-Guzman, "Exploiting map information for driver intention estimation at road intersections," in *Proceeding IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011, pp. 583–588.

[57] T. Streubel and K. H. Hoffman, "Prediction of driver intended path at intersections," in *Proceeding in IEEE Intelligent Vehicles Symposium*, Dearborn, Michigan, June 2014, pp. 134–139.

[58] H. Berndt and K. Dietmayer, "Driver intention inference with vehicle onboard sensors," in *Proceeding International IEEE Vehicular Electronics and Safety*, Pune, November 2009, pp. 102–107.

[59] A. Kurt, J. L. Yester, Y. Mochizuki, and U. Ozguner, "Hybrid-state driver/vehicle modelling, estimation and prediction," in *Proceeding 13th International IEEE Intelligent Transportation Systems*, Madeira Island, Portugal, September 2010, pp. 806–811.

[60] T. J. Gates, D. A. Noyce, L. Laracuente, and E. V. Nordheim, "Analysis of driver behavior in dilemma zones at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2030, pp. 29–39, December 2007.

[61] T. Sato and M. Akamatsu, "Influence of traffic conditions on driver behavior before making a right turn at an intersection: Analysis of driver behavior based on measured data on an actual road," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 10, no. 5, pp. 397–413, September 2007.

[62] "Fhwa intersection safety briefing sheets," Washington DC: U.S. Department of Transportation, Federal Highway Administration, Tech. Rep., March 2010.

[63] P. Goh and Y. Wong, "Driver perception response time during the signal change interval," *Applied Health Economics and Health Policy*, vol. 3, no. 1.

[64] H. Rakha, I. El-Shawarby, A. AMer, J. Setti, V. Inman, and G. Davis, "Characterizing driver behavior on signalized intersection approaches at the onset of a yellow-phase trigger," *IEEE Transactions on Intelligent Transportation System*, vol. 8, no. 4, pp. 630–640, December 2007.

[65] J. R. Setti, H. Rakha, and I. El-Shawarby, "Analysis of brake perception-reaction times on high-speed signalized intersection approaches," in *Proceeding International IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, September 2006, pp. 689–694.

[66] P. Papaioannou, "Driver behaviour, dilemma zone and safety effects at urban signalised intersections in greece," *Accident Analysis & Prevention*, vol. 39, pp. 147–158, June 2007.

[67] Y. J. Moon and F. Coleman, "Dynamic dilemma zone based on driver behavior and car-following model at highwayrail intersections," *Transportation Research Part F*, vol. 37, pp. 323–344, February 2003.

[68] A. Tarko, W. Li, and L. Laracuente, "A probabilistic approach to control dilemma occurence at signalized intersections," in *Proceeding Transportation Research Board 85th Annual Meeting*, Washington, D.C., January 2006, pp. 1–17.

[69] J. Caird, S. Chisholm, C. Edwards, and J. I. Creaser, "The effect of yellow light onset time on older and younger drivers perception response time (prt) and intersection behavior," *Transportation Research Part F*, vol. 10, pp. 383–396, March 2007.

[70] I. A. Kaysi and A. S. Abbany, "Modeling aggressive driver behavior at unsignalized intersections," *Accident Analysis & Prevention*, vol. 39, pp. 671–678, October 2007.

[71] C. Keller and D. Gavrila, "Will the pedestrian cross? a study on pedestrian path prediction," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 2, pp. 494–506, April 2014.

[72] C. F. Wakim, S. Capperon, and J. Oksman, "A markovian model of pedestrian behavior," in *Proceeding International IEEE Conference on Systems, Man and Cybernetics*, October 2004, pp. 4028–4033.

[73] Y. Abramson and B. Steux, "Hardware-friendly pedestrian detection and impact prediction," in *Intelligent Vehicles Symposium, 2004 IEEE*, June 2004, pp. 590–595.

[74] M. M. Hamed, "Analysis of pedestrians behavior at pedestrian crossings," *Safety Science*, vol. 38, no. 1, pp. 63–82, June 2001.

[75] V. Sisiopiku and D. Akin, "Pedestrian behaviors at and perceptions towards various pedestrian facilities: an examination based on observation and survey data," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 6, no. 4, pp. 249–274, December 2003.

[76] W. H. Lam and C. yu Cheung, "Pedestrian speedflow relationships for walking facilities in hong kong," *Journal of Transportation Engineering*, vol. 126, no. 4, pp. 343–349, July 2000.

[77] O. Keegan and M. O. Mahony, "Modifying pedestrian behaviour," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 889–901, December 2003.

[78] I. M. Bernhoft and G. Carstensen, "Preferences and behaviour of pedestrians and cyclists by age and gender," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 11, no. 2, pp. 4–19, March 2008.

[79] M. S. Tarawneh, "Evaluation of pedestrian speed in jordan with investigation of some contributing factors," *Journal of Safety Research*, vol. 32, no. 2.

[80] Y. Malinovskiy, Y.-J. Wu, and Y. Wang, "Video-based monitoring of pedestrian movements at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2073, pp. 11–17, December 2008.

[81] J. Montufar, J. Arango, M. Porter, and S. Nakagawa, "Pedestrians' normal walking speed and speed when crossing a street," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2002, pp. 90–97, December 2007.

[82] R. A. Ferlis, "Infrastructure collision-avoidance concept for straight-crossing-path crashes at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1800, pp. 85–91, January 2007.

[83] C.-Y. Chan, "Defining safety performance measures of driver-assistance systems for intersection left-turn conflicts," in *Proceeding IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, June 2006, pp. 25–30.

[84] J. Alexander, P. Barham, and I. Black, "Factors influencing the probability of an incident at a junction: results from an interactive driving simulator," *Accident Analysis & Prevention*, vol. 34, no. 6, pp. 779–792, November 2002.

[85] X. Yan, E. Radwan, and D. Guo, "Effects of major road vehicle speed and driver age and gender on left-turn gap acceptance," *Accident Analysis & Prevention*, vol. 39, no. 4, pp. 843–852, July 2007.

[86] D. R. Ragland, S. Arroyo, S. E. Shaldover, J. A. Misener, and C.-Y. Chan, "Gap acceptance for vehicles turning left across on-coming traffic: Implications for intersection decision support design," in *Proceeding Transportation Research Board 85th Annual Meeting*, Washington, D.C., January 2006, pp. 1–25.

[87] D. Sun, S. V. Ukkusuri, R. F. Benekohal, and S. T. Waller, "Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks," in *Proceeding Transportation Research Board 82nd Annual Meeting*, Washington, D.C., January 2003, pp. 1–34.

[88] I. Banerjee, S. E. Shladover, J. A. Misener, C.-Y. Chan, and D. R. Ragland, "Impact of pedestrian presence on movement of left-turning vehicles: Method, preliminary results &

possible use in intersection decision support," UC Berkeley Safe Transportation Research & Education Center, Tech. Rep., November 2004.

[89] C.-Y. Chan, D. R. Ragland, S. E. Shladover, J. A. Misener, and D. Macro, "Observations of driver time gap acceptance at intersections in left-turn across-path opposite-direction scenarios," no. 1910, 2005, pp. 10–19.

[90] S. Leung and G. Starmer, "Gap acceptance and risk-taking by young and mature drivers, both sober and alcohol-intoxicated, in a simulated driving task," *Accident Analysis & Prevention*, vol. 37.

[91] G. S. Aoude, B. D. Luders, K. K. H. Lee, D. S. Levine, and J. P. How, "Threat assessment design for driver assistance system at intersections," in *Proceeding 13th International IEEE Conference on Intelligent Transportation Systems*, Madeira Island, Portugal, September 2010, pp. 1855–1862.

[92] G. S. Aoude, B. D. Luders, J. P. How, and T. E. Pilutti, "Sampling-based threat assessment algorithms for intersection collisions involving errant drivers," in *Proceeding IFAC Symposium on Intelligent Autonomous Vehicles*, September 2010, pp. 581–586.

[93] C.-Y. Chan, D. Marco, and J. Misener, "Threat assessment of traffic moving toward a controlled intersection," in *Proceeding IEEE Intelligent Vehicles Symposium*, Parma, Italy, June 2004, pp. 931–936.

[94] S. Lefevre, C. Laugier, and J. Ibanez-Guzma, "Evaluating risk at road intersections by detecting conflicting intentions," in *Proceeding IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura, Algarve, Portugal, October 2012, pp. 4841–4846.

[95] ——, "Risk assessment at road intersections: Comparing intention and expectation," in *Proceeding IEEE Intelligent Vehicles Symposium*, Alcala de Henares, Spain, June 2012, pp. 165–171.

[96] M. J. King, D. Soole, and A. Ghafourian, "Illegal pedestrian crossing at signalised intersections: Incidence and relative risk," *Accident Analysis & Prevention*, vol. 41, no. 3, pp. 485–490, May 2009.

[97] G. Tiwari, S. Bangdiwala, A. Saraswat, and S. Gaurav, "Survival analysis: Pedestrian risk exposure at signalized intersections," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 10, no. 2, pp. 77–89, March 2007.

[98] H.-C. Chin and S.-T. Quek, "Measurement of traffic conflicts," *Journal of Computer Science*, vol. 26, no. 3.

[99] F. Amundsen and C. Hyden, "Proceedings of first workshop on traffic conflicts," Oslo, Institute of Transport Economics.

[100] T. Sayed, M. H. Zaki, and J. Autey, "A novel approach for diagnosing cycling safety issues using automated computer vision techniques," *ransportation Research Board Annual Meeting Compendium of Papers, 2013.*

[101] H. Berndt, S. Wender, and K. Dietmayer, "Driver braking behavior during intersection approaches and implications for warning strategies for driver assistant systems," in *Proceeding of IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, June 2007, pp. 245–251.

[102] W. K. Alhajyaseen, M. Asano, and H. Nakamura, "Estimation of left-turning vehicle maneuvers for the assessment of pedestrian safety at intersections," *IATSS Research*, vol. 36, no. 1, pp. 66–74, July 2012.

[103] K. Vogel, "A comparison of headway and time to collision as safety indicators," *Accident Analysis & Prevention*, vol. 35, no. 3, pp. 427–433, May 2003.

[104] X. Jin, Y. Zhang, F. Wang, L. Li, D. Yao, Y. Su, and Z. Wei, "Departure headways at signalized intersections: A log-normal distribution model approach," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 3, pp. 318 – 327, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X09000102

[105] J. Tan, L. Li, Z. Li, and Y. Zhang, "Distribution models for start-up lost time and effective departure flow rate," *Transportation Research Part A: Policy and Practice*, vol. 51, no. 0, pp. 1 – 11, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0965856413000931

[106] S. Yin, Y. Su, C. Wang, O. Yao, L. Li, and Y. Zhang, "Comparison of vehicle departure headways in beijing and atlanta," *Tsinghua Science and Technology*, vol. 16, no. 3, pp. 332–336, June 2011.

[107] Y. Qi and P. Yuan, "Pedestrian safety at intersections under control of permissive left-turn signal," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2299, pp. 91–99, December 2012.

[108] O. Akoz and M. E. Karsligil, "Traffic event classification at intersections based on the severity of abnormality," *Machine Vision and Applications*, vol. 25, no. 3, pp. 613–632, April 2014.

[109] W. Hu, X. Xiao, D. Xie, T. Tan, and S. Maybank, "Traffic accident prediction using 3-d model-based vehicle tracking," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 677–694, May 2004.

[110] Y.-K. Ki and D.-Y. Lee, "A traffic accident recording and reporting model at intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 188–194, June 2007.

[111] S. Atev, H. Arumugam, O. Masoud, R. Janardan, and N. P. Papanikolopoulos, "A vision-based approach to collision prediction at traffic intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 4, pp. 416–423, December 2005.

[112] S. Atev, O. Masoud, R. Janardan, and N. Papanikolopoulos, "A collision prediction system for traffic intersections," in *Proceeding IEEE/RSJ International Conference on Intelligent Robots and Systems*, August 2005, pp. 169–174.

[113] W. Hu, X. Xiao, D. Xie, and T. Tan, "Traffic accident prediction using vehicle tracking and trajectory analysis," in *Proceeding IEEE Intelligent Transportation Systems*, October 2003, pp. 220–225.

[114] C. A. Harlow and Y. Wang, "Automated accident detection system," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1746.

[115] F. D. Salim, S. W. Loke, A. Rakotonirainy, and S. Krishnaswamy, "U & i aware: A framework using data mining and collision detection to increase awareness for intersection users," in *Proceeding 21st International Conference on Advanced Information Networking and Applications Workshops*, Nigara Falls, Ontario, May 2007, pp. 530–535.

[116] F. D. Salim, S. W. Loke, A. Rakotonirainy, B. Srinivasan, and S. Krishnaswamy, "Simulated intersection environment and learning of collision and traffic data in the u&i aware framework," in *Proceeding 4th International Conference on Ubiquitous Intelligence and Computing*, Hong Kong, china, July 2007, pp. 153–162.

[117] C. Lee and M. Abdel-Aty, "Comprehensive analysis of vehicle pedestrian crashes at intersections in florida," *Accident Analysis and Prevention*, vol. 37, pp. 775–786, March 2005.

[118] A. F. W. H. B. W. David F. Preusser, JoAnn K. Wells, "Pedestrian crashes in washington, dc and baltimore," *Accident Analysis and Prevention*, vol. 34, pp. 703–710, March 2001.

[119] A. S. Al-Ghamdi, "Pedestrian vehicle crashes and analytical techniques for stratified contingency tables," *Accident Analysis and Prevention*, vol. 34, pp. 205–214, January 2001.

[120] J. Cinnamon, N. Schuurman, and S. M. Hameed, "Pedestrian injury and human behaviour: Observing road-rule violations at high-incident intersections," *PLoS ONE*, vol. 6, June 2011.

[121] C. V. Zegeer, J. R. Stewart, H. Huang, and P. Lagerwey, "Safety effects of marked versus unmarked crosswalks at uncontrolled locations: Analysis of pedestrian crashes in 30 cities," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1773, pp. 56–58, February 2007.

[122] E. A. LaScala, D. Gerber, and P. J. Gruenewald, "Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis," *Accident Analysis & Prevention*, vol. 32, no. 5, pp. 651 – 658, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0001457599001001

[123] H. Veeraraghavan, O. Masoud, and N. Papanikolopoulos, "Computer vision algorithms for intersection monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 2, pp. 78–89, June 2003.

[124] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recogn. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2005.11.005

[125] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, June 2008. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2007.09.014

[126] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Transactions on Pattern Recognition*, vol. 29, pp. 51–59, 1996.

[127] S. Sivaraman and M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 267–276, June 2010.

[128] G. B. Z. Sun and R. Miller, "Monocular precrash vehicle detection: Features and classifiers," *IEEE Transactions on Image Processing*, vol. 15, pp. 2019–2034, 2006.

[129] P. Geismann and G. Schneider, "A two-staged approach to vision-based pedestrian recognition using haar and hog features," *Proceeding of IEEE Conference on Intelligent Vehicle Symposium*, pp. 554–559, 2008.

[130] M. T. A. B. A. Mogelmose, A. Prioletti and T. B. Moeslund, "Two-stage part-based pedestrian detection," *Proceeding of IEEE International Conference on Intelligent Transportation Systems*, vol. 4, pp. 73–77, 2012.

[131] C. C. Chang and C. J. Lin, "Libsvm: A library for support vector machines," *ACM Transaction On Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.

[132] P. U. Brad Philip and M. Weber, "Car dataset from the rear, california institute of technology," http://www.vision.caltech.edu/archive.html.

[133] C. S. M. Marszaek, "Accurate object localization with shape masks," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[134] C. Papageorgiou and T. Poggio, "A trainable object detection system: Car detection in static images," Tech. Rep. 1673, October 1999, (CBCL Memo 180).

[135] V. L. M. Ozuysal and P.Fua, "Pose estimation for category specific multiview object localization," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 778–785, 2004.

[136] A. A. Shivani Agarwal and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1475–1490, 2009.

[137] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[138] M. Enzweiler and D. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Transactions on Pattern Analysis and Machine intelligence*, pp. 2179–2195, 2009.

[139] B. L. A. Ess and L. V. Gool, "Depth and appearance for mobile scene analysis," *Proceeding of IEEE International Conference on Computer Vision*, pp. 1–8, 2007.

[140] C. Papageorgio and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, pp. 15–33, 2000.

[141] N. B. L. A. G. Overett, L. Petersson and N. Pettersson, "A new pedestrian dataset for supervised learning," *Proceeding of IEEE Conference on Intelligent Vehicle Symposium*, pp. 373–378, 2008.

[142] B. S. C. Wojek, S. Walk and P. Perona, "Multi cue on board pedestrian detection," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 794–801, 2009.

[143] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 743–761, April 2012.

[144] M. S. Shirazi and B. T. Morris, "Vision-based pedestrian behavior analysis at intersections," *Journal of Electronic Imaging*, vol. 25, no. 5, p. 051203, 2016. [Online]. Available: http://dx.doi.org/10.1117/1.JEI.25.5.051203

[145] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, November 2000. [Online]. Available: http://dx.doi.org/10.1109/34.888718

[146] R. Scheneider, "Comparison of turning movement count data collection methods for a signal optimization study," URS Corporation, Tech. Rep., May 2011.

[147] S. Swann, "Alburycity comparison of traffic data collection methods," AlburyCity, Tech. Rep., October 2010.

[148] A. P. Tarko and R. S. Lyles, "Development of a portable video detection system for counting turning vehicles at intersections," Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, Tech. Rep. FHWA/IN/JTRP-2001/18, 2002.

[149] J. Gerken and B. Guy, "Accuracy comparison of non-intrusive, automated traffic volume counting equipment," Albeck Gerken, Inc., Tech. Rep., October 2009.

[150] M. S. Shirazi and B. Morris, "Vision-based vehicle counting with high accuracy for highways with perspective view," in *11th International Symposium, ISVC 2015*, Las Vegas, USA, December 2015, pp. 809–818.

[151] B. Morris and M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, pp. 425–437, Sept 2008.

[152] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 1999, p. 252 Vol. 2.

[153] L. V. G. Katja Nummiaro, Esther Koller-Meier, "Object tracking with an adaptive color-based particle filter," *Lecture Notes in Computer Science*, vol. 2449, no. 3, pp. 353–360, October 2002.

[154] F. Shahbaz Khan, R. Anwer, J. van de Weijer, A. Bagdanov, M. Vanrell, and A. Lopez, "Color attributes for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012, pp. 3306–3313.

[155] J. J. Bezuidenhout, P. Ranjitkar, and R. Dunn, "Estimating queue length at signalized intersections from single loop detector data," in *Proceedings of the Eastern Asia Society for Transportation Studies*, 2013.

[156] Q. Chai, C. Cheng, C. Liu, and H. Chen, "Vehicle queue length measurement based on a modified local variance and lbp," in *9th International Conference, ICIC 2013*, Nanning, China, July 2013, pp. 123–128.

[157] A. Albiol, A. Albiol, and J. Mossi, "Video-based traffic queue length estimation," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Nov 2011, pp. 1928–1932.

[158] M. Fathy and M. Siyal, "Real-time image processing approach to measure traffic queue parameters," *Vision, Image and Signal Processing, IEE Proceedings* -, vol. 142, no. 5, pp. 297–303, Oct 1995.

[159] M. Zanin, S. Messelodi, and C. Modena, "An efficient vehicle queue detection system based on image processing," in *Image Analysis and Processing, 2003.Proceedings. 12th International Conference on*, Sept 2003, pp. 232–237.

[160] C. Yingfeng, Z. Weigong, and W. Hai, "Measurement of vehicle queue length based on video processing in intelligent traffic signal control system," in *Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on*, vol. 2, March 2010, pp. 615–618.

[161] J. Zhou and L. Cheng, *An Efficient Vehicle Queue and Dissipation Detection Algorithm Based on Spatial-Temporal Markov Random Field*, ch. 49, pp. 468–474. [Online]. Available: http://ascelibrary.org/doi/abs/10.1061/9780784412442.049

[162] P. Venetianer, Z. Zhang, W. Yin, and A. Lipton, "Stationary target detection using the objectvideo surveillance system," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, Sept 2007, pp. 242–247.

[163] A. Bevilacqua and S. Vaccari, "Real time detection of stopped vehicles in traffic scenes," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, Sept 2007, pp. 266–270.

[164] A. Albiol, L. Sanchis, A. Albiol, and J. Mossi, "Detection of parked vehicles using spatiotemporal maps," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 4, pp. 1277–1291, Dec 2011.

[165] W. Wang, T. Gee, J. Price, and H. Qi, "Real time multi-vehicle tracking and counting at intersections from a fisheye camera," in *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, Jan 2015, pp. 17–24.

[166] M. S. Shirazi and B. Morris, "A typical video-based framework for counting, behavior and safety analysis at intersections," in *Intelligent Vehicles Symposium (IV), 2015 IEEE*, June 2015, pp. 1264–1269.

[167] S.-Y. Cho, T. Chow, and C.-T. Leung, "A neural-based crowd estimation by hybrid global learning algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 29, no. 4, pp. 535–541, Aug 1999.

[168] M. S. Shirazi and B. Morris, "Safety quantification of intersections using computer vision techniques," in *11th International Symposium, ISVC 2015*, Las Vegas, USA, December 2015, pp. 752–761.

[169] A. Prioletti, A. Mogelmose, P. Grisleri, M. Trivedi, A. Broggi, and T. Moeslund, "Part-based pedestrian detection and feature-based tracking for driver assistance: Real-time, robust algorithms, and evaluation," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, no. 3, pp. 1346–1359, Sept 2013.

[170] P. Reyad, T. Sayed, M. H. Zaki, and K. Shaaban, "Automated analysis of walking behavior: A case study from qatar," Washington, D.C., pp. 1–18, January 2015.

[171] H. Hediyeh, T. Sayed, and K. Ismail, "Automated analysis of pedestrian crossing speed behavior at scramble-phase signalized intersections using computer vision techniques," *International journal of sustainable transportation*, vol. 8, no. 5.

[172] M. H. Zaki, T. Sayed, and W. Xuesong, "Automatic classification of bike type (motorized vs non-motorized) during busy traffic in the city of shanghai," Washington, D.C., pp. 1–23, January 2015.

[173] A. Mogelmose, A. Prioletti, M. Trivedi, A. Broggi, and T. Moeslund, "Two-stage part-based pedestrian detection," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, Sept 2012, pp. 73–77.

[174] B. T. Morris and M. M. Trivedi, "Understanding vehicular traffic behavior from video: a survey of unsupervised approaches," *Journal of Electronic Imaging*, vol. 22, no. 4, pp. 041 113–041 113, 2013. [Online]. Available: http://dx.doi.org/10.1117/1.JEI.22.4.041113

[175] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, March 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944937

[176] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," *J. American Statistical Assoc.*, vol. 101, pp. 1566–1581, 2006.

[177] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[178] M. S. Shirazi and B. Morris, "Vision-based vehicle queue analysis at junctions," in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, Karlsruhe, Germany, August 2015, pp. 1–6.

[179] S. Yamamoto, K. Mizutani, and M. Sato, "Aichi dsss (driving safety support system) field verification test field verification test on vehicle infrastructure cooperative systems," in *The 13th ITS World Congress, LONDON, 8-12 OCTOBER 2006*, London, UK, 2006.

[180] J. Anaya, P. Merdrignac, O. Shagdar, F. Nashashibi, and J. Naranjo, "Vehicle to pedestrian communications for protection of vulnerable road users," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, June 2014, pp. 1037–1042.

Curriculum Vitae

Graduate College
University of Nevada, Las Vegas

Mohammad Shokrolah Shirazi

Degrees:

Bachelor's of Degree in Computer Engineering 2005
Ferdowsi University of Mashhad, Khorasan, Iran

Master's of Degree in Computer Engineering 2007
Sharif University of Technology, Tehran, Iran

Dissertation Title:Vision-based Intersection Monitoring: Behavior Analysis & Safety Issues

Dissertation Examination Committee:

Chairperson, Dr. Brendan Morris, Ph.D.
Committee Member, Dr. Pushkin Kachroo, Ph.D.
Committee Member, Dr. Emma Regentova, Ph.D.
Committee Member, Dr. Venkatesan Muthukumar, Ph.D.
Graduate Faculty Representative, Dr. Alexander Paz, Ph.D.

Publications:

1. M. S. Shirazi, B. T. Morris, "Vision-based Turning Movement Monitoring: Count, Speed & Waiting Time Estimation," *IEEE Intelligent Transportation Systems Magazine*, Vol. 8, No. 1, pp. 23-34, January, 2016.

2. M. S. Shirazi, B. T. Morris, "Vision-based Pedestrian Behavior Analysis at Intersections," *Journal of Electronic Imaging*, 25 (5), March, 2016.

3. M. S. Shirazi, B. T. Morris, "Looking at Intersections: A Survey of Intersection Monitoring, Behavior & Safety Analysis of Recent Studies," *Accepted in IEEE Transactions on Intelligent Transportation Systems.*

4. M. S. Shirazi, B. T. Morris, "Vision-based Vehicle and Pedestrian Tracking of Intersection Videos," *Accepted in International Journal on Artificial Intelligence Tools.*

5. M. S. Shirazi, B. Morris, "Vision-Based Pedestrian Monitoring at Intersections Including Behavior & Crossing Count," *Accepted in IEEE Intelligent Vehicles Symposium*, (IV 2016), June, 2016, Gothenburg, Sweden.

6. M. S. Shirazi, B. Morris, "Safety Quantification of Intersections Using Computer Vision Techniques," *11th International Symposium on Visual Computing*, (ISVC 2015), pp. 725-761, December, 2015, Las Vegas, USA.

7. M. S. Shirazi, B. Morris, "Vision-based Vehicle Counting with High Accuracy for Highways with Perspective View," *11th International Symposium on Visual Computing*, (ISVC 2015), pp. 809-818, December, 2015, Las Vegas, USA.

8. M. S. Shirazi, B. Morris, "Vision-based Vehicle Queue Analysis at Junctions," *12th IEEE International Conference on Advanced Video and Signal Based Surveillance*, (AVSS 2015), pp. 1-6, August, 2015, Karlsruhe, Germany.

9. M. S. Shirazi, B. Morris, "Observing Behaviors at Intersections: A Review of Recent Studies & Developments," *IEEE Intelligent Vehicles Symposium*, (IV 2015), pp. 1258-1263, June, 2015, Seoul, Korea.

10. M. S. Shirazi, B. Morris, "A Typical Video-based Framework for Counting, Behavior & Safety Analysis at Intersections," *IEEE Intelligent Vehicles Symposium*, (IV 2015), pp. 1264-1269, June, 2015, Seoul, Korea.

11. M. S. Shirazi, B. Morris, "Contextual Combination of Appearance and Motion for Intersection Videos with Vehicles and Pedestrians," *10th International Symposium in Visual Computing*, (ISVC 2014), pp. 708-717, December, 2014, Las Vegas, USA.

12. M. S. Shirazi, B. Morris, "Vision-Based Turning Movement Counting at Intersections by Cooperating Zone and Trajectory Comparison Modules," *IEEE International Conference on Intelligent Transportation Systems*, (ITSC 2014), pp. 3100-3105, October, 2014, Qindao, China.