DEVELOPING A COST-EFFECTIVE BUS-TO-PEDESTRIAN NEAR-MISS DETECTION METHOD USING ONBOARD VIDEO CAMERA DATA

FINAL PROJECT REPORT

by

Yinhai Wang, Ruimin Ke, Weibin Zhang, Ziqiang Zeng
University of Washington

Pacific Northwest Transportation Consortium (PacTrans)
Washington State Transit Insurance Pool (WSTIP)

for

Pacific Northwest Transportation Consortium (PacTrans)
USDOT University Transportation Center for Federal Region 10
University of Washington
More Hall 112, Box 352700
Seattle, WA 98195-2700

In cooperation with US Department of Transportation-Research and Innovative Technology Administration (RITA)

June 2018

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Pacific Northwest Transportation Consortium, the U.S. Government and matching sponsor assume no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
Developing a cost-effective bus-to-pedestrian near-miss detection method using onboard video camera data	06/30/2018
	6. Performing Organization Code
7. Author(s)	**8. Performing Organization Report No.**
Yinhai Wang, Ruimin Ke, Weibin Zhang	
9. Performing Organization Name and Address	**10. Work Unit No. (TRAIS)**
PacTrans Pacific Northwest Transportation Consortium University Transportation Center for Region 10 University of Washington More Hall 112 Seattle, WA 98195-2700	**11. Contract or Grant No.** DTRT13-G-UTC40
12. Sponsoring Organization Name and Address	**13. Type of Report and Period Covered**
United States of America Department of Transportation Research and Innovative Technology Administration	**14. Sponsoring Agency Code**

15. Supplementary Notes

Report uploaded at www.pactrans.org

16. Abstract

Bus-to-pedestrian near-miss data are important surrogate safety data for further pedestrian-related traffic safety studies. However, there is limited existing work on automatically extracting bus-to-pedestrian near-miss data from onboard camera. This project fills the gap by proposing a framework to automatically detect bus-to-pedestrian near-misses through onboard monocular vision with a real-time processing speed. The proposed detection framework has a different processing logic from previous vehicle-to-pedestrian conflict studies. First, our framework does not handle the complex background information in the moving onboard video directly, instead, it tries to locate the pedestrians based on the vision pattern. After the pedestrian being detected and tracked, we conduct the calculation in the 3D real-world coordinate instead of the 2D image coordinate as in previous studies. In the 2D image space no real-world value can be obtained. Specifically, our framework has four main stages: pedestrian detection in onboard video, motion estimation in the image coordinate, relative position, and speed calculation in the real-world coordinate, and near-miss detection.

In the first stage, the well-known Histogram-of-gradient pedestrian detector is used to detect pedestrian within the camera vision. In the second stage, interest points inside the detected bounding box of a pedestrian are tracked with sparse optical flow method. Thus, the motion of the pedestrian in the image coordinate can be estimated. In the third stage, with several camera parameters known and the assumption that the pedestrian detected is on the same plane with the vehicle, pedestrian's relative position and relative speed to the vehicle in the 3D real-world coordinate can be calculated. In the fourth stage, several near-miss indicators are used to determine if there is a potential vehicle-to-pedestrian near-miss event. The results turn out to be reasonably good by comparing with the events detected by a commercial system called MobilEye Shield+, which has multiple camera sensors installed. We have run the system on over one-month data and the overall performance is promising. Over 30 hours of data have been examined in detail for quantified evaluation purposes. The system processes in a nearly real-time manner, and yields over 85% detection overlap rate with the events extracted by the MobilEye Shield+ system. Two additional experiments are conducted to explore the applications of the proposed framework in solving practical traffic safety problems. With the findings and accomplishments in this project, an increasing amount of traffic safety data and a variety of practical applications are expected in the near future to support and advance traffic safety research.

17. Key Words	18. Distribution Statement
Bus-to-pedestrian near-miss, near-miss detection, onboard monocular video	No restrictions.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	31	NA

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Table of Contents

List of Figures

List of Tables

Acknowledgments

We would like to express our gratitude to Pacific Northwest Transportation Consortium (PacTrans) and Washington State Transit Insurance Pool (WSTIP) for the financial supports. We also thank King County Metro, Ben Franklin Transit, Community Transit, Kitsap Transit, Pierce Transit, Rosco Vision Systems, and MobilEye for their supports on the data and experiments.

Executive Summary

Pedestrians are vulnerable road users in multi-modal transportation systems, and pedestrian safety has been addressing more attention recently under the trend of green travel and smart city applications. However, the percentage of pedestrian fatalities increases by 3% in the past decade while the total traffic fatality decreases. Specifically, bus-to-pedestrian collisions often result in severe injuries, fatalities and huge insurance losses. According to Washington State Transit Insurance Pool, a large portion of the large collision-related transit insurance costs involve pedestrians in the Washington region. To a large extent, this is due to the lack of vehicle-to-pedestrian accident data. Researchers have been aware of this issue and trying to find surrogate safety measures. Near-miss is the major surrogate measure; and in order to extract sufficient near-miss events, a huge amount of data needs to be processed. For different data sources, it requires different automated near-miss detection methods. Onboard monocular vision system has been widely deployed in both public and personal vehicles. This data is cost-effective compared to onboard multiple-sensor systems or surveillance videos taken at fixed locations. But extracting events from onboard monocular vision is very challenging, and few works have been done.

This study fills the gap by proposing a framework to automatically detect bus-to-pedestrian near-misses through onboard monocular vision with real-time processing speed. The proposed detection framework has a different processing logic from previous vehicle-to-pedestrian conflict studies. First, our framework does not handle the complex background information in the moving onboard video directly. Instead, it tries to locate the pedestrians based on the vision pattern. After the pedestrian being detected and tracked, we conduct the calculation in the 3D real-world coordinate instead of the 2D image coordinate as in previous studies. In the 2D image space, no

real-world value can be obtained. Specifically, our framework has four main stages: pedestrian detection in onboard video, motion estimation in the image coordinate, relative position, and speed calculation in the real-world coordinate, and near-miss detection. In the first stage, the well-known Histogram-of-gradient pedestrian detector is used to detect pedestrian within the camera vision. In the second stage, interest points inside the detected bounding box of a pedestrian are tracked with sparse optical flow method. Thus, the motion of the pedestrian in the image coordinate can be estimated. In the third stage, with several camera parameters known and the assumption that the pedestrian detected is on the same plane with the vehicle, pedestrian's relative position and relative speed to the vehicle in the 3D real-world coordinate can be calculated. In the fourth stage, several near-miss indicators are used to determine if there is a potential bus-to-pedestrian near-miss event.

The results turn out to be reasonably good by comparing with the events detected by a commercial system, which has multiple camera sensors installed. We have run the system on over one-month data, and the overall performance is promising. Over 30 hours of data have been examined in detail for quantified evaluation. The system processes in a nearly real-time manner and yields over 85% detection overlap rate with the detected events from a well-developed MobilEye Shield+ system. Two additional experiments are conducted to explore the applications of the proposed framework in solving practical traffic safety problems. With the findings and accomplishments in this project, an increasing amount of traffic safety data and a variety of practical applications are expected in the near future to support and advance traffic safety research.

Chapter 1 Introduction

According to a report published by National Highway Traffic Safety Association (NHTSA) in 2015 (NHTSA, 2015), the number of total motor vehicle fatalities in the U.S. keeps decreased from 42,836 in 2004 to 32,719 in 2013. However, the annual number of pedestrian fatalities remained at the about the same level during the past decade. As a result, pedestrian fatalities as a percentage of total fatalities increased from 11% to 14%. And according to Washington State Transit Insurance Pool (WSTIP), despite pedestrian-involved accidents are relatively rare, they cause a large portion of insurance losses in the transit industry. More research is definitely needed to enhance pedestrian safety. Traditional traffic safety research normally relies on data about collisions. But collisions are rare events when considered in the context of normal measures of travel (Ismail et al., 2009). Other data measures of pedestrian activity such as pedestrian volume or speed are relatively rarely available compared with data for motor vehicle use. Consequently, the lack of appropriate pedestrian data makes it very challenging to draw solid conclusions on pedestrian safety improvements.

Researchers and engineers are aware of the lack of pedestrian collision data and started looking for surrogate safety measures. Despite slightly differing definitions in several studies, these surrogate events are commonly called near-misses. A near-miss is the conflict between road users that requires sudden evasive action and has the potential to develop into a collision. Collisions and near-miss events both can be used to measure the safety of certain locations or scenarios. (Guo et al., 2010) Near-misses have attracted more attention and have the potential to be used to explore factors that influence pedestrian safety. Research findings in this area will encourage a walking-friendly environment.

Near misses must be detected and extracted from specific data sources. Typically, the data sources for near-miss extraction will include records spanning a long time period, such as video records (Ismail et al., 2009; Laureshyn et al., 2010), records from in-vehicle sensors (Matsui et al., 2013), or even output from a simulation model of a certain location (Gettman and Head, 2003). Initially, surrogate safety measures were extracted manually, which was very inefficient and inaccurate (Chin and Quek, 1997; Guo et al., 2010; Zegeer and Deen, 1977). Recently, automated near-miss detection methods have been proposed in several studies, but few of them have used onboard monocular cameras (Ismail et al., 2009; Laureshyn et al., 2010; Matsui et al., 2013; Gettman and Head, 2003).

There are several advantages in using onboard monocular cameras as near-miss sensors: compared with surveillance video cameras which are installed at fixed locations with limited view coverage, onboard cameras are moving vision sensors that cover much larger areas; compared with using multiple in-vehicle sensors such as GPS units, radar sensors, and stereo vision systems, onboard monocular cameras are much cheaper, but may need more sophisticated algorithms to reach similar performance. Considering that many personal vehicles and public buses have installed onboard monocular cameras as standalone driver recorders, the recorded videos have huge potential to be turned into valuable datasets for traffic safety research. Since most developed traffic safety models require large volumes of data, the large number of existing onboard videos may be effective data sources if automated near-miss detection methods can be properly developed.

However, challenges do exist in near-miss detection in monocular cameras. First, with the moving background and moving foreground in the video, traditional background segmentation methods would not work as well as for stationary roadway surveillance videos (Zhang et al., 2007); also, in onboard front-facing cameras, the background points in different locations of a video frame do not share a similar motion, thereby identifying background points using "similar motion criterion" would get inaccurate results (Ke et al., 2017). With the recent progress in vision-based pedestrian detection and tracking, several studies have been completed showing that pedestrian detection and tracking algorithms could be potentially applied in vehicle-to-pedestrian collision avoidance and near-miss detection. However, these studies performed all calculations using two-dimensional image coordinates instead of real-world coordinates. Consequently, it was impossible for those algorithms to calculate real near-miss indicators, such as time-to-collision (TTC). To develop the correspondence between image coordinate and real-world coordinate, information from an extra dimension must be added. Two well-known methods use range measuring sensors such as radar or stereo vision, which tend to require expensive hardware. (Mori et al. 2007; Tsuji et al., 2002)

In this study, we propose a novel framework to extract bus-to-pedestrian near-misses from onboard monocular cameras automatically. This framework is composed of four main stages: 1) pedestrian detection, 2) motion estimation, 3) bus-to-pedestrian relative position and speed calculation, and 4) near-miss detection. In pedestrian detection, we made use of the well-known histogram of gradient (HOG) pedestrian detector. (Dalal and Triggs, 2005) A Kanade-Lucas-Tomasi (KLT) interest points tracker (Lucas and Kanade, 1981) is applied to track interest points inside the detection region to estimate the motion of the pedestrian in image coordinates. In the third stage, a camera model is built to find the correspondence between image coordinates and real-world coordinates. Then the relative position and relative speed can be calculated in real-world coordinates. Finally, using several defined thresholds, near-miss events can be detected and extracted from video clips. Our literature review did not reveal any significant published work about bus-to-pedestrian near-miss or conflict detection using onboard monocular videos. The work described in this report appears to be among the first efforts. Our study addresses several challenging issues including the moving video background issue, depth estimation, and real-world motion information extraction only using monocular video.

Chapter 2 Literature Review

2.1 Surrogate Safety Measures

Research on surrogate safety measures for pedestrian collisions has been on-going for several decades. Research initially focused on vehicle-to-vehicle near-misses on highways. (Zegeer and Head, 1977) In terms of research objective, previous studies can be roughly divided into two categories: development of traffic near-miss (conflict) analysis frameworks (Laureshyn et al., 2013; Matsui et al., 2013; Minderhoud and Bovy, 2001; Chin and Quek, 1997; Guo et al., 2010; Zegeer and Deen, 1977; Gettman and Head, 2003) and automated near-miss (conflict) detection methods. (Ismail et al., 2009; Guo et al., 2010; Tsuji et al., 2002; Kaparias et al., 2010; Ismail et al., 2010; Malkhamah et al., 2005; Wannige and Sonnadara, 2010; Kataoka et al., 2013)

Most analysis frameworks include time-to-collision (TTC) as a main indicator, although other indicators may be involved. There is, however, a lack of standard definitions for the indicator and analysis framework (Chin and Quek, 1997). Minderhoud et al. described two extended TTC measures for road traffic safety assessment (Minderhoud and Bovy, 2001). These two indicators consider the full course of vehicles over time and space, thereby giving a more comprehensive picture.

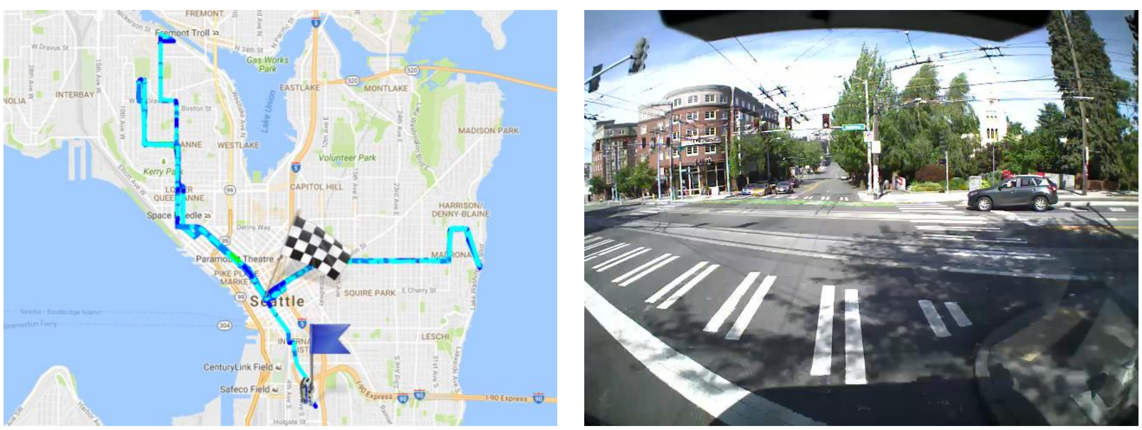
2.2 Vehicle-to-Pedestrian Near-Miss Extraction

When traffic near-miss studies were first conducted, data was manually collected (Chin and Quek, 1997; Guo et al., 2010; Zegeer and Deen, 1977). There are three main disadvantages in manual data collection: first, it is very time consuming for one or several people to stand at a specific location or go through recorded videos to find the near-miss events; second, different people have different judgements of what constitutes a near-miss event, it is very hard to guarantee the accuracy; third, it is impossible to quantitatively obtain safety measures of an event. As the demand for surrogate safety data has been increasing, the manual collection is no longer practical.

Ismail et al. developed an efficient method for vehicle-to-pedestrian near-miss detection using surveillance video data (Ismail et al., 2009). Their work is one of the key milestones in this field since their method could potentially be applied to all roadway surveillance videos. Malkhamah et al. developed an automatic method of safety monitoring using loop data (Malkhamah et al., 2005). Since loop data is still the main data source for traffic monitoring, this work makes it possible to detect conflicts on major freeways and arterials. However, loop detectors and video surveillance cameras installed at fixed locations limit safety monitoring to only those locations. Detectors installed on vehicles can monitor safety situation at many locations or along specific routes of interest. Some research has been conducted focusing on near-miss detection using onboard sensors. Tsuji et al. developed a system working in both day and night (Tsuji et al., 2002). They incorporate multiple sensing technologies including a stereo vision system which, however, is more expensive than a monocular camera. Several studies used vehicle-to-pedestrian conflict detection through monocular vision (Wannige and Sonnadara, 2010; Kataoka et al., 2013), but their methods still work in the 2D image space due to their logical frameworks. Thus, they are actually not able to calculate near-miss indicators.

Chapter 3 Study Site/Data

Data were collected on a King County Metro transit bus operating in downtown Seattle. Figure 3.1(a) replays the map of the bus trip on May 20, 2016. Onboard video data is collected by a Rosco Dual-Vision monocular camera and recorder system. The Rosco/MobilEye Shield Plus system is a vision-based vehicle, pedestrian, and bicyclist collision warning system designed for buses. This system has four cameras installed on the bus and can detect vehicle-pedestrian conflict events. The camera-based Shield+ system does not record video. However, the system detects potential collisions with pedestrians and vehicles based on time to collision and issues alerts and warnings to bus drivers using visible and audible indications. Events triggering the system are time-stamped, geolocated, coded, and transmitted to a server using a 3G telematics unit. Our method uses video from the Rosco onboard front-facing monocular camera (see Fig. 3-1(b)), and our results are compared with the Rosco/MobilEye Shield+ system's event data. The video for testing our method has a resolution of 640 × 480 pixels (width × height), and a frame rate of 7.5 frames-per-second (fps).



(a) (b)

Figure 3.1 (a) Map of King County Metro May 20, 2016 bus trip. (b) Sample frame from the front-facing video Dual-Vision camera.

Chapter 4 Methodology

The proposed detection framework (see Figure 4.1) has a different processing logic from previous vehicle-to-pedestrian conflict studies. Our framework does not handle the complex background information in the moving onboard video, but locates the pedestrian directly. Also, after the pedestrian is detected and tracked, we conduct the calculation in real-world coordinates instead of image coordinates. In the first stage, the HOG pedestrian detector is used to detect pedestrians within the camera's vision (Dalal and Triggs, 2005); in the second stage, the rectangle representing the pedestrian is tracked using KLT, (Lucas and Kanade, 1981) allowing the motion of the pedestrian to be estimated in image coordinates. In the third stage, with several camera parameters known, and the assumption that the pedestrian detected is on the same plane with the vehicle, the pedestrian's relative position and relative speed to the vehicle in 3D real-world coordinates can be calculated; in the fourth stage, several thresholds such as TTC need to be set to determine if there is a potential bus-to-pedestrian near-miss event.

4.1 Pedestrian Detection in Onboard Monocular Video

Pedestrian detection often plays a key role in multi-modal transportation engineering. Efficient and accurate pedestrian detection approaches would benefit traffic surveillance from many perspectives. Pedestrian detection is mainly based on the unique features of pedestrians. Generally, there are three types of single features used in pedestrian detection: gradient-based features, shape-based features, and motion-based features (Dollar et al., 2012). Motion-based features are not suitable for pedestrian detection in onboard videos as a single feature due to the complicated motion of traffic scene which is composed of moving background, and road users with random movements. Gradient-based and shape-based features are more suitable in our case. Our framework has an advantage that it is designed for a wide range of pedestrian detectors as long as they are based on the pedestrian pattern instead of motion information. In this paper, HOG is implemented as the pedestrian detector and the candidate pedestrian windows are identified using the sliding window approach. The input of the pedestrian detection is a video frame and the output is rectangle window(s) representing the pedestrian(s). In order for the following description, we denote p_{1_img} the point where the detected pedestrian's feet on. In other words, p_{1_img} is the midpoint of the pedestrian candidate window's bottom edge.

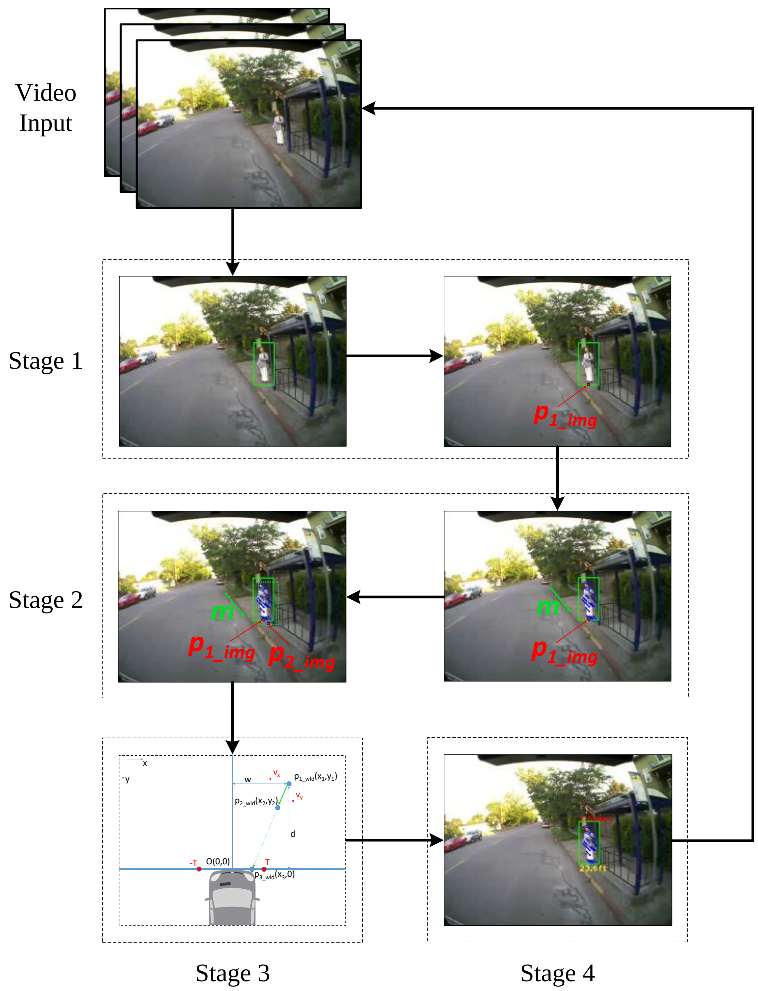


Figure 4.1 The proposed framework for bus-to-pedestrian near-miss detection through onboard monocular vision

4.2 Motion Estimation

In traffic video analysis, KLT tracker is very effective and has been widely used in motion analysis not only in surveillance videos with fixed background (Ismail et al., 2009, Kanhere et al., 2010) but also in aerial videos with moving background (Ke et al., 2016, Shastry and Schowengerdt, 2005). However, in onboard monocular videos, background motion is much more complex than that in either surveillance videos or aerial videos. Thus, instead of tracking points in the background and clustering them, in our framework, only those interest points in the detected region are tracked thereby background motion does not need to be handled. Basically, the average motion of those tracked points represents the relative motion of the detected pedestrians to the vehicle in the image coordinate. If *m* denotes the average motion of all the interest points within the rectangle, and p_{2_img} denotes the location of the pedestrian in the next frame (see Figure 4.1), we have

$$p_{2\_img} = p_{1\_img} + m. \tag{1}$$

4.3 Relative Position and Speed Estimation

With the pedestrian detected and motion *m* obtained, we developed a method to calculate the relative position and speed through monocular vision. In the image coordinate, as defined in the last sub-section, p_{1_img} and p_{2_img} are the pedestrian locations in two frames (see Figure 4.2(a)). We calculate their corresponding points (see Figure 4.2(b)) in the top-view of the real-world coordinate through a camera model as follows.

Let $C(u_0, v_0)$ be the center of the image coordinate and (u_1, v_1) is the position of p_{1_img}, then

$$du = u_1 - u_0 \tag{2}$$

$$dv = v_1 - v_{0,} \tag{3}$$

where *du* and *dv* are the differences between p_{1_img} and the image center.

To find the correspondence, four camera parameters are needed: camera focal length *f*, pixel length *l*, camera installation height *h*, and camera tilt angle θ. In the top-view of the real-world coordinate, the origin $O(0,0)$ is the camera center, whose location and motion are basically the same as the vehicle. Points p_{1_wld} and p_{2_wld} are the correspondences of p_{1_img} and p_{2_img}, respectively. Let x_1 and y_1 be the x-coordinate and y-coordinate of p_{1_img}. Then, x_1 and *y1* are related to *du* and *dv* by the following equations:

$$\phi = \arctan\left(\frac{l \times dv}{f}\right) + \theta, \tag{4}$$

where ϕ is the angle between the ground and the line connecting p_{1_wld} and $O(0,0)$. Thus, the depth value y_1 can be obtained, that is,

$$y_1 = \frac{h}{\arctan(\phi)}. \tag{5}$$

Then, with y_1 and *du* known, x_1 can be computed by the relation

$$x_1 = \frac{l \times du}{f} \times y_1. \tag{6}$$

In this way, the relative position of the pedestrian to the vehicle is obtained. Similar to the calculation of x_1 and y_1, x_2 and y_2 can be calculated. Let *fr* be the frame rate, then the relative speed *v* between pedestrian and the vehicle is

$$v = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \times fr. \tag{7}$$

Specifically, for relative speed components v_x and v_y in the *x*-axis and *y*-axis respectively, we have

$$v_x = (x_2 - x_1) \times fr, \tag{8}$$

$$v_y = (y_2 - y_1) \times fr. \tag{9}$$

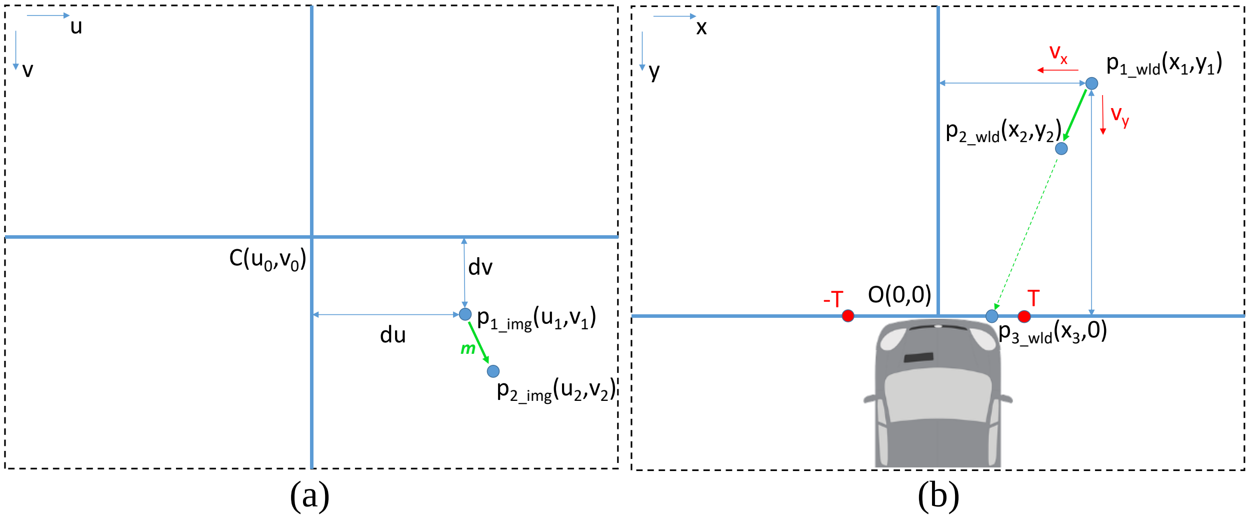


Figure 4.2 Method to find the correspondence between image coordinates and real-world coordinates.

<u>Near-miss Detection</u>

With the relative position and speed estimated through monocular vision, events can be judged by calculating near-miss indicators. The most commonly used indicator is TTC *(2-5)* and we also use TTC as the major near-miss indicator in this study, which can be obtained with the following equation

$$TTC = \frac{y_1}{v_y}, \tag{10}$$

where y_1 is the y-coordinate of the detected pedestrian in the real-world coordinate (see Figure 4.2(b)).

However, Eq. (10) alone is not sufficient to determine whether there is a near-miss, because even if the value got by Eq. (10) is very small, it is possible the horizontal component of the relative speed, i.e., v_x, is very large so that the pedestrian would not hit the vehicle following the current moving direction. Thus, another indicator is needed to be set to judge if the conflict will happen following the current relative speed on the x-axis. We define this indicator as distance-to-safety (DTS), which can be calculated as follows

$$DTS = v_x \times \frac{y_1}{v_y}. \tag{11}$$

Therefore, if both TTC and DTS are within their respective ranges for near-miss detection, i.e., $TTC < TTC_{threshold}$ and $-T < DTS < T$, where $TTC_{threshold}$ and T are the thresholds. T is shown in Figure 4.2(b), and it should be set no smaller than half of the vehicle width.

Chapter 5 Results

More than 30 hours of onboard monocular video data were used to test the performance of the proposed near-miss detection method. Figure 5.1 shows two representative samples identified as near-misses by our system. In (a), the vehicle was approaching a stop sign when two pedestrians were crossing the street. One of the pedestrians was detected as having the potential to collide with the vehicle if no evasive action was taken. In (b), a pedestrian standing at a bus stop was detected when the bus approached the stop and changed lanes.

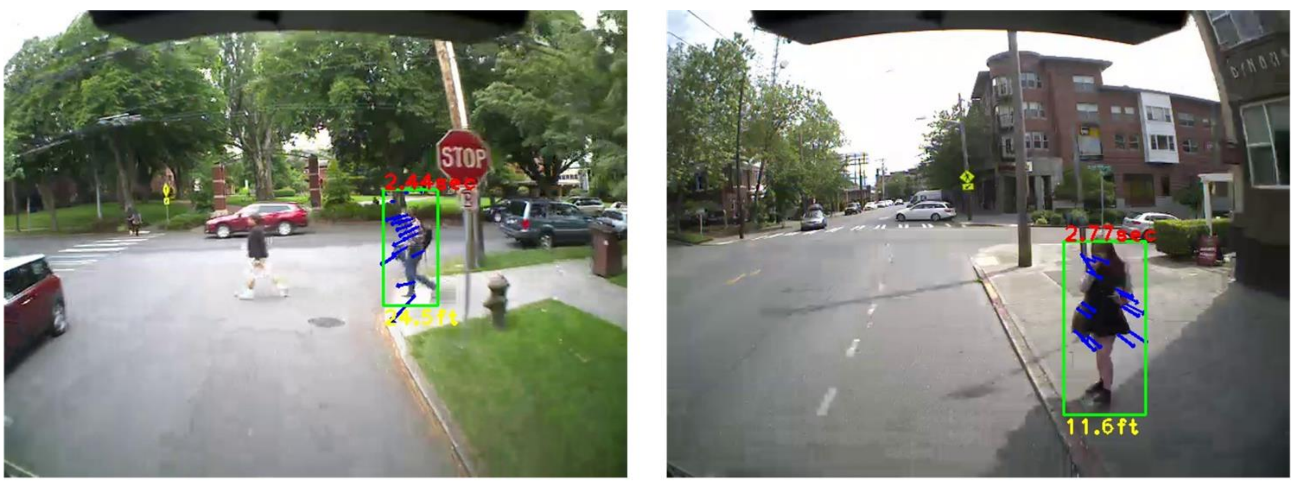
Video detection results are compared with events logged by the Rosco/MobilEye Shield+ system with multiple camera sensors. Different TTC thresholds are used in the experiments, and the results are presented in Table 5.1. In general, the corresponding detection overlap rate ($Overlap\ rate = (N_{TotalDetection} - N_{DifferentDetection}) / N_{TotalDetection}$) between the two systems ranges from 81.5% to 90.7%, with an average overlap rate of 86.9%. The largest overlap rate occurs when the TTC threshold is set to 2s. The results show that our video system detects the majority of near-misses picked up by the Shield+ system, but the difference still exists. We manually checked those video clips showing events that are not detected by both systems at the same time. Generally, we find there are three main reasons:

1) Some events occur at the side of the bus, and these events are not recorded by the onboard monocular camera. These events cannot be detected by our system because the target object (i.e., the pedestrian) does not appear in the view of the front-facing camera.
2) Some events detected by our system involve a pedestrian running towards the front of a stopped bus; a bus with no speed deactivates the Rosco/MobilEye system's bus-to-pedestrian near-miss detection function but the relative motion calculated by our system still indicates a potential conflict.

3) Some interest points inside the detected rectangle may come from objects other than the pedestrian such as corner points of lane markings, which could result in inaccurate motion estimation.

Table 5.1 Summary of the comparison results with the Rosco/MobilEye Shield+ system

$TTC_{threshold}$	4s	3s	2s	1s
Number of different detections	20	10	4	1
Number of total detections	108	81	43	8
Detection overlap rate	81.5%	87.7%	90.7%	87.5%



(a) (b)

Figure 5.1 Sample frames showing the representative near-miss events detected by the proposed system.

Chapter 6 Application of the Framework for Near-Miss Hotspot Identification

Besides safety surrogate data collection, another purpose for developing a cost-effective bus-to-pedestrian near-miss detection framework is to automatically identify hotspots and geographic distributions of events, to help drivers anticipate potential collisions in higher-risk locations. With the event data collected by our system, several plots displaying the distribution of the events are shown in Figure 6.1. It can be seen that most events occur at the right of the vehicle. This is reasonable since when a vehicle travels on the roadway, normally pedestrians appear to the right of it; the left of the vehicle is traffic moving along the opposite direction thereby few pedestrians appear. However, at intersections, pedestrians are likely to appear at different spots (rather than just right of the vehicle) from the driver’s perspective. By manually checking those frames with near-misses occurring at the left or middle of the vehicle, we find most of them do occur at intersections. For example, an event may occur when a left-turning vehicle has a conflict with a pedestrian crossing the street.

Also, we can see that the region with densest events are different in the image coordinate ((a), (b)) and the real-world coordinate ((c), (d)): the densest region in the image coordinate is the top right region, but in the real-world coordinate it is the bottom right region. That is to say, most near-misses occur at a relatively farther distance to the vehicle in the image coordinate intuitively, but closer to the vehicle in the real-world coordinate. This result is surprising at first glance, but the reason is that in the image coordinate, objects of the same size at a farther distance to the camera occupy fewer pixels than those closer; in other words, a pixel represents larger real-world size at a farther location to the camera. Thus, although the fact is more near-miss events occur in the region closer to the vehicle, it looks like more near-misses occur at a relatively farther distance in the image space. These findings may help drivers improve driving behavior and overall safety by knowing the distribution of near-misses.

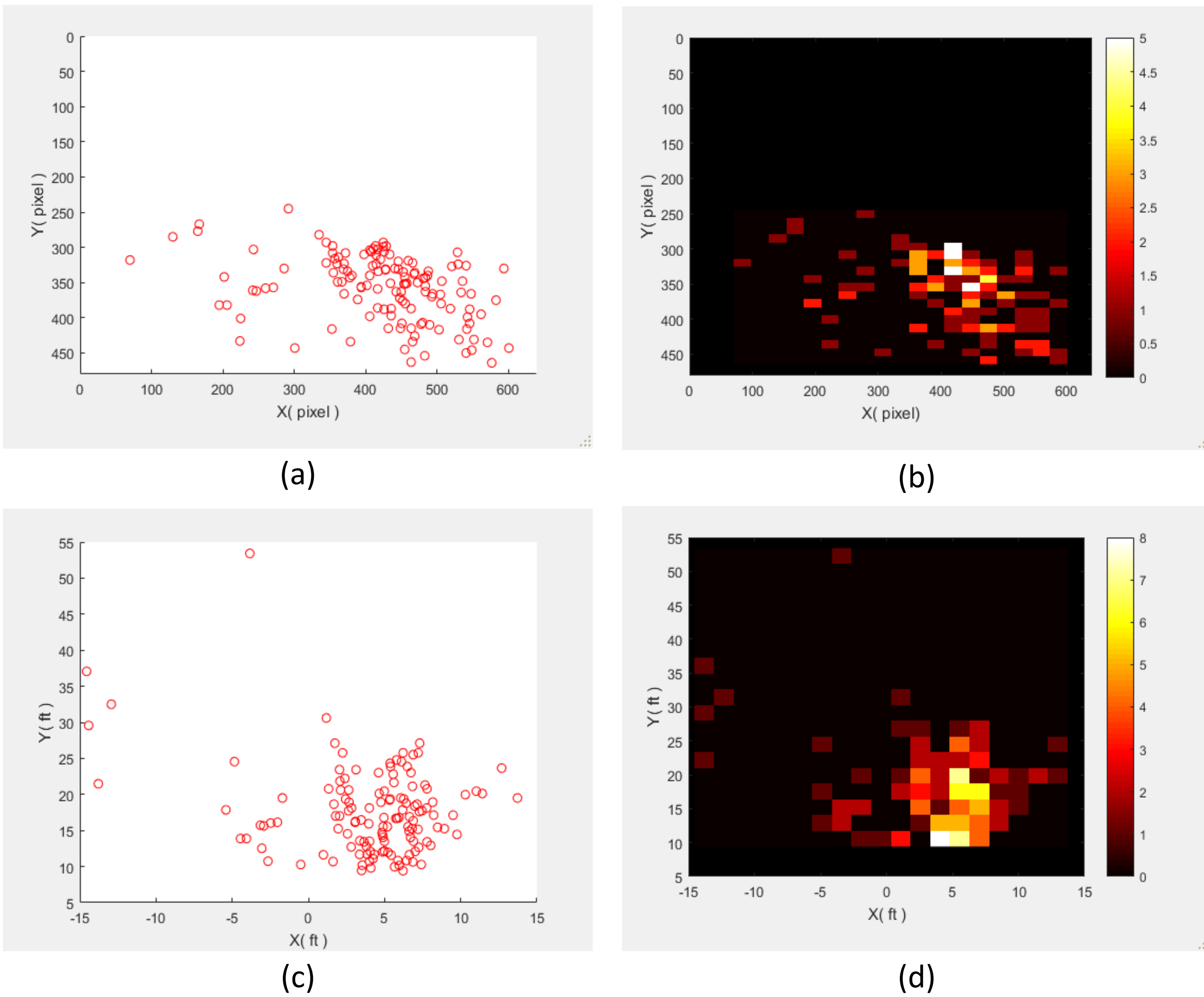


Figure 6.1 Scatter plots and heat maps showing the distribution of near-misses in image coordinates (a) and (b) and top-view of real-world coordinates, (c) and (d).

Chapter 7 Application of the Framework for Evaluating Emerging Technologies

Due to the cost-effectiveness of the proposed near-miss detection framework, it becomes a potential tool to assist in efficiently evaluating emerging technologies in intelligent transportation systems and autonomous driving. We explore this potential by using the framework to evaluate the detection performance of the Rosco/MobilEye Shield+ collision avoidance system. Specifically, we collect over 10 TB of monocular video data and bus-to-pedestrian near-miss events data on 38 transit buses equipped with the Shield+ system. The main purpose of our proposed framework/system in the evaluation process is data reduction. With proper parameters settings in our system, the system is able to efficiently reduce over 99% of the videos and only keep those with possible bus-to-pedestrian near-misses. With another round of manual checking by traffic safety experts, the bus-to-pedestrian near-miss detection performance of any collision avoidance system can be estimated. The two main metrics we use in the evaluation are false-positive (FP) rate and false-negative (FN) rate, which are standard measures of a detector. In this study, a FP is defined as a situation where the detector incorrectly indicates a presence of a near-miss event, and a FN indicates a missed near-miss event by the collision avoidance system.

It is important to determine how to define a false positive (FP) and false negative (FN) in order to maximize the identification accuracy. As aforementioned, the major indicator in our framework for near-miss detection is TTC. In order to set an appropriate TTC threshold for evaluation, we define a detection overlap rate (OR) to find the TTC threshold that would maximize OR, which is described with Eq. (12),

$$OR = \frac{A \cap B}{A \cup B} \tag{12}$$

where A and B are the sets of detections by each of the two technologies, i.e., our system and Shield+ system (see Figure 7.1). OR ranges from 0 to 1, and a larger OR would indicate a better TTC threshold with our tool to approximate the detection performance of Shield+ system. We have tested TTC threshold values from 1s to 5s, and find that 2s results in the largest OR about 0.9, which basically indicates 90 percent of the total events detected by either one of the two systems are also detected by both systems. However, in the practical evaluation process, we suggest setting the thresholds a little larger than 2s. Especially considering that the evaluation process includes both automatic detection and manual checking, setting the TTC larger would significantly reduce the probability of missing an event.

The identification of FP's is described in this paragraph. We run our framework on video clips labeled with events. If the Shield+ collision avoidance system also detects the same event, it is considered this is a true-positive ($A \cap B$). However, if no event is detected by our tool in the video clip, further checking is required. With the ROSCO video player, audio alert is available to precisely locate the time of an event in the video clip. Then the further checking process for FP's runs as follows: 1) locate the time of audio alert; 2) check the video around the alert to see if there is an observable near-miss; 3) if there is no observable near-miss such as no appearance of pedestrians around the bus or no obvious aggressive movement around the time of alert, this labeled event would be considered a FP ($FP \in A \cup B\text{-}A$).

Given the storage of videos over 10 TB, the identification of FN's is more time consuming. Our method aims at minimizing the manual checking time and maximizing the probability of finding all the FN's. The first step in identifying FN's is to run our software on the entire video dataset, thus to largely reduce the manual checking on outputted events. False detections of road users would be then filtered out first in the manual checking process. For example, a tree mistakenly recognized as a pedestrian will be discarded from the output immediately. Then the remaining detected events by our system are considered true events with the assumption that the KLT-based motion estimation process has no significant error. In the end, the events detected by our system but not the Shield+ are regarded as the FN's (FN$\in$A$\cup$B-B). Generally, given the limitation of the data available, perfectly recognizing all the FN's may be challenging. On the one hand, the union set of two systems' detected events (i.e, A$\cup$B) may not accurately represent the ground truth of all events that should be extracted (S may not be exactly equal to A$\cup$B). On the other hand, some of the "FN's" may be generated just because of the differences in the detection framework of the two systems. However, with the proposed cost-effective bus-to-pedestrian near-miss detection framework, this evaluation pipeline is efficient and effective to estimate the overall system detection performance.

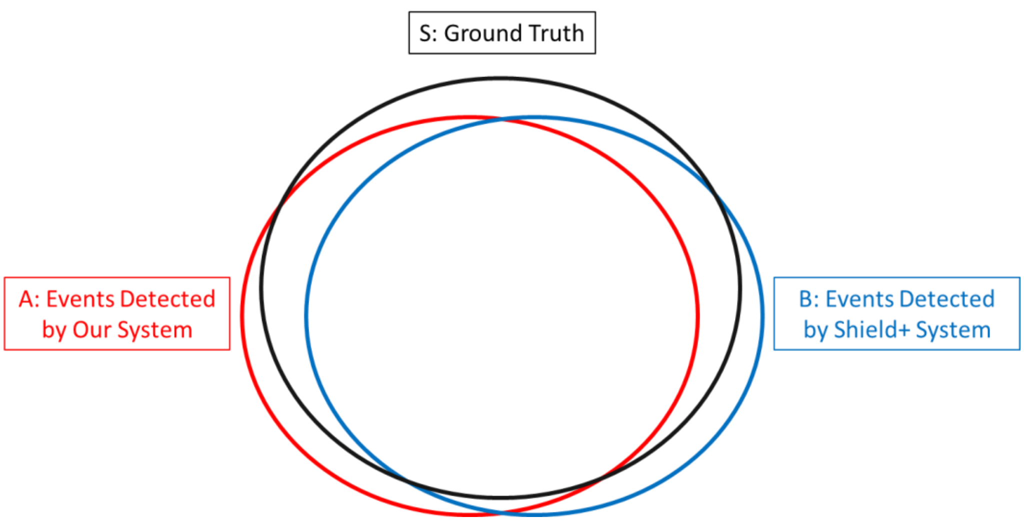


Figure 7.1 Relationships among three event sets, i.e., A: events detected by our system, B: events detected by Shield+ system, and S: ground truth.

Events and video data from five transit agencies are utilized: Ben Franklin Transit, Community Transit, King County Metro, Kitsap Transit, and Pierce Transit. Table 7.1 shows the summary statistics for the performance evaluation based on the videos have been fully processed so far. The total FP rate is about 3.21% and the FN rate is about 0.30%. In general, the Shield+ collision avoidance system demonstrates great performance. Although both the FP rate and FN rate are very low, the current system generates more FP's than FN's based on our evaluation. There are two typical FP patterns found during the testing period (See Figure 7.2). The first pattern is the false detection of road users, in which a collision warning is generated by the movement of the bus toward an object similar in shape to a pedestrian. For example, a standalone stop sign does not tend to trigger a warning, but a stop sign with cylinders around it would possibly does. The second typical pattern for FP's involves pedestrians/bicyclists moving parallel to and on the left of the bus either in the same or opposite direction. In some instances, pedestrians are on sidewalks at some distance and not on a trajectory to collide with the bus. The second pattern does not generate FP's for all buses, and may be caused by individual installation or parameter settings. Very few FN's were identified and no strong patterns emerged. Late detections are defined as FN's. Two example false-negatives identified by the proposed system are shown in Figure 7.3. Both (a) and (b) are detected by the Shield+ system but the warnings are late.

Table 7.1 1 Evaluation on the detection performances of the collision avoidance system

	Ben Franklin Transit	Community Transit	King County Metro	Kitsap Transit	Pierce Transit	Total
Events	1640	1062	430	1477	1461	6070
FP	111	24	7	39	14	195
FN	3	4	4	2	5	18
FP Rate	6.77%	2.26%	1.63%	2.64%	0.96%	3.21%
FN Rate	0.18%	0.38%	0.93%	0.14%	0.34%	0.30%

(a) (b)

Figure 7.2 Two typical patterns of false-positive have been observed.

(a) (b)

Figure 7.3 Two false-negatives identified by our system. Both (a) and (b) are detected by the Shield+ system but the warnings are late.

Chapter 8 Conclusions and Recommendations

A cost-effective framework for automated bus-to-pedestrian near-miss detection through onboard monocular vision is proposed in this project. It aims at automatically extracting bus-to-pedestrian surrogate safety measure data, i.e., bus-to-pedestrian near-miss events, using onboard monocular camera. The framework incorporates a HOG pedestrian detector and KLT tracker to detect and track pedestrians appearing in the monocular camera. Then it calculates the region of interest and estimates motion in image coordinates. With known camera parameters, a camera model is built to find the correspondence between image coordinates and real-world coordinates of detected pedestrians. Using this correspondence, we calculate the relative speed and relative position information and then manage to obtain the near-miss indicators. This framework is among the first efforts for detecting bus-to-pedestrian near-misses by using onboard monocular video. It can be applied to both safety surrogate data collection and collision avoidance tasks for most types of vehicles. The experiment shows our system works reasonably well by the comparison with Rosco/MobilEye Shield+ system which includes four camera sensors.

Based on the experimental results and analyses in this study, future work is currently planned for the following aspects. First, future work will involve testing the system in more challenging scenarios such as bus approaching a crowd of pedestrians thus to further improve the overall performance. Second, errors in motion estimation may occur due to that some of the interest points may not come from the pedestrians but other objects appearing in the candidate windows. Hence, in future work, we plan to implement a method to filter out those extraneous interest points. Third, instead of validating the proposed framework with a vision-based system, it would be helpful to also compare it with more advanced systems such as a system incorporating both vision and radar sensors.

References

National Highway Traffic Safety Association, US Department of Transportation: “Traffic Safety Facts 2013.” Available at http://www.nhtsa.gov/lifesavers/assets/812124.pdf, accessed on July 18, 2017.

Ismail, K., Sayed, T., Saunier, N. and Lim, C., 2009. “Automated analysis of pedestrian-vehicle conflicts using video data.” *Transportation Research Record: Journal of the Transportation Research Board*, (2140), pp.44-54.

Laureshyn, A., Svensson, Å. and Hydén, C., 2010. “Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation.” *Accident Analysis & Prevention*, *42*(6), pp.1637-1646.

Matsui, Y., Hitosugi, M., Doi, T., Oikawa, S., Takahashi, K. and Ando, K., 2013. “Features of pedestrian behavior in car-to-pedestrian contact situations in near-miss incidents in Japan.” *Traffic injury prevention*, *14*(sup1), pp.S58-S63.

Minderhoud, M.M. and Bovy, P.H., 2001. “Extended time-to-collision measures for road traffic safety assessment.” *Accident Analysis & Prevention*, *33*(1), pp.89-97.

Chin, H.C. and Quek, S.T., 1997. “Measurement of traffic conflicts.” *Safety Science*, *26*(3), pp.169-185.

Guo, F., Klauer, S.G., McGill, M.T. and Dingus, T.A., 2010. “Evaluating the relationship between near-crashes and crashes: Can near-crashes serve as a surrogate safety metric for crashes?.”

Zegeer, C.V. and Deen, R.C., 1977. “Traffic conflicts as a diagnostic tool in highway safety.”

Gettman, D. and Head, L., 2003. “Surrogate safety measures from traffic simulation models.” *Transportation Research Record: Journal of the Transportation Research Board*, (1840), pp.104-115.

Zhang, G., Avery, R. and Wang, Y., 2007. “Video-based vehicle detection and classification system for real-time traffic data collection using uncalibrated video cameras.” *Transportation Research Record: Journal of the Transportation Research Board*, (1993), pp.138-147.

Ke, R., Li, Z., Kim, S., Ash, J., Cui, Z. and Wang, Y., 2017. “Real-time bidirectional traffic flow parameter estimation from aerial videos.” *IEEE Transactions on Intelligent Transportation Systems*, *18*(4), pp.890-901.

Mori, K., Takahashi, T., Ide, I., Murase, H., Miyahara, T. and Tamatsu, Y., 2007, June. "Recognition of foggy conditions by in-vehicle camera and millimeter wave radar." In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 87-92). IEEE.

Tsuji, T., Hattori, H., Watanabe, M. and Nagaoka, N., 2002. "Development of night-vision system." *IEEE Transactions on Intelligent Transportation Systems*, *3*(3), pp.203-209.

Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893). IEEE.

Lucas, B.D. and Kanade, T., 1981. "An iterative image registration technique with an application to stereo vision." In *IJCAI* (Vol. 81, pp. 674-679).

Kaparias, I., Bell, M., Greensted, J., Cheng, S., Miri, A., Taylor, C. and Mount, B., 2010. "Development and implementation of a vehicle-pedestrian conflict analysis method: Adaptation of a vehicle-vehicle technique." *Transportation Research Record: Journal of the Transportation Research Board*, (2198), pp.75-82.

Ismail, K., Sayed, T. and Saunier, N., 2010. "Automated analysis of pedestrian-vehicle: conflicts context for before-and-after studies." *Transportation Research Record: Journal of the Transportation Research Board*, (2198), pp.52-64.

Malkhamah, S., Tight, M. and Montgomery, F., 2005. "The development of an automatic method of safety monitoring at Pelican crossings." *Accident Analysis & Prevention*, *37*(5), pp.938-946.

Wannige, C.T. and Sonnadara, D.U.J., 2010. "Pedestrian Collision Detection through Monocular Vision."

Kataoka, H., Tamura, K., Aoki, Y., Matsui, Y., Iwata, K. and Satoh, Y., 2013, "November. Robust feature descriptor and vehicle motion model with tracking-by-detection for active safety." In *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE* (pp. 2472-2477). IEEE.

Dollar, P., Wojek, C., Schiele, B. and Perona, P., 2012. "Pedestrian detection: An evaluation of the state of the art." *IEEE transactions on pattern analysis and machine intelligence*, *34*(4), pp.743-761.

Shastry, A.C. and Schowengerdt, R.A., 2005. "Airborne video registration and traffic-flow parameter estimation." *IEEE Transactions on Intelligent Transportation Systems*, *6*(4), pp.391-405.

Kanhere, N., Birchfield, S., Sarasua, W. and Khoeini, S., 2010. “Traffic monitoring of motorcycles during special events using video detection.” *Transportation Research Record: Journal of the Transportation Research Board*, (2160), pp.69-76.