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)

Disclaimer

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Developing a cost-effective bus-to-pedestrian near-miss detection method using onboard video camera data		**5. Report Date** 01/31/2018 **6. Performing Organization Code**
7. Author(s) Yinhai Wang, Ruimin Ke, Weibin Zhang		**8. Performing Organization Report No.**
9. Performing Organization Name and Address PacTrans Pacific Northwest Transportation Consortium University Transportation Center for Region 10 University of Washington More Hall 112 Seattle, WA 98195-2700		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** DTRT13-G-UTC40
12. Sponsoring Organization Name and Address United States of America Department of Transportation Research and Innovative Technology Administration		**13. Type of Report and Period Covered** **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 cameras. This project fills the gap by proposing a framework to automatically detect bus-to-pedestrian near-misses through an onboard monocular vision system with real-time processing speed. The proposed detection framework has a different processing logic than 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 on the basis of the vision pattern. After the pedestrian has been detected and tracked, the calculation occurs in 3D, real-world coordinates instead of 2D image coordinates, 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 with onboard video, motion estimation in image coordinates, relative position and speed calculation in real-world coordinates, and near-miss detection.

In the first stage, the well-known Histogram-of-gradient pedestrian detector is used to detect pedestrians within the camera's vision. In the second stage, interest points inside the detected bounding box of a pedestrian are tracked with the sparse optical flow method. Thus, the motion of the pedestrian in the image coordinates can be estimated. In the third stage, with several camera parameters known and the assumption that the detected pedestrian is on the same plane as the vehicle, the pedestrian's position and speed relative to the vehicle in the 3D real-world coordinates can be calculated. In the fourth stage, several near-miss indicators are used to determine whether a vehicle-to-pedestrian near-miss event is possible. In comparison with events detected by a commercial system called MobilEye Shield+ that had multiple camera sensors installed, the results turned out to be reasonably good. We ran the system with over one-month of data, and the overall performance was promising. Over 30 hours of data were examined in detail for quantified evaluation purposes. The system processed in a nearly real-time manner and yielded an over 85 percent detection overlap rate with the events extracted by the MobilEye Shield+ system. With the findings and accomplishments in this project, a much larger number of bus-to-pedestrian conflict data are expected to be collected from onboard videos to support and advance future 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 want to express our gratitude to the Pacific Northwest Transportation Consortium (PacTrans) and the Washington State Transit Insurance Pool (WSTIP) for their financial support. We also thank King County Metro, Rosco Vision Systems, and MobilEye for their support on the data and experiments.

Executive Summary

Pedestrians are vulnerable road users in multi-modal transportation systems, and pedestrian safety has been given more attention recently under the umbrella of green travel and smart city applications. However, the percentage of pedestrian fatalities increased by 3 percent in the past decade, while the total traffic fatalities decreased. As one example, bus-to-pedestrian collisions often result in severe injuries, fatalities, and huge insurance losses. According to the Washington State Transit Insurance Pool, a large portion of high collision-related transit insurance costs in Washington state involve pedestrians. To a large extent, this is due to a lack of vehicle-to-pedestrian accident data. Researchers have been aware of this issue and have been trying to find surrogate safety measures. The "near-miss" is the primary surrogate measure, and in order to extract sufficient near-miss events, a huge number of data need to be processed. For different data sources, different automated near-miss detection methods are required. Onboard monocular vision systems have been widely deployed in both public and personal vehicles. The resulting data are cost effective in comparison to data from onboard multiple-sensor systems or surveillance videos taken at fixed locations. But extracting events from onboard monocular vision systems is very challenging, and few efforts have been undertaken.

This study fills the gap by proposing a framework to automatically detect bus-to-pedestrian near-misses through onboard monocular vision systems with real-time processing speed. The proposed detection framework has a different processing logic than 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 pedestrians on the basis of vision patterns. After pedestrians have been detected and tracked, calculations are

conducted in 3D real-world coordinates instead of with 2D image coordinates as in previous studies. In the 2D image space, no real-world values can be obtained.

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

A comparison with events detected by a commercial system with multiple camera sensors installed showed that he results were reasonably good. We ran the system on over one month of data, and the overall performance was promising. Over 30 hours of data were examined in detail for a quantified evaluation. The system processed in a nearly real-time manner and yielded an over 85 percent detection overlap rate with the detected events from a well-developed MobilEye Shield+ system. With the findings and accomplishments in this project, a much larger number of bus-to-pedestrian conflict data can be collected from onboard videos to support and advance future traffic safety research.

1. Introduction

According to a report published by National Highway Traffic Safety Association (NHTSA) in 2013 (NHTSA, 2013), the number of total motor vehicle fatalities in the U.S. has decreased from 42,836 in 2004 to 32,719 in 2013. However, the annual number of pedestrian fatalities has 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 percent to 14 percent. And according to the Washington State Transit Insurance Pool (WSTIP), although 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 in comparison with data for motor vehicle use. Consequently, a lack of appropriate pedestrian data makes it very challenging to draw solid conclusions about pedestrian safety improvements.

Researchers and engineers are aware of the lack of pedestrian collision data and have 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 a 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 attention and have the potential to be used to explore factors that influence pedestrian safety. Research findings in this area will encourage a more pedestrian-friendly environment.

Near-misses must be detected and extracted from specific data sources. Typically, the data sources for near-miss extraction 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), and 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 data from onboard monocular cameras (Ismail et al., 2009; Laureshyn et al., 2010; Matsui et al., 2013; Gettman and Head, 2003).

There are several advantages to using onboard monocular cameras as near-miss sensors. In comparison with surveillance video cameras that are installed at fixed locations with limited view coverage, onboard cameras are moving vision sensors that cover much larger areas. Furthermore, in comparison with multiple in-vehicle sensors such as GPS units, radar sensors, and stereo vision systems, onboard monocular cameras are much cheaper, although they may need more sophisticated algorithms to reach similar performance. Given that many personal vehicles and public buses have onboard monocular cameras installed as standalone driver recorders, the recorded videos have huge potential to be turned into valuable data sets for traffic safety research. Because 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 for near-miss detection with monocular cameras. First, because of 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, with onboard front-facing cameras, the background points in different locations of a video frame do not share a similar motion; therefore, identifying background points using a "similar motion criterion" would produce inaccurate results (Ke et al., 2017). With recent progress in vision-based pedestrian detection and tracking, several studies have shown that pedestrian detection and tracking algorithms could be potentially applied to vehicle-to-pedestrian collision avoidance and near-miss detection. However, these studies performed all calculations using two-dimensional (2D) image coordinates instead of real-world coordinates. Consequently, it is impossible for those algorithms to calculate real near-miss indicators, such as time-to-collision (TTC). To develop correspondence between image coordinates and real-world coordinates, 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-miss data from onboard monocular cameras automatically. This framework comprises four main stages: 1) pedestrian detection, 2) motion estimation, 3) bus-to-pedestrian relative position and speed calculation, and 4) near-miss detection.

1. In pedestrian detection, we make use of the well-known histogram of gradient (HOG) pedestrian detector (Dalal and Triggs, 2005).
2. 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.

3. In the third stage, a developed camera model is finds the correspondence between image coordinates and real-world coordinates. Then the relative position and speed can be calculated in real-world coordinates.
4. 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 using only monocular video.

2. Literature Review

2.1 Surrogate Safety Measures

Research to find surrogate safety measures for pedestrian collisions has been ongoing for several decades. Research initially focused on vehicle-to-vehicle near-misses on highways (Zegeer and Head, 1977). In terms of research objectives, previous studies can be roughly divided into two categories: development of traffic near-miss (conflict) analysis frameworks (Laureshyn et al., 2015; 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 and Bovy 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 were manually collected (Chin and Quek, 1997; Guo et al., 2010; Zegeer and Deen, 1977). There are three main disadvantages to manual data collection:

1. It is very time consuming for one or several people to stand at a specific location or go through recorded videos to find near-miss events.

2. Different people have different judgments of what constitutes a near-miss event, so it is very hard to guarantee accuracy.
3. It is impossible to quantitatively obtain safety measures of an event.

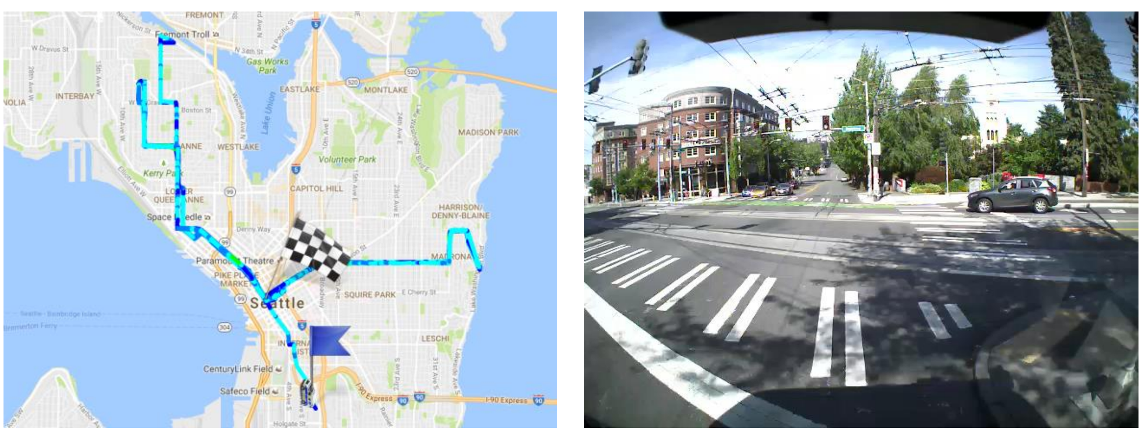
As the demand for surrogate safety data has been increasing, manual collection is no longer practical.

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 for safety monitoring using loop data (Malkhamah et al., 2005). Because loop data are 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 situations at many locations or along specific routes of interest.

Some research has been conducted on near-miss detection using onboard sensors. Tsuji et al. developed a system that works in both day and night (Tsuji et al., 2002). They incorporated multiple sensing technologies, including a stereo vision system, which, however, is more expensive than a monocular camera. Several studies have 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 because of their logical frameworks. Therefore, they are actually not able to calculate near-miss indicators.

3. Study Site and Data

Data were collected on a King County Metro transit bus operating in downtown Seattle. Figure 3.1(a) shows the map of the bus trip on May 20, 2016. Onboard video data were collected by a Rosco Dual-Vision monocular camera and recorder system. The Rosco/MobilEye Shield+ system is a vision-based vehicle, pedestrian, and bicyclist collision warning system designed for buses. This system includes four cameras 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 on the basis of time to collision and issues alerts and warnings to bus drivers via visible and audible indicators. Events that trigger the system are time-stamped, geolocated, coded, and transmitted to a server using a 3G telematics unit. Our method used video from the Rosco onboard front-facing monocular camera (see figure 3-1(b)), and our results were compared with the Rosco/MobilEye Shield+ system’s event data. The video for testing our method had 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.

4. Method

The proposed detection framework (see figure 4.1) has a different processing logic than previous vehicle-to-pedestrian conflict studies. Our framework does not handle the complex background information in the moving onboard video, but rather locates the pedestrian directly. Also, after the pedestrian has been detected and tracked, the calculation is conducted in real-world coordinates instead of image coordinates. In the first stage, the histogram of gradient (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 a Kanade-Lucas-Tomasi (KLT) interest points tracker (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 detected pedestrian is on the same plane as the vehicle, the pedestrian's position and speed relative to the vehicle can be calculated in 3D real-world coordinates. In the fourth stage, several thresholds such as TTC need to be set to determine whether there is potential for a bus-to-pedestrian near-miss event.

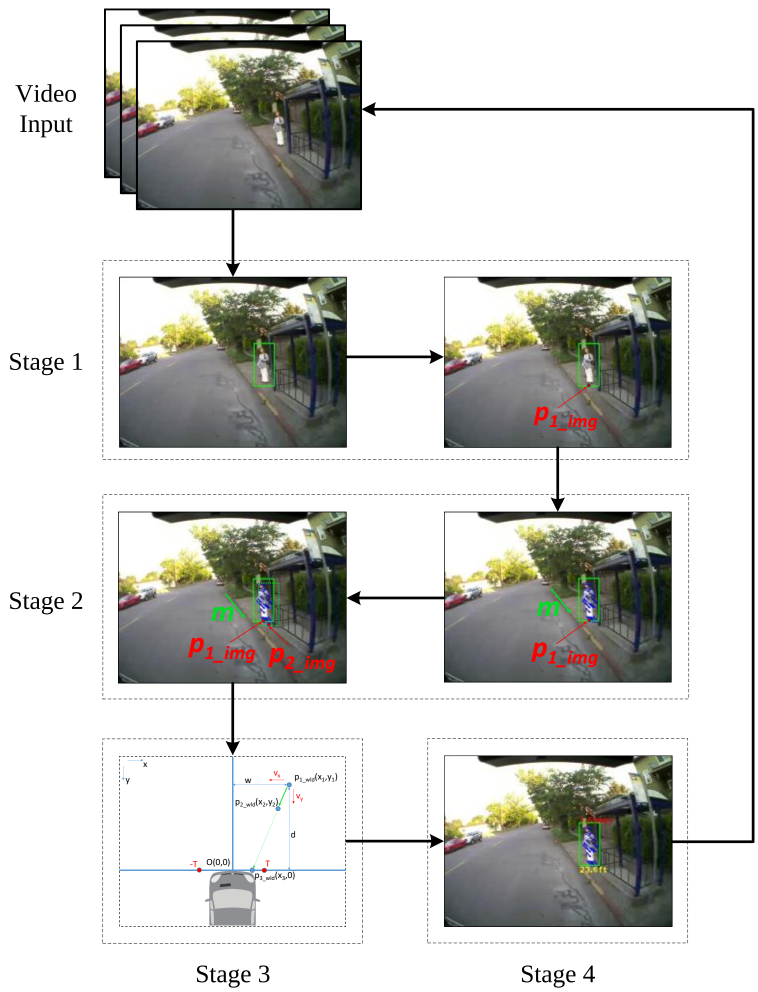


Figure 4.1 The proposed framework for bus-to-pedestrian near-miss detection through onboard monocular vision

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, three types of single features are 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 because the complicated motion of the traffic scene is composed of a moving background and road users with random movements. Gradient-based and shape-based features are more suitable for our use. Our framework has an advantage in that it is designed for a wide range of pedestrian detectors, as long as they are based on pedestrian patterns instead of motion information. For this study, HOG was implemented as the pedestrian detector, and the candidate pedestrian windows were identified by using the sliding window approach. The input for pedestrian detection is a video frame, and the output is a rectangular window(s) representing the pedestrian(s). For this work, we denote p_{1_img} the point of the detected pedestrian's feet. In other words, $p_{1_img}$is the midpoint of the pedestrian candidate window's bottom edge.

4.2 Motion Estimation

In traffic video analysis, the KLT tracker is very effective and has been widely used in motion analysis, not only in surveillance videos with a fixed background (Ismail et al., 2009, Kanhere et al., 2010) but also in aerial videos with a moving background (Ke et al., 2017, Shastry and Schowengerdt, 2005). However, in onboard monocular videos, background motion is much more complex than it is in either surveillance videos or aerial videos. Therefore, instead of tracking points in the background and clustering them, in our framework, only points of interest in the detected region are tracked. In this way, background motion does not need to be handled. Basically, the average motion of the tracked points represents the motion of the detected pedestrians relative to the vehicle in the image coordinates. 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), then 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 use a method developed 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 coordinates through a camera model as follows.

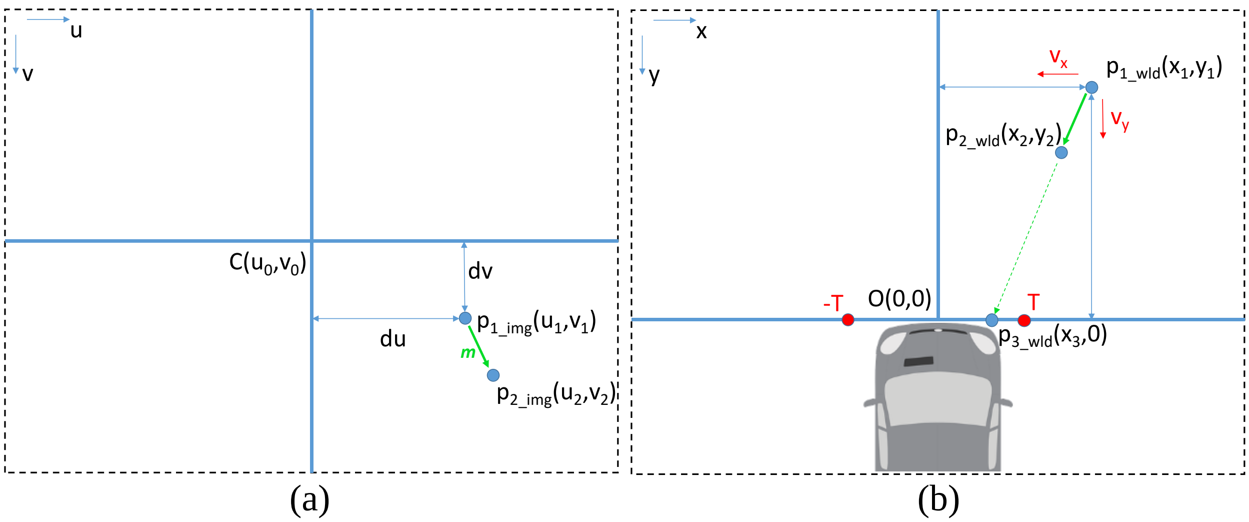


Figure 4.2 Method to find the correspondence between image coordinates and real-world coordinates.

If $C(u_0, v_0)$ is 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 coordinates, 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. If x_1 and y_1 are 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. If *fr* is 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

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

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

4.4 Near-Miss Detection

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 (Ismail et al. 2009, Laureshyn et al. 2010, Matsui et al. 2013, Minderhoud and Bovy 2001), and we also used TTC as the major near-miss indicator in this study. TTC 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 coordinates (see figure 4.2(b)).

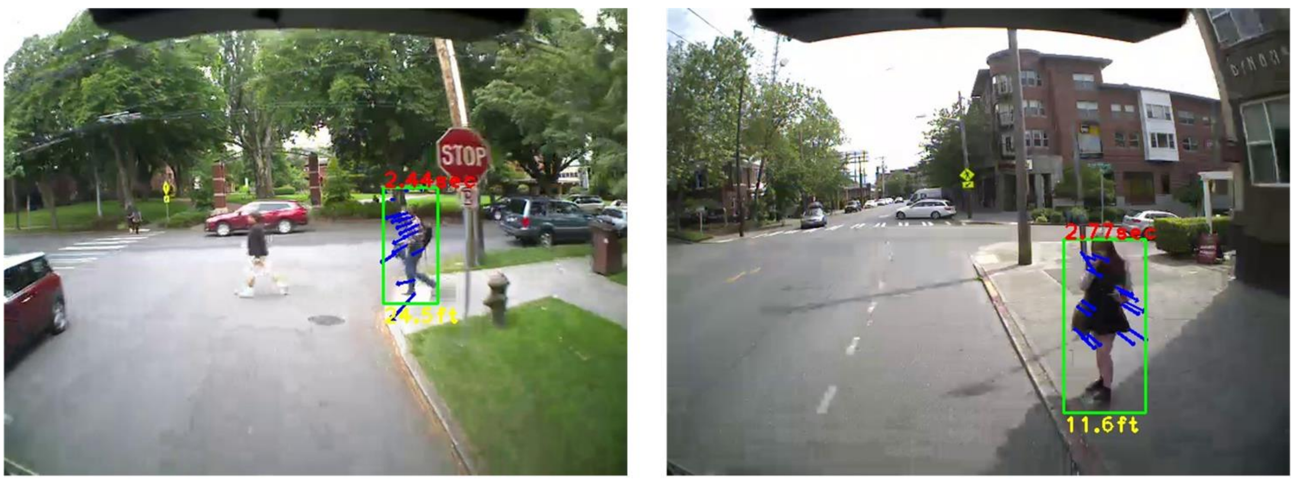
However, equation (10) alone is not sufficient to determine whether a near-miss is possible, because even if the value obtained by equation (10) is very small, it is possible that the horizontal component of the relative speed, i.e., v_x, is very large, so that the pedestrian would not connect with the vehicle following the current moving direction. Therefore, another indicator is needed to judge whether a 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:

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

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, then a near-miss is possible. T is shown in figure 4.2(b), and it should be set no smaller than half of the vehicle width.

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.



(a) (b)

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

Video detection results were compared with events logged by the Rosco/MobilEye Shield+ system with multiple camera sensors. Different TTC thresholds were 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 ranged from 81.5 percent to 90.7 percent, with an average overlap rate of 86.9 percent. The largest overlap rate occurred when the TTC threshold was set to 2s. The results showed that our video system detected the majority of near-misses picked up by the Shield+ system, but differences still existed. We manually checked those video clips showing events that were not detected by both systems at the same time. Generally, we found there were three main reasons:

1) Some events occurred at the side of the bus, and those events were not recorded by the onboard monocular camera. Those events could not be detected by our system because the target object (i.e., the pedestrian) did not appear in the view of the front-facing camera.
2) Some events detected by our system involved a pedestrian running toward 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 indicated a potential conflict.
3) Some interest points inside the detected rectangle may have come from objects other than a pedestrian, such as corner points of lane markings, which might have resulted 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%

6. Discussion

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 distributions of the events were developed, as shown in figure 6.1. The figure shows that most events occurred at the right of the vehicle. This is reasonable because when a vehicle travels on the roadway, normally pedestrians appear to the right of it; to the left of the vehicle traffic moves in the opposite direction, so 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 found that most of them did occur at intersections. For example, an event may occur when a left-turning vehicle has a conflict with a pedestrian crossing the street.

Also, were able to see that the regions with densest events were different in the image coordinates ((a), (b)) and the real-world coordinates ((c), (d)). The densest region in the image coordinates was the top right region, but in the real-world coordinates it was the bottom right region. That is to say, most near-misses occur at a relatively farther distance from the vehicle in the image coordinates intuitively, but closer to the vehicle in the real-world coordinates. This result was surprising at first glance, but the reason is that in the image coordinates, objects of the same size at a farther distance from the camera occupy fewer pixels than those closer. In other words, a pixel represents a larger real-world size at a farther location from the camera. Therefore, although in fact more near-miss events occurred in the region closer to the vehicle, it looked like more near-misses occurred at a relatively farther distance in the image space. Knowing the distribution of near-misses may help drivers improve driving behavior and overall safety.

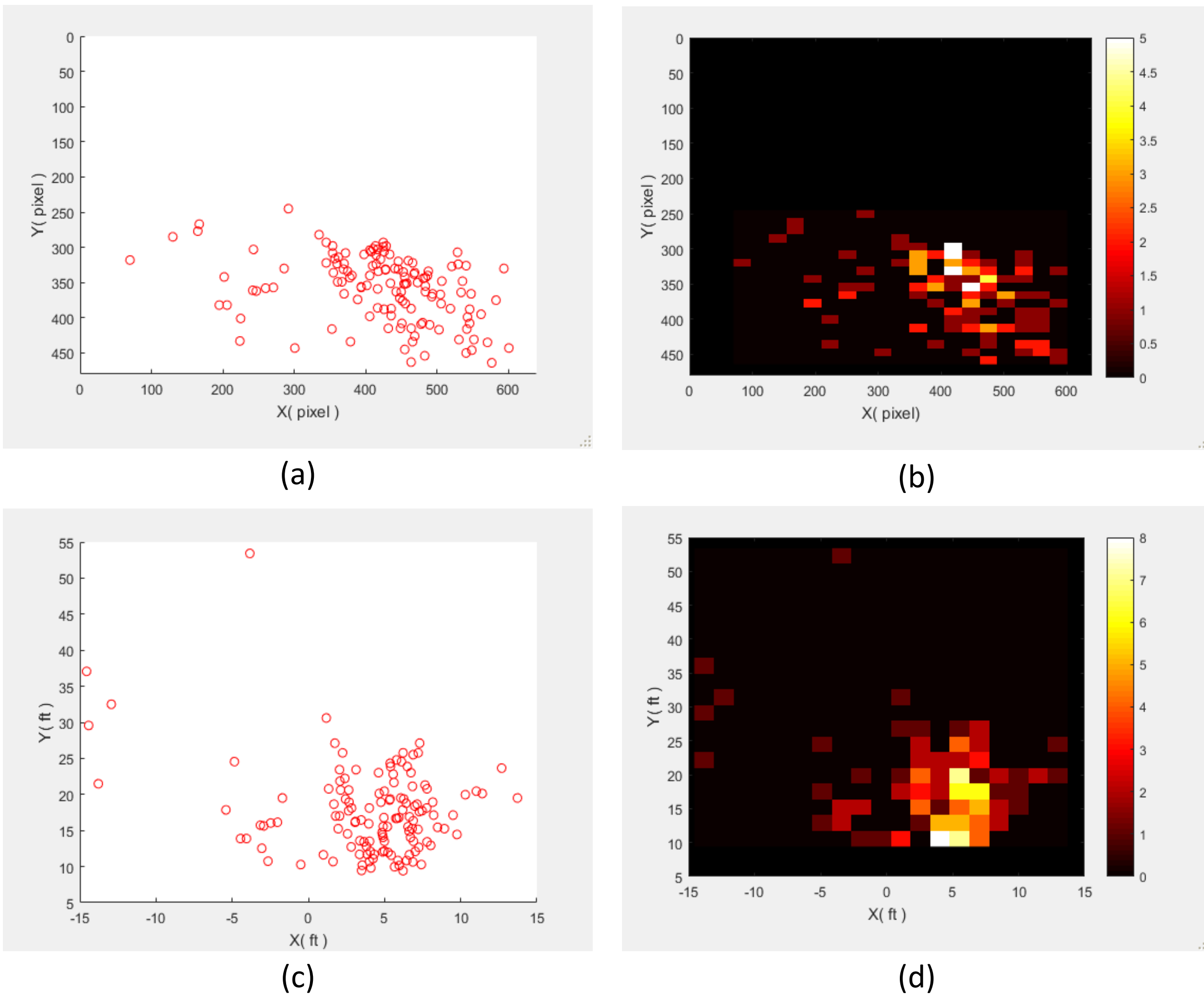


Figure 6.1 Scatter plots and heat maps showing the distributions of near-misses in image coordinates (a) and (b) and top-view of real-world coordinates (c) and (d).

7. Conclusions and Recommendations

A cost-effective framework for automated bus-to-pedestrian near-miss detection through an onboard monocular vision system was developed in this project. Its purpose is to automatically extract bus-to-pedestrian surrogate safety measure data using onboard monocular video. 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 developed camera model finds the correspondence between the image coordinates and real-world coordinates of the detected pedestrians. Using this correspondence, relative speed and position information is calculated, and then the near-miss indicators can be found. This framework is among the first efforts for detecting bus-to-pedestrian near-misses by using onboard monocular video. It is applicable to both safety surrogate data collection and collision avoidance tasks for most types of vehicles. Comparison with a Rosco/MobilEye Shield+ system that included four camera sensors showed that our system works reasonably well.

On the basis of the experimental results and analysis 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 a bus approaching a crowd of pedestrians, to further improve overall performance. Second, errors in motion estimation may occur because some interest points may not derive from the pedestrians but from 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.

8. References

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

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.

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.

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.

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?."

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.

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.

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.

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.

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.

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.

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.

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).

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.

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.

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.

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.

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.

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.

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

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.

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

Zegeer, Charles V. "Traffic Conflicts, Erratic Maneuvers, and Near-Miss Accidents: State-of-the-Art." (1977).