A Cost-effective Framework for Automated Vehicle-pedestrian Near-miss Detection through Onboard Monocular Vision

Ruimin Ke[1] Jerome Lutin[2] Jerry Spears[3] Yinhai Wang[1,*]

[1]University of Washington [2]New Jersey Transit [3]Washington State Transit Insurance Pool

{ker27, yinhai}@uw.edu jerome.lutin@verizon.net jerry@wstip.org

Abstract

Onboard monocular cameras have been widely deployed in both public transit and personal vehicles. Obtaining vehicle-pedestrian near-miss event data from onboard monocular vision systems may be cost-effective compared with onboard multiple-sensor systems or traffic surveillance videos. But extracting near-misses from onboard monocular vision is challenging and little work has been published. This paper fills the gap by developing a framework to automatically detect vehicle-pedestrian near-misses through onboard monocular vision. The proposed framework can estimate depth and real-world motion information through monocular vision with a moving video background. The experimental results based on processing over 30-hours video data demonstrate the ability of the system to capture near-misses by comparison with the events logged by the Rosco/MobilEye Shield+ system which includes four cameras working cooperatively. The detection overlap rate reaches over 90% with the thresholds properly set.

1. Introduction

According to a report published by National Highway Traffic Safety Association (NHTSA) in 2016 [1], the number of total motor vehicle fatalities in the U.S. keeps decreased from 43510, in 2005 to 32,675 in 2014. However, the annual number of pedestrian fatalities remained at about same level during the past decade. As a result, pedestrian fatalities as a percentage of total fatalities increased from 11% to 15%. More research is needed to enhance pedestrian safety.

Traditional traffic safety research normally relies on data about collisions, which are rare events when considered in the context of normal measures of travel [2]. 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 about pedestrian safety improvements.

Researchers and engineers are aware of the lack of pedestrian collision data and started looking for surrogate safety measures [2-9, 16-18]. Despite slightly different definitions in several studies, these surrogate events are commonly called near-misses. Basically, 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 [7]. 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, such as video records [2-3], records from in-vehicle sensors [4], or even output from a simulation model of a certain location [9]. Initially, safety surrogate measures were extracted manually, which was very inefficient and inaccurate [6-8]. Recently, automated near-miss detection methods have been proposed in several studies but few of them have used onboard monocular cameras [2-4, 9].

There are several advantages in using onboard monocular camera as the near-miss sensor: 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 moving background and moving foreground in the video, traditional background

segmentation methods would not work as well as for stationary roadway surveillance videos [10]; 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 [11]. 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 applied in vehicle-pedestrian collision avoidance and near-miss detection. However, these studies performed all calculations using two-dimensional image coordinates instead of real-world coordinates. Consequently, those algorithms are not able to calculate true 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 [12-13].

In this paper, we propose a cost-effective framework to automatically extract vehicle-pedestrian near-misses from onboard monocular cameras. This framework is composed of four main stages: 1) pedestrian detection, 2) motion estimation, 3) vehicle-pedestrian relative position and speed calculation, and 4) near-miss detection. Our study addresses several challenging issues in near-miss detection including the moving video background issue, depth estimation, and real-world motion information extraction only using monocular video. The experimental results show that the proposed system is comparable to a commercial system with multiple camera sensors in terms of accuracy. Further analysis such as near-miss distribution estimation can be conducted with the proposed system. Our literature review did not reveal any significant published work about vehicle-pedestrian near-miss detection and extraction using onboard monocular videos. The work described in this paper appears to be among the first efforts.

2. Methodology

2.1. Overview

The proposed detection framework has a different processing logic from previous vehicle-pedestrian conflict studies. First, our framework does not handle the complex background information in the moving onboard video, but tries to locate the pedestrian directly. Also, 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, which are 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. The processing pipeline is shown in Figure 1. In the first stage, the well-known HOG pedestrian detector is used to detect pedestrian within the camera vision [14]; in the second stage, interest points inside the detected rectangle region which basically represents the pedestrian is tracked with KLT method [15], 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 thresholds such as TTC need to be set to determine if there is a potential vehicle-pedestrian near-miss event.

2.2. Pedestrian Detection

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 [21]. 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 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.

2.3. 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 [2, 23] but also in aerial videos with moving background [11, 22, 24].

However, in onboard monocular videos, background motion is 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 directly handled. Basically, the average motion of the top 20 interest point with the least errors is used to represent 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 1), we have

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

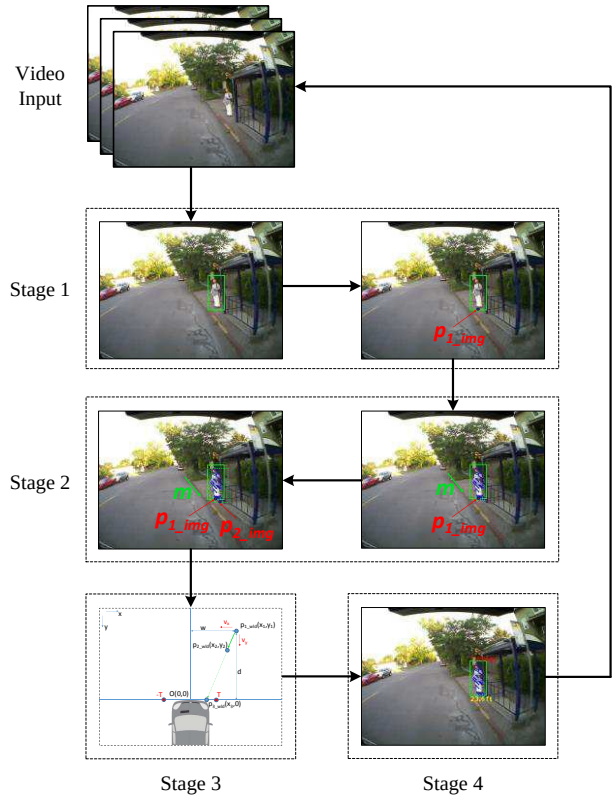

Figure 1. The proposed framework for vehicle-pedestrian near-miss detection through onboard monocular vision

2.4. Relative Position and Speed Calculation

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 last sub-section, p_{1_img} and p_{2_img} are the pedestrian locations in two frames (see Figure 2(a)). We calculate their corresponding points (see Figure 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 y_1 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 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

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

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

$$v_x = fr \times (x_2 - x_1) \quad (8)$$

$$v_y = fr \times (y_2 - y_1) \quad (9)$$

2.5. 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 [2-5] and we also use TTC as the major near-miss indicator in this study, which can be obtained with the following equation

$$TTC = \frac{y_1}{v_y} \quad (10)$$

where y_1 is the *y*-coordinate of the detected pedestrian in the real-world coordinate (see Figure 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 x-axis. We define this indicator as distance-to-safety (DTS), which can be calculated as follows

$$DTS = v_x \times \frac{y_1}{v_y} \quad (11)$$

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), an event is detected. T is shown in Figure 2(b) and it should be set not smaller than half of the vehicle width.

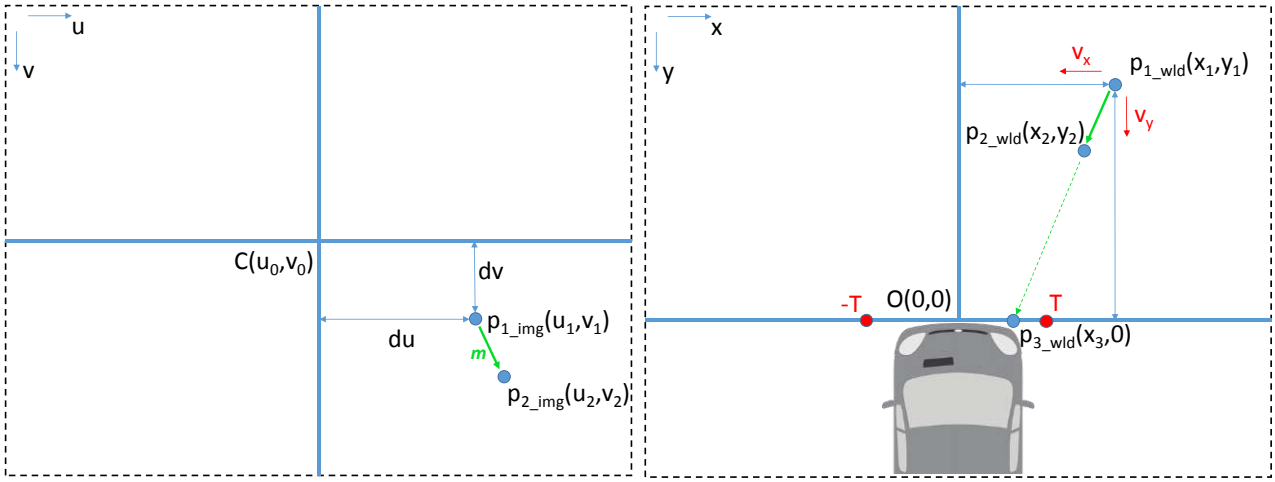

Figure 2. Method to find the correspondence between image coordinates and real-world coordinates.

3. Case Study

3.1. Data Description

The data was collected on a King County Metro transit bus, which was operated in Seattle area. The onboard monocular video data used as input to our system was part of the Rosco Dual-Vision system. The comparison dataset of vehicle-pedestrian conflict was collected by the Rosco/MobilEye Shield+ system. Rosco/MobilEye Shield+ system is a vision-based collision avoidance warning system specifically designed for large vehicles (e.g., buses, trucks) which includes four cameras working cooperatively. The video for testing our method, i.e., the

onboard front-facing video collected by Dual-Vision system, has a resolution of 640 × 480 pixels (width × height).and a frame rate of 7.5 frames-per-second (fps).

3.2. Results and Validation

More than 30 hours of onboard monocular video data was used to test the performance of the proposed near-miss detection method. Figure 3 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 by 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 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 at when the TTC threshold is set to 2s. The results show that our video system detects majority of near-misses picked up by the Shield+ system but 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 vehicle-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.

(a) (b)

Figure 3. Sample frames showing the representative near-miss events detected by the proposed system.

Table 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%

Besides safety surrogate data collection, another purpose for developing a cost-effective vehicle-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 4. It can be seen that most events occur at the right of the vehicle. This is reasonable since when a vehicle travels on 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 same size at a farther distance to the camera occupy less 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.

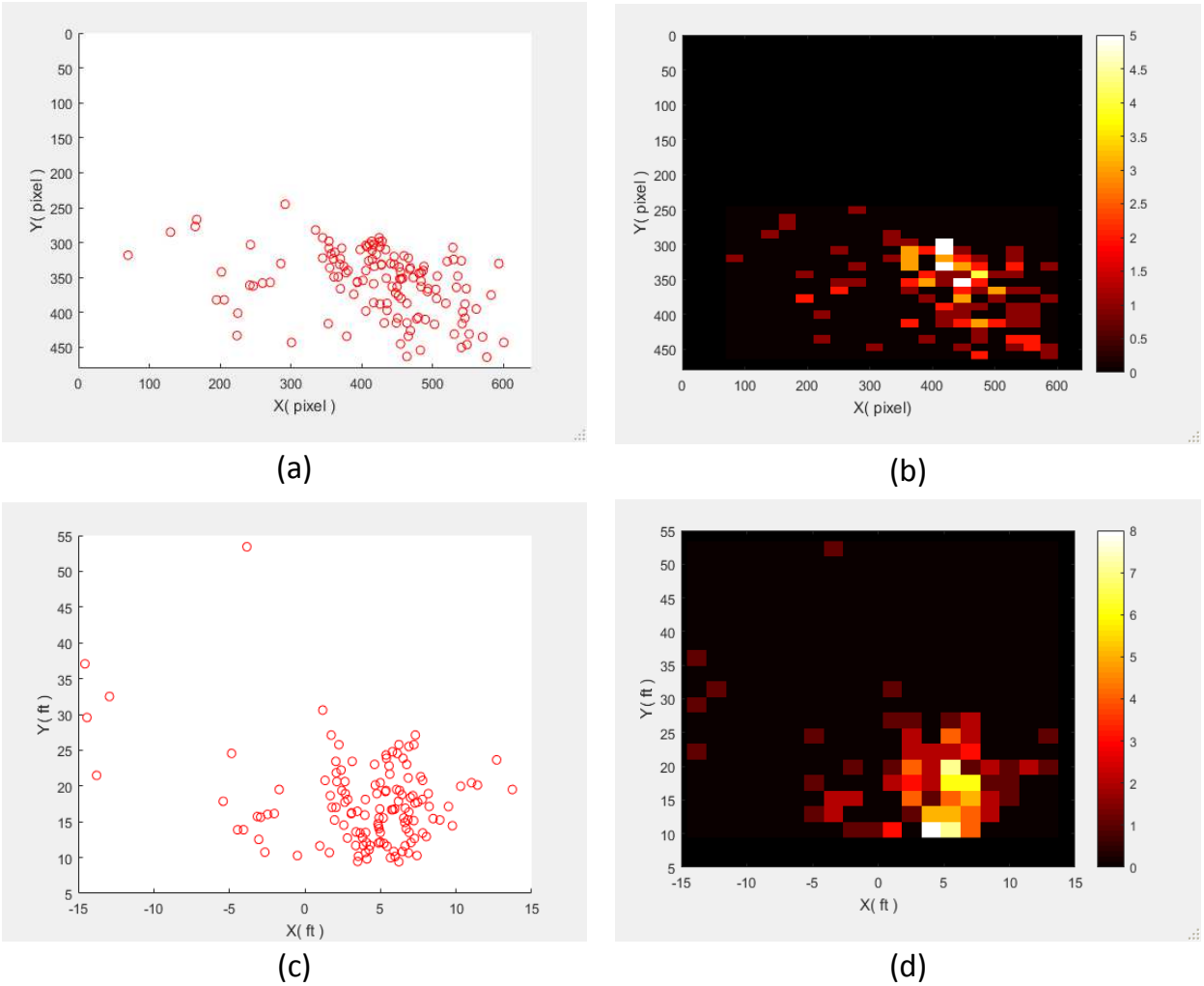

Figure 4. Scatter plots and heat maps showing the distribution of near-misses in image coordinates (a) and (b) and top-view of real-world coordinates, (c) and (d).

4. Conclusion and Future Work

A cost-effective framework for automated vehicle-pedestrian near-miss detection through onboard monocular vision is proposed in this paper. It aims at automatically extracting vehicle-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 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 are able to obtain the near-miss indicators. This framework is among the first efforts for detecting vehicle-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 analysis in this study, future work is currently planned on the following aspects. First, future work will involve testing the system in more challenging scenarios such as vehicle 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 the 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.

Acknowledgement

This study was supported in part by TRB Transit IDEA J-04/IDEA 82, by the Pacific Northwest Transportation Consortium, US Department of Transportation University Transportation Center for Federal Region 10, and by the National Natural Science Foundation of China (Grant No. 51329801).

References

[1] National Highway Traffic Safety Association. US Department of Transportation: Traffic Safety Facts, 2016.

[2] 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*, (2140): 44-54, 2009.

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

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

[5] M. Minderhoud, and H. Bovy. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention*, 33(1): 89-97, 2001.

[6] C. Chin, and T. Quek. Measurement of traffic conflicts. *Safety Science*, 26(3): 169-185, 1997.

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

[8] V. Zegeer, & C. Deen. Traffic conflicts as a diagnostic tool in highway safety, 1977.

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

[10] G. Zhang, R. Avery, and Y. Wang. 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): 138-147, 2007.

[11] R. Ke, Z. Li, S. Kim, J. Ash, Z. Cui, and Y. Wang. Real-time bi-directional traffic flow parameter estimation from aerial videos. *IEEE Transactions on Intelligent Transportation Systems*, 18(4): 890-901, 2017.

[12] K. Mori, T. Takahashi, I. Ide, H. Murase, T. Miyahara, and Y. Tamatsu. Recognition of foggy conditions by in-vehicle camera and millimeter wave radar. In *2007 IEEE Intelligent Vehicles Symposium*, pages 87-92, 2007.

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

[14] N. Dalal, and B. Triggs (2005, June). Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886-893, 2005.

[15] D. Lucas, and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI* , volume 81, pages 674-679, 1981.

[16] I. Kaparias, M. Bell, J. Greensted, S. Cheng, A. Miri, C. Taylor, and B. Mount. 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):75-82, 2010.

[17] 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*, (2198): 52-64, 2010.

[18] S. Malkhamah, M. Tight, and F. Montgomery. The development of an automatic method of safety monitoring at

Pelican crossings. *Accident Analysis & Prevention*, 37(5): 938-946, 2005.

[19] T. Wannige, and J. Sonnadara. Pedestrian Collision Detection through Monocular Vision. In *Proceedings of the Technical Sessions*, volume 26, pages 17-24, 2010.

[20] H. Kataoka, K. Tamura, Y. Aoki, Y. Matsui, K. Iwata, and Y. Satoh. 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*, pages 2472-2477, 2013.

[21] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4): 743-761, 2012.

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

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

[24] R. Ke, S. Kim, Z. Li, and Y. Wang, Motion-Vector Clustering for Traffic Speed Detection from UAV Video. In *Proceedings of the IEEE Conference on Smart Cities*, Guadalajara, Mexico, 2015.