

Single-Variable Scenario Analysis of Vehicle-Pedestrian Potential Crash Based on Video Analysis Results of Large-Scale Naturalistic Driving Data

Renran Tian¹(✉), Lingxi Li¹, Kai Yang¹, Feng Jiang¹, Yaobin Chen¹,
and Rini Sherony²

¹ Transportation Active Safety Institute, Department of Electrical and Computer
Engineering, Indiana University-Purdue University, Indianapolis, USA

{rtian, LL7, kaiyang, fejiang, ychen}@iupui.edu

² Collaborative Safety Research Center,

Toyota Motor Engineering and Manufacturing North America Inc.,
Erlanger, USA

rini.sherony@tema.toyota.com

Abstract. Vehicle-pedestrian crashes are big concerns in transportation safety, and it is important to study the vehicle-pedestrian crash scenarios in order to facilitate the development and evaluation of pedestrian crash mitigation systems. Many researchers have tried to investigate the pedestrian crash scenarios relying on crash databases or pedestrian behavior prediction models, both of which have some limitations like limited generalizability of the results, missing of important information, biased results. In this study, we propose to study the potential crash scenarios as one surrogate targets of the actual pedestrian crash scenarios. Extended from several previous studies, one single-variable scenario analysis is completed based on the video analysis results of one large-scale naturalistic driving data collection focusing on recording pedestrian behaviors in all kinds of situations. Through calculating potential conflict rates and applying chi-square tests for around 40 attributes from 12 scenario variables individually, this study has found out that number of pedestrians, pedestrian moving speed, pedestrian moving direction, vehicle moving direction, road type, road location, and existence of road separator/median are all important scenario variables for potential pedestrian-vehicle crashes.

Keywords: Naturalistic driving data collection · Pedestrian crash · Potential crash scenarios · Video analysis · Pedestrian behavior modeling

1 Introduction

Pedestrian safety is always a big concern in traffic safety, and pedestrian crashes are associated with a large number of injuries and fatalities every year. In the United States, the number for pedestrian fatalities has kept increasing in the past several years. In 2011, there are approximately 4,432 pedestrians killed which account for 14 percent of all traffic fatalities, and 69,000 pedestrian injured [1]. The fatality number has increased

7 % to 4,743 in the following year of 2012 [2]. In order to alleviate the situation, more and more vehicle safety systems for pedestrian detection and pedestrian crash mitigation have been developed and installed in the modern vehicles to improve the effectiveness of vehicular active safety functions [3]. As summarized in the literature [4], it is important to understand pedestrian behaviors as well as the most common vehicle-pedestrian crash scenarios for developing and evaluating these pedestrian crash mitigation systems, and typically the researchers have been following three routes in studying the scenario variables: (1) relying on crash database, (2) constructing pedestrian behavior prediction models, or (3) relying on the large-scale naturalistic driving data collection.

Towards using the crash database, researchers have tried to use real crash data to aid the development of pedestrian detection systems [7], and to generate test scenarios and key variables for pedestrian mitigation system evaluation [8, 9]. As pointed out in [4], although using crash data can guarantee the generated crash scenarios representing real crash situations in various road types, these scenarios usually miss some important details and contain biased information. Towards the second route listed above, many published studies are focusing on systematically predicting pedestrian behaviors under different traffic conditions and modeling general vehicle-pedestrian crash scenarios [10–14]. These mathematical models are clearly good solutions to provide convenient and detailed pedestrian behavior information. However, researchers have also criticized that most of the current pedestrian behavior prediction models are focusing on specified traffic scenarios with limited capability to generalize the prediction results to overall road environment [4, 14].

To overcome the limitations of crash databases and pedestrian behavior prediction models, some efforts are made towards studying pedestrian-vehicle crashes using naturalistic driving data in the Transportation Active Safety Institute (TASI), Indiana University – Purdue University Indianapolis. Based on one large-scale naturalistic driving data collection focusing on pedestrian behaviors, researchers have tried to study single pedestrian behavior variable of walking step frequency [5], estimate the pedestrian/vehicle encountering risks in various traffic scenarios [4], and provide preliminary results of pedestrian behavior analysis as well as pedestrian-vehicle crash scenarios [6]. Based on the literature [15], it is reasonable in majority of cases to use near-miss events or potential conflict events for crash scenario construction. Thus, the authors have briefly discussed the potential conflict definitions, potential conflict cases achieving process, and the vehicle-pedestrian crash scenarios based on the potential conflict rates in different situations [6]. However, the discussions in [6] have some limitations:

- As one preliminary study, the dataset used for analysis is neither complete nor accurate;
- Several important factors are missing during the data analysis;
- There is no statistical analysis performed to test the significance of each factor.

Thus, in this study, we will extend the results described in [6] with using the whole dataset of the large-scale naturalistic driving data collection after reducing errors, including more variables in the analysis, and apply statistical analyses to test each factor.

2 TASI 110-Car Naturalistic Driving Study

From 2012 to 2013, the Transportation Active Safety Institute at the Indiana University – Purdue University Indianapolis has completed one large-scale naturalistic driving data collection focusing on understanding pedestrians' behavior during pre-crash and crash scenarios. A total of 110 cars with subject drivers were recruited around the city of Indianapolis starting from spring, 2012. Since the recruitments, every driver's driving data for the following one year were continuously recorded. The recorded data mainly include three parts that are synchronized during the whole data collection period:

1. Forward-view video data: relying on one video camera installed on the windshield of every data collection vehicle facing forward, high-definition video data were continuously recorded when the vehicle is moving representing the driver's view during all the time.
2. Vehicle GPS coordinates: relying on the GPS receiver installed in each vehicle, the longitude and latitude coordinates of the vehicle were recorded all the time when the subject was driving.
3. Vehicle acceleration data: relying on the accelerometer (g-sensor) installed in each vehicle, the vehicle accelerations in x-y-z directions were continuously retrieved and recorded at the rate of 10 Hz when the subject was driving.

After the one-year of data collection, a total of approximately 90 TBs of driving video data for over 1.44 million driving miles accumulatively across all the 110 subjects were recorded, with corresponding vehicle GPS coordinates and vehicle acceleration data also synchronized and recorded in the database. Upon the construction of the database, several steps of data analyses have been completed. Because the main purpose of the research focuses on pedestrian behaviors, one automatic pedestrian detection algorithm has been firstly applied on the collected video data to locate every scene with pedestrian(s). After manually checking all the pedestrian scenes detected automatically by the image-processing based algorithm to remove errors, no-interest scenes, and duplicated cases, short video clips were created towards each confirmed pedestrian scene. Then manual video analysis has been applied to the video clips by a group of trained data reductionists using several programs to detect potential conflict cases, to study pedestrian and driver behaviors, and to assign different scenario variables.

For more detailed information about the data acquisition apparatus, the constructed database structure, and the video analysis results, please refers to the literatures [4–6]. For more detailed information about the image processing based automatic pedestrian detection, please also refer to the literatures [16, 17].

3 Methodology for Data Analysis

The main purpose of this study is to investigate the potential conflict scenarios between the vehicles and pedestrians in the naturalistic road environment. As described above, literatures have shown that the potential conflict cases can represent the situations of

the real crashes very well [15], and thus the potential conflict scenarios found in this research can also be used as the surrogate scenarios for real vehicle-pedestrian accidents to some extents.

The definition of potential conflict in this study follows the literature [4]: potential conflict case refers to the case that “real crash between the vehicle and pedestrian(s) will occur if neither the driver nor the pedestrian(s) changes the moving speed/moving direction, or the trajectories of the vehicle and the pedestrian are adjacent to each other during the time period which results in the movement responses from the driver and/or the pedestrian(s) to avoid the contact, although the responses may not be necessary.”

As part of the overall data analysis process, the above Fig. 1 shows the whole picture of the project, as well as the data analysis process of the current study, marked as single-variable scenario analysis in the yellow solid line. The parts with green dashed line in the chart show the previous work, and the parts with red dotted line show the future work. Starting from the collected naturalistic driving video data, multiple steps of data analyses have been performed to achieve the potential conflict and non-potential conflict cases with values for different scenario variables assigned to each of them via intensive video analysis. Two types of scenario analysis can be done towards these processed cases including:

- Single-variable analysis: each scenario variable is treated to be independent, and the effect of individual scenario variable on the potential conflict rates can be calculated.
- Multi-variable analysis: multiple scenario variables are studied together, and compressive potential conflict scenarios can be achieved based on the calculated potential conflict rates.

Both of these two types of scenario analyses rely on the calculation of potential conflict rates using the following formula (1). When applying the formula (1) for single-variable analysis, certain scenario refers to one particular value for one individual scenario variable; however when applying the formula for multi-variable analysis, certain scenario refers to the combined values for a group of scenario variables.

$$\text{Potential Conflict Rate} = \frac{\text{Number of Potential Conflict Cases Under Certain Scenario}}{\text{Total Number of Cases Under Certain Scenario}} \quad (1)$$

Due to the scope of this paper, only single-variable analysis will be discussed, and the multi-variable analysis as well as matching the potential crash scenarios into accident scenarios will be completed in the future works. The following Table 1 shows the 12 variables studied for the single-variable analysis in this work, with descriptions about the variable definitions and the corresponding attributes. These variables include mainly pedestrian-related variables, road-related variables, vehicle-related variables, and environment-related variables. For each of the variable, we will calculate the potential conflict rate for every attribute respectively, and then apply chi-square analysis to test the significances of the effects.

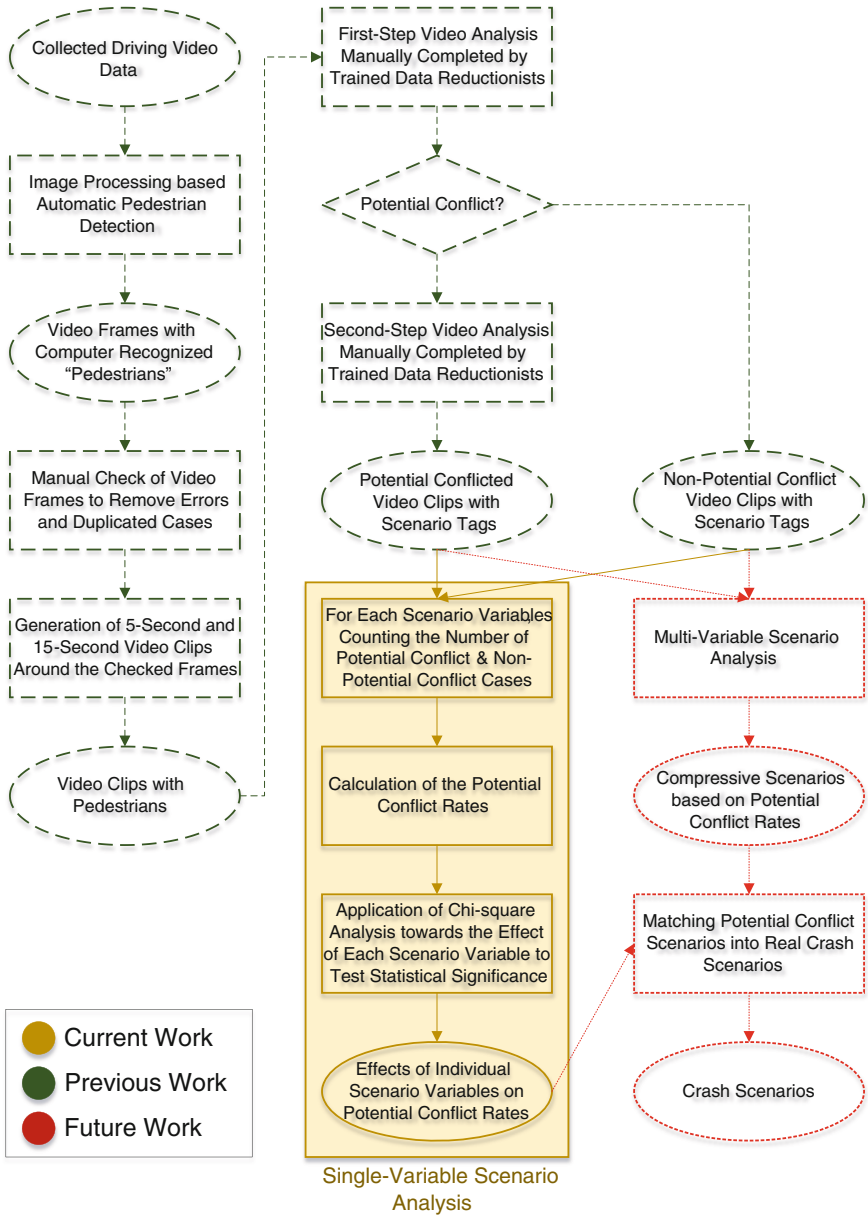


Fig. 1. Single-variable scenario analysis process as a part of the overall data analysis

Table 1. Scenario variables for the single-variable analysis

Variable name	Variable definition and attributes
Adult number	Definition: the number of adult pedestrians in the scene. Attributes: 0, 1, 2, >=3
Child number	Definition: the number of child pedestrians in the scene. Attributes: 0, 1, 2, >=3
Pedestrian walking speed	Definition: estimated pedestrian walking speed based on the video. Attributes: slow, normal, fast
Driving environment	Definition: brief categorization of the environment where the pedestrian-vehicle encounter happens. Attributes: rural, urban
Pedestrian moving direction	Definition: pedestrian moving trajectory during the pedestrian-vehicle encounter. Attributes: cross the road, with traffic (along the road), against traffic (along the road)
Pedestrian moving status	Definition: Movement status of the pedestrian before the event. Attributes: walking, standing, running, entering/exiting vehicle
Vehicle movement	Definition: moving behavior of the vehicle during the pedestrian-vehicle encounter Attributes: stop, straight, left turn, right turn, lane changing
Road alignment	Definition: the direction of the road that the vehicle is moving in. Attributes: straight, right curve, left curve
Road type	Definition: different places of the road where the encounter happens. Attributes: junction, intersection, mid-block cross walk
Road shoulder	Definition: shoulder is the part of a trafficway contiguous with the roadway for emergency use, for accommodation stopped vehicles, and for lateral support of the roadway structure. Attributes: with road shoulder, without road shoulder
Road median/separator	Definition: separator is the area of a trafficway between parallel roads separating travel in the same direction or separating a frontage road from other roads. Median is an area of a divided trafficway between parallel roads separating travel in opposite directions. Attributes: with road median/separator, without road median/separator
Driving condition	Definition: atmospheric conditions during the encounter. Attributes: clear, cloudy, rain, snow, backlight

4 Results

The above Fig. 2 shows the potential conflict rates calculated for all the attributes from the scenario variables, and the corresponding chi-square analysis results for each scenario variable individually. First of all, it looks like for both the number of adult pedestrians and the number of child pedestrians, their effects on potential conflict rate are significant. The charts show that with either more adult pedestrians or more child pedestrians, the chances to have potential conflict rates increase.



Fig. 2. Potential conflict rates and chi-square analysis results for different attributes from the 12 scenario variables individually.

Then for the pedestrian and vehicle behavior related variables, all of the pedestrian moving direction, pedestrian walking speed, pedestrian moving status, and vehicle movement are significant factors affecting the potential conflict rates. If looking into them separately, it looks like crossing the road is more dangerous to have potential conflict compared to walking along the road. Also when the pedestrian is running or

moving in faster speed will be associated with much higher chance of potential conflict situations, compared to slow speed walking or entering/exiting vehicles. For vehicle movements, making right turn and left turn will result in much higher potential conflict rates for the vehicle-pedestrian encounter, followed by changing lanes.

Finally for road and environment related variables, driving conditions and road shoulder existence are not proved to have significant effects on potential conflict rates. Although the chi-square analysis for road alignment has achieved one p -value less than 0.05, the effects are not as strong as other factors, and may not have practical meanings since all the potential conflict rates for the three road alignment attributes are relatively small. The other three strong factors include driving environment, road type, and existence of median/separator. The results have shown that in the rural environment, the vehicle-pedestrian encounter has doubled chances to be potential conflict case compared to the urban environment. Also among different road types, mid-block cross walk has a significantly higher chance to be associated with potential conflict cases compared to intersections and junctions, which means that the mid-block cross walk location is more dangerous. Another interesting finding is that when there is median or separator for the road, the potential conflict chance decreases compared to the road without such infrastructure. This may be caused by the situations that the pedestrian can wait safely at the middle of the road to reduce the chance to get into potential crash with the vehicles.

5 Conclusions and Future Work

In this study, one single-variable scenario analysis has been applied on the video analysis results of a series of potential conflict and non-potential conflict cases with scenario variable attributes assigned. These cases are retrieved from the TASI 110-car naturalistic driving data collection through the previous works. By assuming that all the interested scenario variables are independent from each other, the single-variable analysis studies the effect of each variable individually by calculating potential conflict rates and applying chi-square analysis. The results have shown that for vehicle-pedestrian encounter in the road, several factors may significantly increase the chance for the encounter to become a potential crash, including: (1) larger numbers of adult or child pedestrian, (2) faster pedestrian moving speed, (3) pedestrian crossing the road, (4) vehicle turning, (5) at mid-block cross walk location, (6) in the rural area, and (7) at the road without separator/median in the mid of the road.

As for future work, multi-variable analysis may be applied towards the same dataset used in this research to achieve more comprehensive scenarios. The multi-variable analysis will not require the assumption that all the scenario variables are independent from each other, and will consider the combinations of different variables/attributes together as one scenario for the analysis. Also, the current results are only regards to potential crash scenarios between the vehicles and pedestrians in the road. It will be beneficial to match the scenarios with real crash scenarios to achieve more applicable results.

Acknowledgment. The authors would like to thank Toyota's Collaborative Safety Research Center for funding this project.

References

1. National Highway Traffic Safety Administration: Traffic safety factors – pedestrians (2011 Data). NHTSA report DOT HS 811 748 (2013)
2. National Highway Traffic Safety Administration: Traffic safety factors – pedestrians (2012 Data). NHTSA report DOT HS 811 999 (2014)
3. Gandhi, T., Trivedi, M.M.: Pedestrian protection systems: issues, survey, and challenges. *IEEE Trans. Intell. Transp. Syst.* **8**(3), 413–430 (2007)
4. Tian, R., Li, L., Yang, K., Chien, S., Chen, Y., Sherony, R.: Estimation of the vehicle-pedestrian encounter/conflict risk on the road based on TASI 110-car naturalistic driving data collection. In: *IEEE Intelligent Vehicles Symposium*, 8–11 June 2014, Dearborn, MI, pp. 623–629 (2014)
5. Tian, R., Du, E.Y., Yang, K., Jiang, P., Jiang, F., Chen, Y., Sherony, R., Takahashi, H.: Pilot study on pedestrian step frequency in naturalistic driving environment. In: *IEEE Intelligent Vehicles Symposium (IV)*, 23–26 June 2013, Gold Coast, Australia (2013)
6. Du, E.Y., Yang, K., Jiang, F., Jiang, P., Tian, R., Luzetski, M., Chen, Y.: Pedestrian behavior analysis using 110-car naturalistic driving data in USA. In: *23rd International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, 27–30 May 2013, Seoul, Korea (2013)
7. Kong, C., Nei, J., Yang, J.: A study on pedestrian detection models based on real accident data from IVAC database in Changsha of China. In: *2012 Third International Conference on Digital Manufacturing & Automation*, July 2012, Guilin, China, pp. 136–139 (2012)
8. Good, D.H., Abrahams, R.L.: Vehicle speeds for pedestrian pre-crash system test scenarios based on US data. In: *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, 8–11 October 2014, Qingdao, China (2014)
9. Coelingh, E., Eidehall, A., Bengtsson, M.: Collision warning with full auto brake and pedestrian detection – a practical example of automatic emergency braking. In: *IEEE International Conference on Intelligent Transportation Systems*, September 2010, Madeira Island, Portugal, pp. 155–160 (2010)
10. Wakim, C.F., Capperon, S., Oksman, J.: Design of pedestrian detection systems for the prediction of car-to-pedestrian accidents. In: *IEEE International Conference on Intelligent Transportation Systems*, October 2004, Washington, D.C., USA, pp. 696–701 (2004)
11. Flach, A., David, K.: A physical analysis of an accident scenario between cars and pedestrians. In: *the IEEE 70th Vehicular Technology Conference*, pp. 1–5, Anchorage, AK (2009)
12. Zhuang, X., Wu, C.: Modeling pedestrian crossing paths at unmarked roadways. *IEEE Trans. Intell. Transp. Syst.* **14**(3), 1438–1448 (2013)
13. Chen, Z., Ngai, D.C.K., Yung, N.H.C.: Pedestrian behavior prediction based on motion patterns for vehicle-to-pedestrian collision avoidance. In: *IEEE International Conference on Intelligent Transportation Systems*, October 2008, Beijing, China, pp. 316–321 (2008)
14. Papadimitriou, E., Yannis, G., Golias, J.: A critical assessment of pedestrian behavior models. *Transp. Res. Part F Traffic Psychol. Behav.* **12**(3), 242–255 (2009)
15. National Highway Traffic Safety Administration: Evaluating the relationship between near-crashes and crashes: can near-crashes serve as a surrogate safety metric for crashes? NHTSA report DOT HS 811 382 (2010)

16. Yang, K., Du, E.Y., Delp, E.J., Jiang, P., Chen, Y., Sherony, R., Takahashi, H.: A new approach of visual clutter analysis for pedestrian detection. In: 16th International IEEE Conference on Intelligent Transportation Systems, 6–9 October 2013, The Hague, The Netherlands (2013)
17. Yang, K., Du, E.Y., Delp, E.J., Jiang, P., Jiang, F., Chen, Y., Sherony, R., Takahashi, H.: An Extreme Learning Machine-based Pedestrian Detection Method. In: IEEE Intelligent Vehicles Symposium (IV), 23–26 June 2013, Gold Coast, Australia (2013)