2018 3rd International Science of Smart City Operations and Platforms Engineering in Partnership with Global City Teams Challenge

Crowd motion detection and prediction for transportation efficiency in shared spaces

Dongfang Yang, John M. Maroli, Linhui Li, Menna El-Shaer, Bander A. Jabr, Keith Redmill, Füsun Özgüner, and Ümit Özgüner

***Abstract*—In the shared space scenario where pedestrian crowds and autonomous vehicles coexist, the transportation efficiency of the shared space can be improved by predicting the intention of the crowd and adjusting the driving strategy of the autonomous vehicles. This study proposes a framework that consists of the detection of individual pedestrians in a crowd via both on-vehicle and infrastructure sensors, the prediction of the crowd motion given the vehicle driving strategy, and the evaluation of the transportation efficiency in shared spaces. Methods for pedestrian detection and scenario prediction are introduced. Several aspects for improving transportation efficiency in shared spaces are discussed. Preliminary results of pedestrian detection on individual sensors and a simulation case study for estimating the desired time for an autonomous vehicle to pass the a shared space scenario demonstrate the potential of the proposed framework.**

***Index Terms*—smart city, intelligent transportation, autonomous vehicle, shared space, pedestrian, crowd, scenario simulation**

I. Introduction

THE intention of crowd pedestrians plays an important role for autonomous vehicles or intelligent systems in transportation. This intention is especially critical to shared space scenarios that involve crowd pedestrians and autonomous vehicles. One of the most common issues that relies on the crowd intention is how to improve the transportation efficiency in shared spaces when autonomous vehicles traverse the shared spaces that are partially or mostly occupied by crowd pedestrians. To do this, it is necessary to evaluate the status of individual pedestrians as well as their interactions with other pedestrians and the autonomous vehicles. This evaluation requires simultaneous handling of pedestrian motion, vehicle action, and the driving efficiency of the vehicle.

This study proposes a framework that aims to (a) detect individual pedestrian's state in the crowd via multiple sensors, (b) predict crowd pedestrians' motion given the driving strategy of the autonomous vehicles, and (c) evaluate the vehicle driving efficiency based on the scenario simulation, which eventually contributes to the transportation efficiency in shared spaces. A multi-sensor strategy was introduced for accurately detecting and estimating the individual pedestrian's state. Initial results of pedestrian detection on each separate sensor are presented. A social force based vehicle-crowd interaction model was combined with a vehicle model to predict and evaluate the scenario, and consequently adjust the driving efficiency of the autonomous vehicles. The approaches to address several types of driving efficiency problems are discussed. Specifically, a simulation case study was done for the prediction of the desired time for an autonomous vehicle to pass through crowds of different densities.

Pedestrian detection using on-vehicle sensors has been widely studied for a long time. Among these on-vehicle sensors, much progress has been achieved with monocular cameras [1] [2] [3]. LiDAR-based [4] and stereo-based [5] [6] pedestrian detection approaches usually work in conjunction with monocular cameras. On-vehicle sensors provide instant detection results of pedestrians in the neighborhood of the autonomous vehicle. However, when the crowd density is high enough, it is generally difficult for on-vehicle sensors to detect all individual pedestrians due to massive occlusions. Nowadays, with the commercialization and the decreasing prices of unmanned aerial vehicles (UAVs), it is possible to use UAVs with downward facing aerial cameras as infrastructure sensors hovering above the interested area so that the individual pedestrians can be more easily detected. Therefore, we propose a multi-sensor pedestrian detection strategy that relies on both UAV-based infrastructure sensors and the on-vehicle sensors to handle the massive occlusion problems.

Scenario prediction provides necessary information for adjusting the driving efficiency of autonomous vehicles traversing the crowd in shared spaces. The prediction requires analyzing the interactive behavior of both the crowd pedestrians and the autonomous vehicles. Several studies have inspected this interaction mechanism [7] [8] [9]. Due to the complexity of the interaction mechanism, scenario simulation [10] is an effective approach to address the driving efficiency problem. Assuming the correctness of pedestrian detection in the previous stage and the validity of the vehicle-crowd interaction mechanism, analyzing the simulation results gives useful information for improving the driving efficiency, on which the transportation efficiency in shared spaces can be improved.

There are two main contributions in the study. First, the combination of UAV-based infrastructure sensors and the on-vehicle sensors provides a prospective scheme to deal with the pedestrian detection task in crowded scenarios. Second, utilizing the interactive scenario simulation, the transportation efficiency problems in shared spaces can be effectively analyzed. The proposed framework will eventually contribute towards furthering intelligent transportation systems (ITS) and

Dongfang Yang, John M. Maroli, Menna El-Shaer, Bander A. Jabr, Keith Redmill, Füsun Özgüner, and Ümit Özgüner are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, 43210, USA

Linhui Li is with Dalian University of Technology, Dalian, Liaoning, China.

Manuscript received February 04, 2018.

978-1-5386-6746-0/18/$31.00 ©2018 IEEE
DOI 10.1109/SCOPE-GCTC.2018.00007

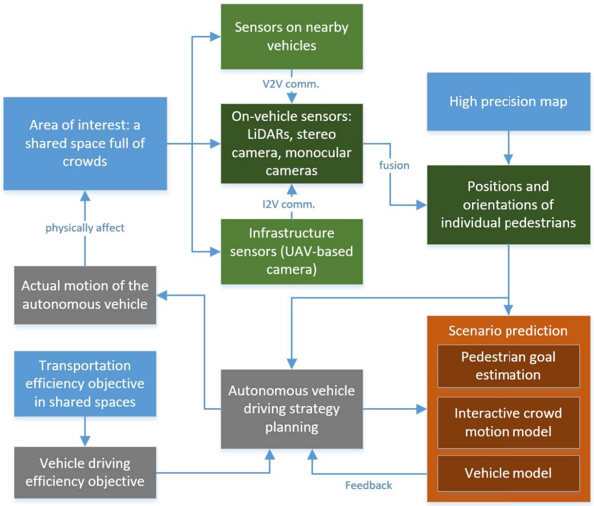

Fig. 1. The proposed framework to improve transportation efficiency.

smart city operations.

For the rest of the paper, section 2 presents the proposed overall framework. Section 3 describes the multi-sensor strategy and the methods for pedestrian detection on individual sensors with the corresponding initial results. Section 4 details the process and the applied models to achieve scenario prediction. Section 5 proposes the approach to evaluate and improve the transportation efficiency in shared spaces, with a simulation case study for the estimated time to pass through the crowd. Lastly, conclusions and future work are discussed.

II. Framework

Figure 1 shows the proposed overall framework. First, for an area of interest, individual pedestrians in the crowd are detected via both on-vehicle sensors and infrastructure sensors. Second, for the subject vehicle, the detection results are fused with the results from nearby vehicles (if they exist) and the results of the infrastructure sensors. The current pedestrian states, i.e., the positions and orientations of individual pedestrians at the current time step, are determined by combining the fused detection results with the high precision map. Next, the autonomous vehicle plans an initial driving strategy based on current pedestrian states. Both the driving strategy and the pedestrian states are sent to the scenario prediction module. The output of the scenario prediction is fed back to the autonomous vehicle so that the driving strategy can be updated based based on a specific driving efficiency objective, which is generated according to the transportation efficiency objective in shared spaces. Finally, the actual motion of the autonomous vehicle physically affects the area of interest.

Transportation efficiency objectives in shared spaces should be pre-specified as an input or a criteria for the framework. The transportation efficiency objective is then translated into the vehicle driving efficiency objective, because the autonomous vehicle is the primary and the most influential participant in the scenario.

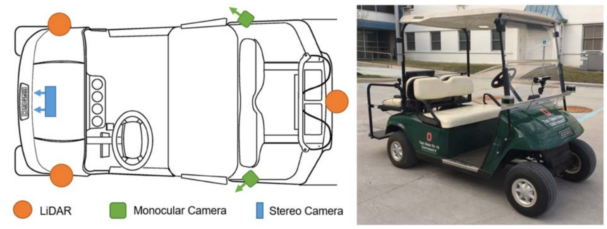

Fig. 2. Experimental vehicle (right) and the configuration of on-vehicle sensors (left).

III. Pedestrian Detection in Crowd

A. On-vehicle Sensors

An E-Z-GO® golf cart is used as our experimental vehicle. The vehicle contains a front facing stereo camera, two front/side facing monocular cameras, and three LiDAR (Light Detection And Ranging) sensors as shown in figure 2.

1) Monocular camera: Monocular camera vision comes from two FLEA®3 GigE Vision cameras. It also comes from either one of the channels of the stereo camera. Our study uses the approach in [11], which relies on extracting a Histogram of Oriented Gradient (HOG) features from the image, followed by a linear classifier using Support Vector Machines (SVMs). The overall detection process is illustrated in figure 3. Those methods usually work, specifically for pedestrians, as HOG features are robust against illumination and small local pose differences due to the fact that pixel gradients are normalized locally within blocks in the image. Adding texture information using Local Binary Pattern (LBP) descriptors to HOG features is a well-known method of detecting shapes and textures in the image feature space and could be applied to our system for better detection accuracy [12].

2) Stereo camera: A ZED™ 2K stereo camera is used primarily for detecting pedestrians in front of the autonomous vehicle. As shown in figure 4, the detection task is divided into two parallel processes: UV-disparity map based object detection, and the semantic segmentation of a monocular vision based on a convolution neural network (CNN). The U-disparity map can be used to detect the ground plane and find the upper and lower edges of the object while the V-disparity map can be used to find the left and right edges. Once the objects are identified, they are compared with the semantic segmentation result, which is achieved by ICNet [13]. Pedestrians can be identified by combining the semantic segmentation result and the objects found using the UV-disparity map.

3) LiDAR: Three Velodyne® VLP-16 LiDAR sensors are used in conjunction to form a single 3D point cloud. Each LiDAR unit has 16 vertical layers covering a $\pm 15°$ vertical field of view and a $360°$ horizontal field of view with a 100 meter range. The point cloud data is received from each LiDAR at 10 Hz and then translational and rotational offsets are applied before combining the point clouds to properly account for their different mounting locations. The translational offset is measured manually and the rotational offset is measured by the extrinsic rotational calibration method presented in [14].

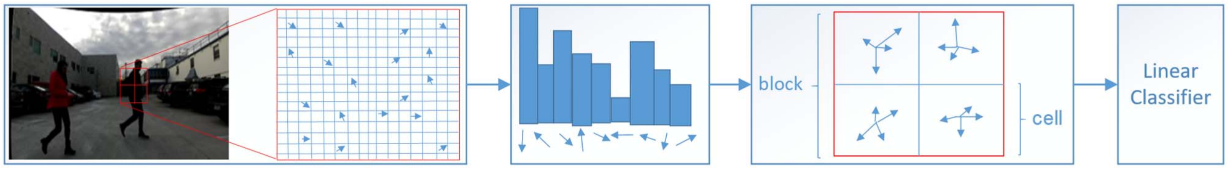

Fig. 3. The pedestrian detection process of a monocular camera. Step 1: calculate pixel gradients, i.e., magnitude and direction in 8x8 cells. Step 2: each pixel votes for its cell gradient orientation depending on its gradient magnitude. Step 3: concatenate cell histograms into blocks of 2x2 that describe the final HOG feature vector of the whole image. Step 4: classify resulting HOG feature vector into pedestrians and other.

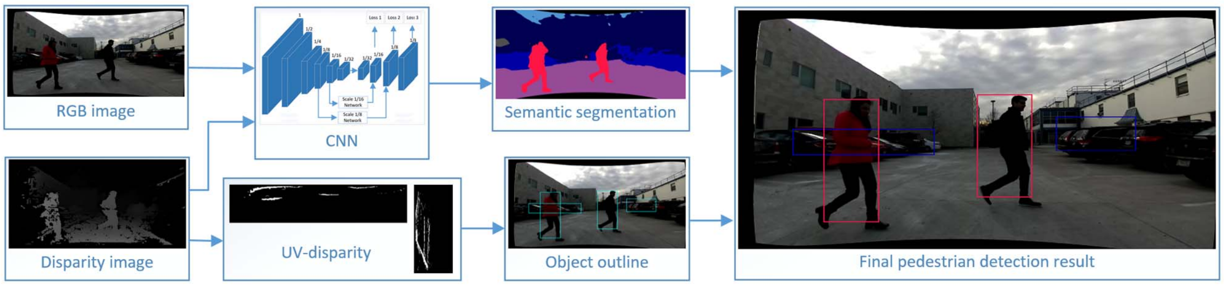

Fig. 4. Stereo vision based pedestrian detection. The top row shows the process of semantic segmentation while the bottom row shows the process of UV-disparity map based object detection.

The pedestrian detection method is similar to that used in [4] in regards to the LiDAR data being used in conjunction with monocular camera data. The overall process is illustrated in figure 5. The ground plane is first removed from the combined point cloud using the ground plane extraction algorithm from [15]. After the ground plane removal, object segmentation is performed as also done in [15]. The objects found in the point cloud after segmentation guide the camera-based pedestrian detection by providing regions of interest and narrowing the search space.

B. Infrastructure Sensors

Infrastructure sensors could be any combination of dedicated cameras mounted on nearby buildings, regular surveillance cameras, and downward facing aerial cameras mounted on UAVs. This section only focuses on UAV-based infrastructure sensors, as the detection methods apply similarly to others. In this study, a DJI® Phantom 3 SE with a built-in camera, as shown in figure 6, is used as the UAV-based infrastructure sensor.

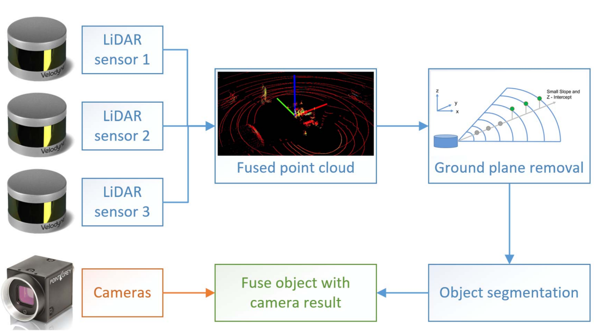

Fig. 5. The flowchart of LiDAR based pedestrian detection.

Since the UAV is part of the infrastructure, it is safe to assume that the background of the area of interest is known. Individual pedestrian detection is done based on each new frame calibrated and subtracted with respect to the background. For each background-removed frame, a series of image processing operations (thresholding, opening, and closing) are applied and the contours and bounding boxes of all objects are then generated. Using the contours, positions of individual pedestrians can be easily determined. The above detection process is illustrated in figure 7.

C. Sensor fusion

Once we have performed the pedestrian detection for both on-vehicle sensors and UAV-based infrastructure sensors, the next step is to fuse the detection results. The purpose of sensor fusion is to exploit the complementary and redundant characteristics of the sensors for increasing the reliability and accuracy of the pedestrian detection. The Dempster-Shafer theory (DST) is applied for the sensor fusion task, which combines the sources of evidence while avoiding counter-intuitive results [16]. Figure 8 shows how multiple sources

Fig. 6. The DJI® Phantom 3 SE unmanned aerial vehicle (UAV).

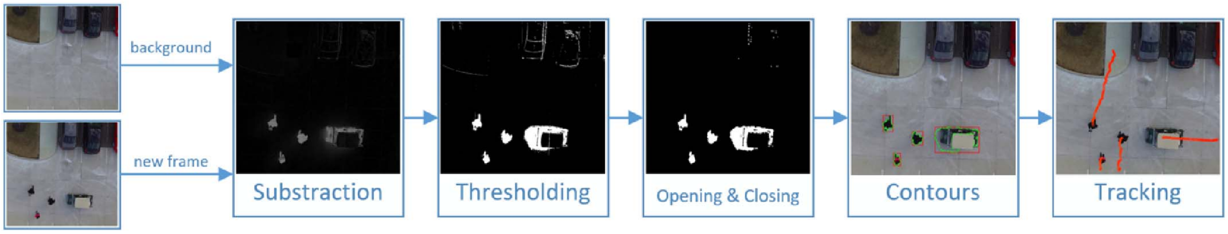

Fig. 7. Pedestrian detection of UAV-based infrastructure camera. In the scenario, 3 pedestrians are walking in shared space while the vehicle tries to traverse.

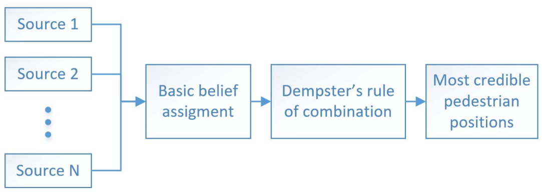

Fig. 8. The flowchart of the Dempster-Shafer theory (DST).

of evidence are processed by DST. First, the basic belief assignment (BBA) is done based on multiple sources. Then the most credible pedestrian positions are generated by applying Dempster's rule of combination.

IV. The Scenario Prediction

The scenario in a short time horizon, for example 10 to 30 seconds, is predicted in simulation given the current states of the individual pedestrians, models that describe the motion of both the pedestrians and the vehicle, and the driving efficiency objective of the autonomous vehicle.

A. Pedestrian Goals

Once the states of all individual pedestrians are available, the next step is to estimate the current pedestrian goals. It is common to assume that a pedestrian walks linearly at a constant speed for a short horizon. Under this linear assumption, a near-term goal of the pedestrian can be inferred. If a more precise estimation is desired, a nonlinear assumption could be applied. For example, a nonlinear Bayesian estimation filter with the hospitality map and the synthetic inclination map could be a solution [17] if provided with the terrain information.

B. Pedestrian Motion

The social force based vehicle-crowd interaction model [9] is applied to describe the motion of pedestrians under the influence of vehicles in shared spaces. Generally, the motion of a pedestrian in the crowd is governed by Newtonian dynamics and subject to the interactive forces from surrounding pedestrians and vehicles. The total force applied in the Newtonian dynamics can be expressed as

$$F_i = \sum_{j \in \mathbb{Q}(i)} (f_r^{ij} + f_c^{ij} + f_n^{ij}) + f_g^i + f_v^i + f_b^i + \epsilon_i. \qquad (1)$$

where $\sum_{j \in \mathbb{Q}(i)} (f_r^{ij} + f_c^{ij} + f_n^{ij})$ stands for the repulsive, collision, and navigational forces from nearby pedestrians, and f_g^i, f_v^i, f_b^i, and ϵ_i are the goal driven force, the interactive forces from vehicles, the repulsive forces of all boundaries, and a random noise, respectively. All forces are modeled mathematically to represent reasonable individual effect on the subject pedestrian.

C. Vehicle Motion

A kinematic bicycle model [18] is utilized to describe the vehicle motion due to the effectiveness of representing the slowly moving vehicle and the availability of adjusting vehicle driving strategy by changing its steering and gas/brake control action. This model assumes planner motion, the same left and right wheel steering angle, and no slip on all tires, which is compatible with the low speed shared space scenario. The model is expressed as

$$\dot{x} = v \cdot \cos(\theta + \beta) \qquad (2)$$
$$\dot{y} = v \cdot \sin(\theta + \beta) \qquad (3)$$
$$\dot{v} = f(u) \qquad (4)$$
$$\dot{\theta} = \frac{v}{l_r} \cdot \sin\beta \qquad (5)$$
$$\beta = \arctan\left(\frac{l_r}{l_f + l_r} \tan\delta_f\right) \qquad (6)$$

where v is the longitudinal speed, β is the velocity angle with respect to the vehicle center of gravity (C.G.), l_f, l_r are the distances from C.G. to the front wheel and the rear wheel respectively, u is the longitudinal control action (brake/gas), and δ_f is the lateral control action (steering angle of the front wheel).

V. Transportation Efficiency in Shared Spaces

As the pedestrian motion can only be predicted but hardly controlled by intelligent systems, the driving efficiency of the autonomous vehicle plays the main role in affecting the transportation efficiency in shared spaces. Depending on the driving efficiency objective, the driving strategy can be determined by either online or offline approaches.

A. Approaches

Figure 9 shows the process of online driving strategy planning. Once the driving efficiency objective has been chosen and the current pedestrian states are available, modules of

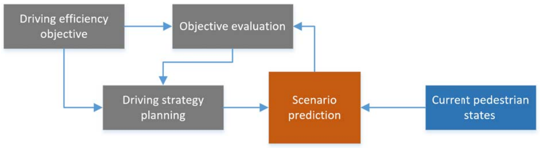

Fig. 9. Online driving strategy planning process.

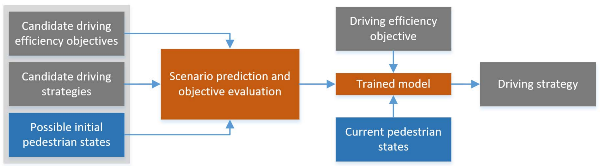

Fig. 10. Offline driving strategy determination process.

scenario prediction, objective evaluation, and driving strategy planning are executed sequentially and iteratively until a driving strategy that satisfies the driving efficiency objective is generated. This can be achieved by forming a model predictive control (MPC) problem and solving the objective function. Specific vehicle driving objectives could be, for example, finding the shortest passing time through the crowd while guaranteeing the pedestrian safety.

The driving strategy can also be determined offline as shown in figure 10. This approach requires training of a model that represents the relationship between the inputs (the driving efficiency objective and the current pedestrian states), and the output (the driving strategy). Training is achieved by the scenario prediction and the objective evaluation module given the data of possible initial pedestrian states and different candidates of driving efficiency objectives and driving strategies. The following simulation case study shows how this approach works to estimate the desired time for the autonomous vehicle to pass an area of interest in a crowded shared space.

B. *Offline Approach Case Study: Desired Time to Pass*

A crowded shared space scenario in front of the Thompson Library at The Ohio State University (OSU) was constructed in simulation, as shown in figure 11, to demonstrate how to estimate the desired time for the autonomous vehicle to pass this area. In this scenario, pedestrians are constantly entering into or exiting from the front door of the library while an autonomous vehicle is trying to pass through the shared space. The pedestrian density is defined as the number of pedestrians inside the area of interest at the time the autonomous vehicle initially enters this area. There is a random pattern of how pedestrians enter and exit the front door, so that different values of the pedestrian density can be generated. The vehicle driving strategy is fixed at the current stage to estimate the relationship between the pedestrian density and the desired passing time. Different driving strategies can be incorporated in the training model in the future.

For this simulation, the vehicle driving strategy generates the desired reference speed in terms of the distance from the closest pedestrian in the front ($FOV = 60°$) to the vehicle

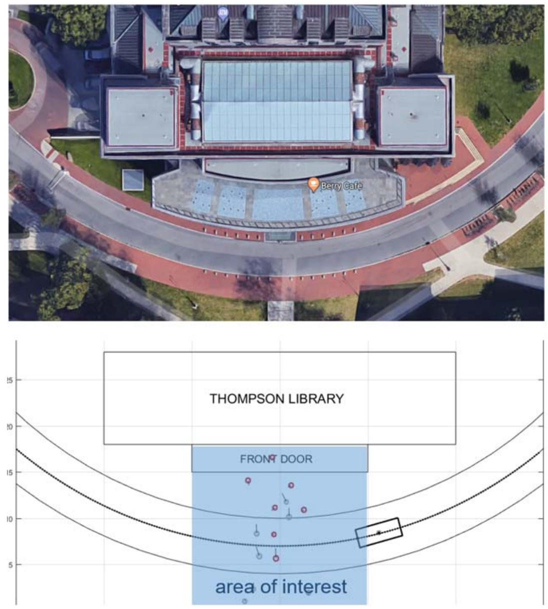

Fig. 11. Shared space layout in front of the Thompson Library at OSU (top) and the corresponding simulation environment (bottom).

center. The relationship between the distance and the desired reference speed is shown in figure 12, which guarantees the pedestrian safety due to the interactive behavior of pedestrians when facing the vehicle, as described in the vehicle-crowd interaction model in the previous section. A proportional controller is applied to regulate the longitudinal speed of the autonomous vehicle with the acceleration control gain $K_a = 0.5$ and the braking control gain $K_b = 5$.

The simulation was run 40 times. The initial pedestrian density and the total simulated time for the vehicle to pass this area were recorded. Based on the recorded data, a simple linear regression was conducted and the fitting results are shown in figure 13. The fitted line equation is

$$f(x) = \hat{p}_1 \cdot x + \hat{p}_2 = 0.3156 \cdot x + 3.095. \qquad (7)$$

The 95% confidence bounds for the two parameters are: $\hat{p}_1 \in (0.1369, 0.4944)$, $\hat{p}_2 \in (0.6222, 5.568)$.

Using the regression result, the desired passing time can be predicted when an autonomous vehicle with the same driving

Fig. 12. The action strategy of the autonomous vehicle for the time-to-pass case. In our simulation, $d_1 = 6$, $d_2 = 10$, $v_{normal} = 10mph$, and $v_{min} = 3mph$.

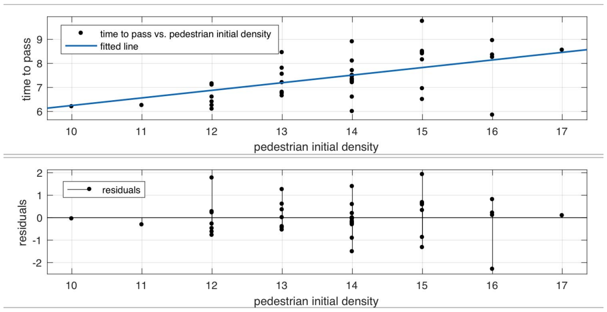

Fig. 13. Fitting results and the residuals for the Thompson Library scenario.

strategy wants to pass through the shared space scenario provided with the detected pedestrian density.

VI. Conclusion

This study constructed a framework that combines pedestrian detection via multiple sensors, vehicle-crowd interactive scenario prediction, and approaches to improve the driving efficiency of autonomous vehicles, which finally affects the transportation efficiency in shared spaces. Methods of pedestrian detection on different types of sensors were introduced and the corresponding initial results were presented. A simulation case study was conducted to demonstrate one of the proposed approaches for improving driving efficiency. The proposed framework has the potential to solve transportation problems in shared spaces where crowd pedestrians and autonomous vehicles interact with each other.

Future works, possible improvements, and challenges include the following:

- Pedestrian Detection: The fusion of detection results from different sensors should be further improved. As a fundamental structure for sensor fusion, Dempster-Shafer theory should be adapted to fit the specific situation. The communication between UAV-based infrastructure sensors and on-vehicle sensors also requires further exploration, especially for how to guarantee real-time information exchange.
- Pedestrian Goal Estimation: Although linear assumption is generally acceptable in practice, applying a high-fidelity estimation model or incorporating more environment information can improve the estimation result. However, high-fidelity models and additional information require high computational capability. It is necessary to find an approach that can balance the estimation performance and the computational load.
- Improving Driving Efficiency: More case studies are desired, especially for online driving strategy planning. In the simulation, replacing some variables, for example the initial pedestrian density, with realistic data can improve the reliability of the simulation. For the offline approach, methods other than linear regression are desired to better describe the relationship. It is also necessary to evaluate the computational complexity of all the approaches to address the driving efficiency problems.

Acknowledgment

Material reported here was partially supported by the NSF under the CPS Program (Award 1528489 and 1446735) and partially by the United States Department of Transportation under Award Number 69A3551747111 for the Mobility21 University Transportation Center.

Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the United States Department of Transportation or Carnegie Mellon University.

References

[1] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 12, pp. 2179–2195, 2009.

[2] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 7, pp. 1239–1258, 2010.

[3] 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*, vol. 34, no. 4, pp. 743–761, 2012.

[4] D. Matti, H. K. Ekenel, and J. P. Thiran, "Combining lidar space clustering and convolutional neural networks for pedestrian detection," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, Aug 2017.

[5] L. Zhao and C. E. Thorpe, "Stereo-and neural network-based pedestrian detection," *IEEE Transactions on intelligent transportation systems*, vol. 1, no. 3, pp. 148–154, 2000.

[6] S. Nedevschi, S. Bota, and C. Tomiuc, "Stereo-based pedestrian detection for collision-avoidance applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 380–391, 2009.

[7] B. Anvari, M. G. Bell, A. Sivakumar, and W. Y. Ochieng, "Modelling shared space users via rule-based social force model," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 83–103, 2015.

[8] W. Zeng, P. Chen, G. Yu, and Y. Wang, "Specification and calibration of a microscopic model for pedestrian dynamic simulation at signalized intersections: A hybrid approach," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 37–70, 2017.

[9] D. Yang, Ü. Özgüner, and K. Redmill, "Social force based microscopic modeling of vehicle-crowd interaction," in *Intelligent Vehicles Symposium (IV), 2018 IEEE*, IEEE, under review.

[10] D. Yang, A. Kurt, K. Redmill, and Ü. Özgüner, "Agent-based microscopic pedestrian interaction with intelligent vehicles in shared space," in *Proceedings of the 2nd International Workshop on Science of Smart City Operations and Platforms Engineering*, pp. 69–74, ACM, 2017.

[11] N. Dalal and B. Triggs, "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, 2005.

[12] S. Tang, M. Andriluka, and B. Schiele, "Detection and tracking of occluded people," *International Journal of Computer Vision*, vol. 110, no. 1, pp. 58–69, 2014.

[13] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "Icnet for real-time semantic segmentation on high-resolution images," *arXiv preprint arXiv:1704.08545*, 2017.

[14] J. M. Maroli, Ü. Özgüner, K. Redmill, and A. Kurt, "Automated rotational calibration of multiple 3d lidar units for intelligent vehicles," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, Oct 2017.

[15] M. Himmelsbach, F. v. Hundelshausen, and H. J. Wuensche, "Fast segmentation of 3d point clouds for ground vehicles," in *2010 IEEE Intelligent Vehicles Symposium*, pp. 560–565, June 2010.

[16] G. Shafer, *A mathematical theory of evidence*, vol. 42. Princeton university press, 1976.

[17] Z. M. Kassas, Ü. Özgüner, and J. Layne, "Out-of-surveillance target state estimation: A combined hospitability and synthetic inclination approach," in *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, vol. 1, pp. 710–715, IEEE, 2004.

[18] Ü. Özgüner, T. Acarman, and K. A. Redmill, *Autonomous ground vehicles*. Artech House, 2011.