Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Real Time Obstacle Estimation Based on Dense Stereo Vision for Robotic Lawn Mowers*

Jie Li[1], Huan Liu[3], Zhenglong Sun[1,2] and Rui Huang[1,2]

Abstract— **Information on obstacles (e.g., distances, scales, and categories) in the unstructured outdoor environment is crucial for a variety of robotic navigation tasks, such as obstacle avoidance, path planning, and security mechanism for pedestrians. However, most of the previous obstacle estimation methods merely focus on the detection of obstacles. In this paper, we propose a novel automatic obstacle estimation system that not only estimates distance and scale information but also can distinguish pedestrians from other barriers. The designed system comprises two branches, namely obstacle estimation and pedestrian detection. Obstacle estimation infers scales and depth clues from the disparity map using stereo vision, while pedestrian detection detects the pedestrian category by machine learning algorithms. We conduct the experiments on the mowing robot platform, and results show that our proposed system is highly effective. The furthest measurable distance is over 10 meters, while the maximum distance error rate is less than 5%. In addition, the average accuracies of the obstacle estimation and pedestrian detection are 94% and 97.6%, respectively.**

I. INTRODUCTION

There are numerous lawns that need regular maintenance, such as park lawns, football field lawns, and golf course lawns. If the mowing work must be completed by the worker assisted mowers, it will consume a lot of workforce and financial resources. Therefore, the automation of mowers becomes particularly important [1], [2]. Besides, the real-time obstacle information will help robotic lawn mowers to avoid obstacles and plan the path, thereby improving the overall automation and intelligence level of the robots.

Traditionally, sensor technologies such as ultrasonic sensors, infrared light, lasers, D-GPS, and sonars have been used to estimate obstacles [3], [4], [5], [6]. However, these sensors have some disadvantages, such as the long measurement period of the ultrasonic, the high price of the lasers, and the high sensitivity to the illumination of the infrared sensors. Because of the significant cost-effectiveness, cameras play an increasingly important role in the obstacle detection area. Cameras can have a much more full field of view, and they can gain information at higher resolutions and obtain more semantic information. In the past forty years, computer vision technologies have seen significant breakthroughs in robotic vehicle applications. We also believe that using computer vision to get environmental information is the primary trend in the future. In this paper, we will estimate the information on obstacles based on a stereo vision system. The reason we choose stereo vision is that both the monocular camera and the RGBD camera have obvious defects. For example, the ranging of the RGBD camera is relatively short, while it is difficult to measure the real distance of the obstacles using the monocular camera.

*This paper is partially supported by Shenzhen Institute of Artificial Intelligence and Robotics for Society, and Robotics Discipline Development Fund from Shenzhen Municipal Government (Grant No. 2016-1418).

[1]Jie Li, Zhenglong Sun and Rui Huang are with the School of Science and Engineering, The Chinese university of Hong Kong, Shenzhen, Guangdong, 518172, P.R. China jieli1@link.cuhk.edu.cn, sunzhenglong@cuhk.edu.cn, the correspondence author is Rui Huang, ruihuang@cuhk.edu.cn.

[2]Zhenglong Sun and Rui Huang are also with Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen, Guangdong, 518172, P.R. China.

[3]Huan Liu is with the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, Guangdong, 518055, P.R. China, hhandliu@gmail.com.

II. RELATED WORK

For the outdoor obstacle estimation, there are many different scenes, such as parks, apple orchards, and lawns [7]. It contains all kinds of obstacles, and the ground is probably potholed. In these scenarios, robots need to consider vehicles and pedestrians as well as animals, trees, and trash cans; therefore, an automatic obstacle estimation system is necessary. Based on the above situations, many researchers apply vision-based algorithms to detect obstacles for robotic lawn mowers, such as [8], [9]. But these two methods only consider the static objects, neither dynamic obstacles are considered, nor the distance and size of obstacles are calculated. Other researchers detect obstacles by using the UV disparity map [10]. However, this kind of approaches may lose the information of the original disparity map, and it takes a long time to calculate the UV disparity map. To simplify the processing of disparity maps, in this paper, we directly take advantage of the dense disparity map to identify obstacles. Besides, the proposed system can detect both the static and dynamic obstacles, and it can also calculate the distances and sizes of the obstacles.

For pedestrian detection, there are both traditional machine learning methods and emerging deep learning methods. Histogram of Oriented Gradient (HOG) was first proposed by Dalal in 2005 [11], and the SVM classifiers or the cascade classifiers [12] were trained using the descriptors. This method is still one of the most effective methods in the field of pedestrian detection, even though many variants have been developed in the following years. Since 2012, The Convolutional Neural Network (CNN) has been designed for object detection, such as Faster R-CNN [13], [14], [15] and SSD (Single Shot multibox Detector) [16]. Because SSD outperforms other methods in the processing time for mowing robot applications, the pedestrian detection network in this paper will be implemented based on the SSD network.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

The main contributions of this paper can be summarized as follows:

- We have proposed a novel autonomous obstacle estimation system that estimates distance and scale information of obstacles and can also distinguish pedestrians from other obstacles.
- We have improved the accuracy of the real-time pedestrian detection by using the enhanced SSD-based pedestrian detection algorithm.

III. SYSTEM OVERVIEW

An overview of the whole obstacle estimation system is shown in Fig. 1. The obstacle estimation system based on the binocular stereo vision includes two branches, namely the obstacle estimation and the pedestrian detection. First, the pipeline starts at the image pre-processing stage, which provides camera initialization, image acquisition, and pre-processing. The second stage includes two parallel modules. In the obstacle estimation module, the system will detect obstacles and calculate the distance as well as the size information. More details are described in section IV. Section V introduces different learning algorithms for pedestrian detection module. After that, we further fuse obstacle estimation and pedestrian detection information to help mowing robots better understand the environment.

IV. STEREO VISION BASED OBSTACLE ESTIMATION

In this part, we will estimate the obstacle information based on the disparity map. There are various obstacles on the path of the mowing robot, including the static obstacles (such as trash cans, cartons) and dynamic obstacles (for instance, pedestrians). The obstacle estimation branch can estimate the depths and sizes of these obstacles. Our method includes the following four components.

A. Stereo Rectifying and Matching

Firstly, we can rectify the left and the right images to compute the dense disparity map. This operation aligns these two images so that the 2D search can be reduced to a 1D search, which can improve the calculation speed of the algorithm. When we get the rectified images, we can obtain the disparity map through stereo matching. The rectified images are shown in Fig. 2.

B. Obstacle Detection

To improve the obstacle detection efficiency, we first remove the ground and other unrelated information in the disparity map. As shown in Fig. 3, h is the height of $\boldsymbol{P}$ from the ground, while $y_{\boldsymbol{P}}$ is the Y coordinate value of point $\boldsymbol{P}$, and h_c is the camera height, which is a constant.

According to Fig. 3, we can obtain the height h of a point $\boldsymbol{P}$:

$$h = h_c - y_{\boldsymbol{P}} \tag{3}$$

In order to reduce the computational complexity of the algorithm, we will remove as much useless information as possible. In this practical application, the obstacles above a certain height can be considered harmless to the robot so that we can remove the associated pixels in the disparity map. For example, in this paper, when the h is less than 0.01 or higher than 3 meters, we will discard the corresponding pixels in the disparity map. Because these kinds of pixels belong to the ground area or have no effect on the robotic lawn mower. Fig. 4 is the effect of the interference removal.

Then we carry out the morphological operation and it will output the convex hulls of all obstacles. At the same time, we also need to calculate the areas of the convex hulls. When the area is smaller than a specific threshold value, we will abandon it. Finally, the system will output the convex hull

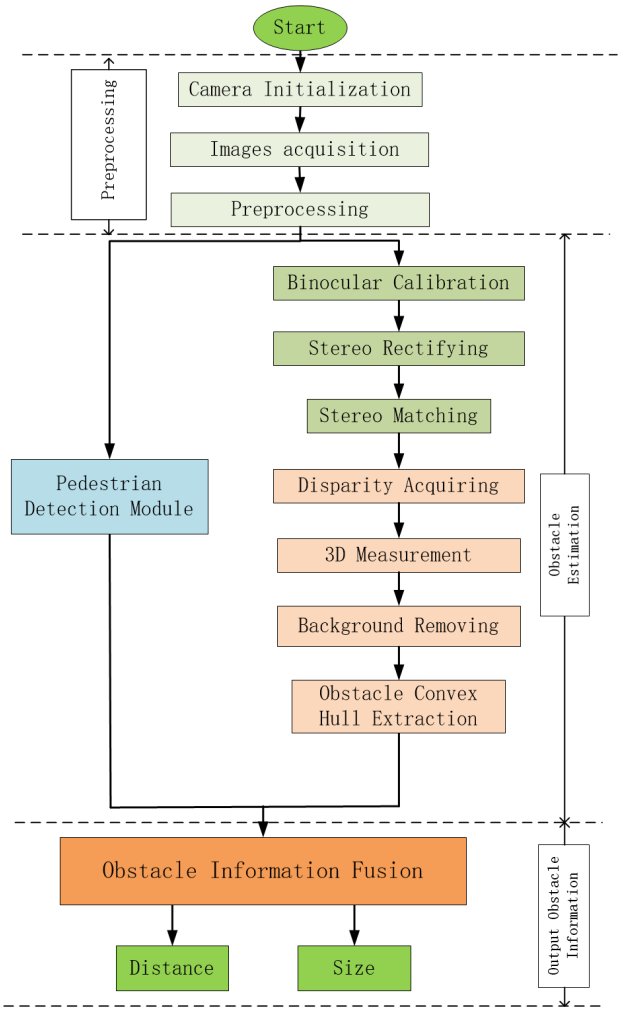


Fig. 1. The obstacle estimation system overview.

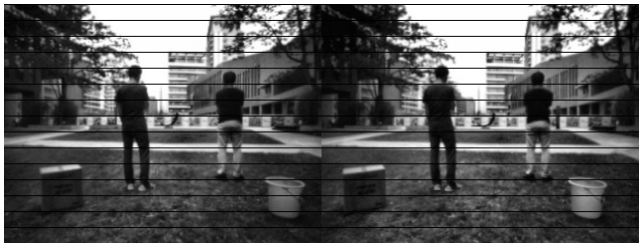


Fig. 2. Left and right images after rectifying.

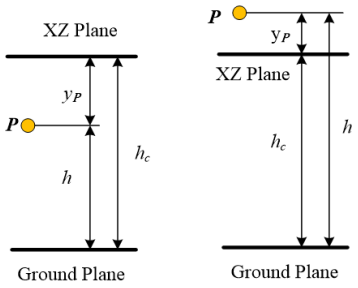
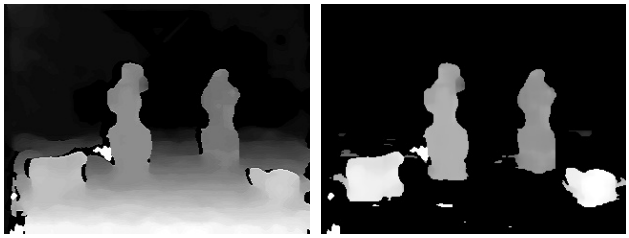


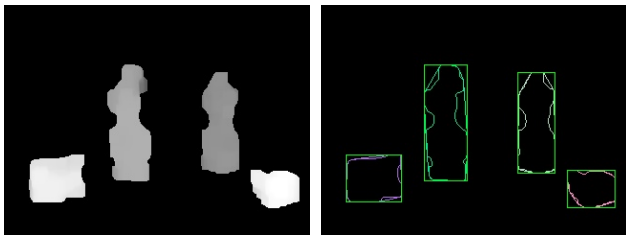
Fig. 3. The height h of point $\boldsymbol{P}$.



(a) (b)

Fig. 4. Remove interference information. (a) The disparity map, and (b) The ground information removal.

coordinates of the obstacles. Fig. 5 shows the above process:



(a) (b)

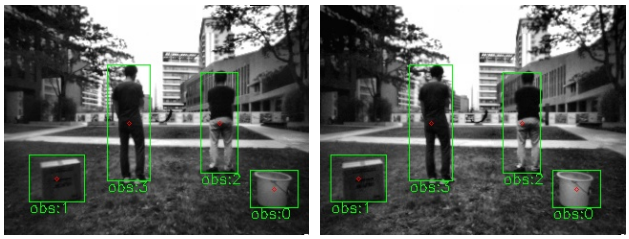
Fig. 5. Obstacle bounding boxes. (a) After the morphological operation, and (b) The obstacle convex hulls extraction.

The detected obstacles are displayed in Fig. 6 with the green rectangle boxes. The obstacle detection algorithm based on the disparity map can detect all front obstacles, both static obstacles and dynamic ones. In addition, the average precision rate of the obstacle estimation module is 94%.

C. Binocular Distance Measurement

The principle diagram of the binocular distance measurement is shown in Fig. 7:

In Fig. 7, $\boldsymbol{P}$ is a 3D point in the space, while $\boldsymbol{P_R}$ and $\boldsymbol{P_T}$ are two projection points of $\boldsymbol{P}$ on the two image planes. $\boldsymbol{R}$ and $\boldsymbol{T}$ are two camera optical centers, and f is the focal length. Besides, b is the baseline between the two cameras. In this stereo vision module, the optical axes of the two cameras are parallel. x_R is the distance between $\boldsymbol{P_R}$ and the left edge of the left image plane, while the x_T is the distance between $\boldsymbol{P_T}$ and the left edge of the right image plane.



(a) (b)

Fig. 6. Obstacle detection. (a) The left image obstacle detection, and (b) The right image obstacle detection.

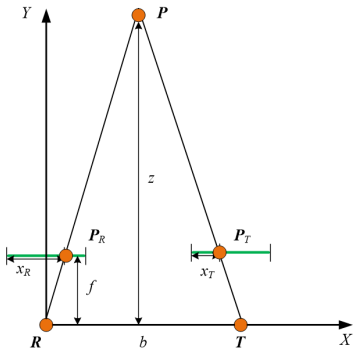


Fig. 7. The stereo distance measurement.

Because the two cameras have been rectified, the polar lines are parallel, and the directions of the two optical axes are also parallel. The relationship between the disparity and the depth of the obstacle is shown in the disparity-depth formula:

$$\frac{b}{z} = \frac{b + X_T - X_R}{z - f} \tag{1}$$

From (1), we can derive the depth formula (2):

$$z = \frac{b \times f}{X_R - X_T} \tag{2}$$

Since b and f are known, given the disparity $(X_R - X_T)$ in section IV.A, the real distance of the corresponding point can be obtained through the depth formula (2). Next we will calculate the distance of each obstacle.

Firstly, we calculate the three-dimensional coordinates of all boundary points of the obstacles. Then we can gain the real-world distances of the obstacles based on the three-dimensional coordinates. To get the nearest distance between the mowing robot and the obstacle, we select the minimum value of all distance values as the obstacle distance. The distance will be transmitted to the robot and assist the robot in operating the next action.

In order to further verify the accuracy of the binocular stereo vision system, we measure 10 sets of data within 10 meters and compare the measured distance with the real distances. Table I shows the results of the tests, and in this paper, GT indicates the ground truth. From the results, it can be seen that the ranging errors within 1 m - 7 m are all less than 30 cm, and the error rate within 10 m is less than

5%. We can conclude that the accuracy will decrease as the distance increases.

TABLE I
DISTANCE MEASUREMENTS

No.	GT (m)	Calculated (m)	Error (cm)	Error rate
1	1.342	1.365	2.3	1.714%
2	2.229	2.271	4.2	1.884%
3	3.728	3.808	8	2.146%
4	4.568	4.701	13.3	2.912%
5	5.108	5.297	18.99	3.718%
6	6.223	6.466	23.3	3.738%
7	7.018	7.292	27.4	3.904%
8	8.466	8.801	33.5	3.957%
9	9.643	10.038	39.5	4.096%
10	10.556	11.042	48.6	4.608%

D. Obstacle Size Calculation

When a lawn mower is working, the size information of the obstacles will prevent it from colliding with them, or help it avoid long-distance bypasses. Thus obstacle sizes are beneficial for path planning. The general method for calculating the obstacle size is the Triangle Similarity Method (TSM). The obstacle size is calculated based on three parameters, i.e., the camera focal length (*f*), the width of the rectangle on the image (*del*), and the depth of the obstacle (*Z*). The calculation principle diagram is shown in Fig. 8.

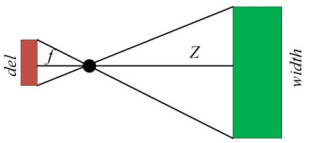
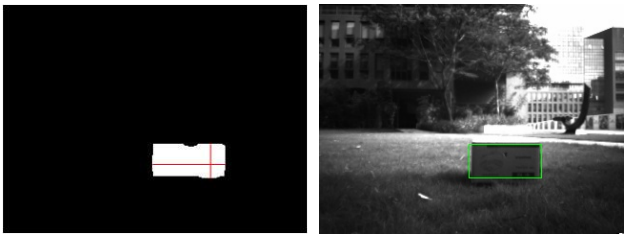


Fig. 8. The TSM diagram.

According to Fig. 8, the obstacle size can be calculated as in (4):

$$width = \frac{del \times Z}{f} \tag{4}$$

Take the box size calculation as an example, and the result is shown in Fig. 9 and Table II.



(a) (b)

Fig. 9. The TSM size calculation. (a) The box in the disparity map, and (b) The box in the grey image.

For the sizes of every object, we measure them ten times and then take the average. From Table II, we can see that the measured average error of TSM is relatively low.

TABLE II
MEASUREMENT ERRORS OF TSM

Method	Real size (cm)	Measured average size (cm)	Errors (cm)
TSM	44×26.5	48.56×25.42	$4.56 \times (-1.08)$
	55.8×58.9	62.87×55.42	$7.07 \times (-3.48)$

Table III shows the height and width measurements of a box at different distances, and the real width and height are 58.9 cm and 55.8 cm, respectively. We can find that the maximum errors of size calculations are both less than 6 cm at different distances.

TABLE III
BOX (58.9CM × 55.8CM) MEASUREMENTS

No.	GT(m)	Width(cm)	Height(cm)	Max error(cm)
1	1.342	60.33	50.98	4.82
2	2.229	60.7	53.2	2.6
3	3.728	61.4	53.99	2.24
4	4.568	62.86	54.59	3.96
5	5.108	63.04	55.66	4.14
6	6.223	63.15	56.22	4.25
7	7.018	64.03	56.3	5.13
8	8.466	64.15	57.4	5.25
9	9.643	64.22	57.66	5.32
10	10.556	64.8	58.19	5.9

V. PEDESTRIAN DETECTION

Because the safety of pedestrians is the highest, it is necessary to distinguish the pedestrians from other obstacles further. In this paper, our system designs a module specific for pedestrians. In this module, a traditional machine learning-based algorithm and a deep learning-based algorithm are presented. They will be introduced next.

A. Machine Learning-Based Pedestrian Detection

Generally, pedestrian detection algorithms are based on feature extractions and classifiers. In this part, HOG features and cascade classifiers are combined to detect pedestrians. The pipeline of this algorithm is shown in Fig. 10.

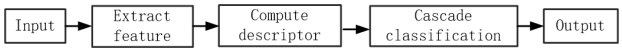


Fig. 10. The cascade classification algorithm.

- Extract HOG Features: HOG features have been widely used in image recognition, especially in pedestrian detection. To calculate HOG descriptors, we choose 64×128 as the window size, and the cell size is 6×6. Besides, there are 3×3 cells per block and nine gradients bins per cell. Therefore, a 3780-dimensional descriptor vector can be obtained.
- Cascade classifier: After extracting the HOG features of positive and negative samples, the cascade classifier is trained. When the trained pedestrian classifier is used to distinguish pedestrians, a sliding window is used

to scan the whole image and then the HOG features are extracted from the windows. Finally, the detected pedestrians will be displayed in red rectangular boxes.

B. Deep Learning-Based Pedestrian Detection

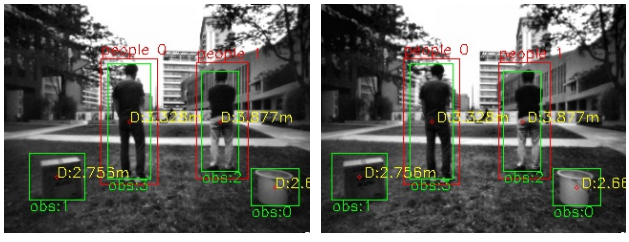
The deep learning algorithm is the current state of the art in the target detection domain. It uses the end-to-end network model for training and prediction, and eliminates the process of feature extraction. It is faster and more robust than the traditional machine learning algorithms.

In this paper, we choose the SSD as the backbone network for pedestrian detection. The reason why we prefer SSD is that SSD is faster than the Fast R-CNN and Faster R-CNN. To improve the real-time performance, we modify the basic network of SSD and replace the VGG16 [17] with MobileNetV2 [18]. MobileNetV2 is based on the depthwise separable convolution, so it has fewer parameters as well as faster training and detection speed. At the same time, it can improve the accuracy rate. To further improve the speed, the standard convolution of the auxiliary network layer is modified to the depthwise separable convolution. The model is trained on the PASCAL VOC2012 dataset [19] because VOC2012 has a large number of people, and the environment is diverse.

We train the Tensorflow model with a batch size of 32 frames and with 0.01 initial learning rate, which is decayed by 0.1 at 5k, 10, 15k epochs, respectively. The proposed model can be trained in our own laptop, and the detailed configurations are shown as follows:

- Operating System: 64-bit Ubuntu 16.04
- CPU: i7-8700k
- GPU: one Nvidia GTX TITAN XP
- RAM: 32G

As shown in Fig. 11, the red rectangles represent the detected pedestrians, which can accurately determine whether the obstacle is a pedestrian. Therefore, the proposed algorithm can detect all obstacles in front of the mowing robot and output the distances and sizes of obstacles. Besides, it can also detect pedestrians in real time.



(a) (b)

Fig. 11. Pedestrian recognition. (a) The left image pedestrian recognition, and (b) The right image pedestrian recognition.

VI. EXPERIMENTS

A. Experiment Setup

The 3D model and the setup of the robotic lawn mower are shown in Fig. 12 and Fig. 13, respectively. The stereo camera is mounted on the front end of the mowing robot.

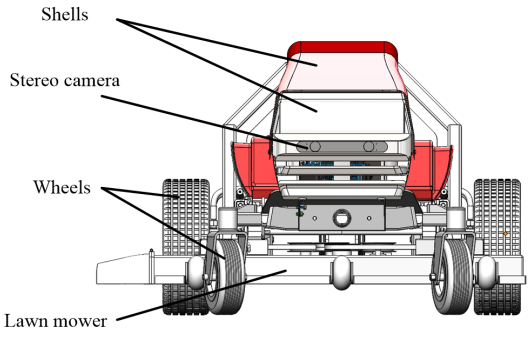


Fig. 12. The 3D model of the robotic lawn mower.

Fig. 13. The setup of the robotic lawn mower.

To verify our algorithm, we choose one experiment area in the central park, and the park is located in Chongqing, China. This park needs regular mowing, and it contains kinds of obstacles, such as trees, pedestrians, trash cans, and so on. The environment and device information about the experiments are shown in Table IV.

TABLE IV
THE ENVIRONMENT AND CAMERA INFORMATION

Test environment	
Temperature	About 19°
Humidity	$\leq 60\%$
Test interval	From 08:30 am to 19:00 pm
Weather	Sunny and rainy days
Device information	
Camera type	Stereo camera
Visual angle	Horizontal: 60°, Vertical: 56°
Sensing range	12 meters
Resolution	320 × 240

B. Experiment Results

We test the Pedestrian Detection (PD) algorithm under different moving speeds of the mowing robot, and the PD algorithm is based on the traditional Machine Learning (PD-ML) or Deep Learning (PD-DL) algorithms. The detection accuracy is shown in Fig. 14. Fig. 14 (a) shows the case

in which the robotic lawn mower is moving while the pedestrians are static, and Fig. 14 (b) shows the case in which the robotic lawn mower and pedestrians are moving together. We can quickly observe that the accuracy of the PD-DL is much higher than the PD-ML in both situations. The average precision of PD-DL in Fig. 14 (a) and (b) are 98.92% and 97.6%, respectively, so we can say that when the environment becomes complex (e.g., the people are moving), there will be a slight accuracy drop. Through the observation, it can be found that when the mower is stationary or moving slowly, the accuracy of the pedestrian recognition is close to 100%, and the accuracy decreases with the increase of the speed. Through the experiments, we have found that the PD-ML is faster than PD-DL, and the average detection time of PD-ML is about 0.041 seconds per frame. While the PD-DL is more accurate than PD-ML.

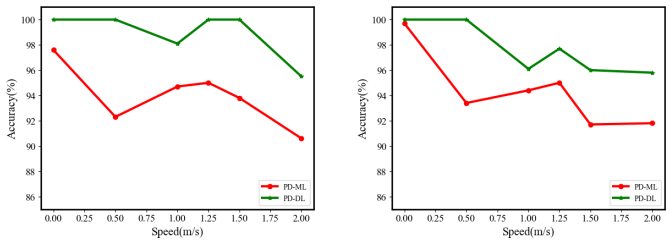


Fig. 14. The accuracy of pedestrian detection. (a) The accuracy of PD-ML and PD-DL when people are static, and (b) The accuracy of PD-ML and PD-DL when people are moving.

C. Comparison with Other Robotic Lawn Mowers

For the obstacle detection of the lawn mower robot, [8] detected three different obstacles, and the detection rate is about 80%. However, all three obstacles are static, and also it does not consider the distances and sizes of obstacles. In [9], it obtains a higher detection rate, which is 88%. However, it also neglects the distances and sizes of obstacles. In this paper, we implement the obstacle detection algorithm with higher accuracy (94%), and also calculate the distances and sizes with low errors, see Table V.

TABLE V

COMPARISON WITH OTHER METHODS

Method	Zin et al. [8]	Franzius et al. [9]	Ours
Accuracy	80%	88%	**94%**

VII. CONCLUSION

In this paper, we design a system for real-time obstacle estimation and pedestrian detection. For the obstacle estimation, the robotic lawn mower can detect all obstacles in front of it, including dynamic and static obstacles. The obstacle estimation module can also estimate distances and sizes, and the average detection accuracy is 94%. The maximum measured distance can reach 10 meters, and the distance error within 7 m is less than 30 cm. The size error of the obstacle is also less than 6 cm when the maximum distance is over 10 meters. For pedestrian detection, the traditional cascade classification algorithm is simpler and faster, while the deep learning end-to-end method has higher accuracy. Experimental results show that the created system can detect pedestrians in real time with an accuracy of 97.6%.

REFERENCES

[1] H. Sahin and L. Guvenc, "Household robotics: autonomous devices for vacuuming and lawn mowing [applications of control]," *IEEE Control Systems Magazine*, vol. 27, no. 2, pp. 20–96, 2007.

[2] R. W. Hicks II and E. L. Hall, "Survey of robot lawn mowers," in *Intelligent Robots and Computer Vision XIX: Algorithms, Techniques, and Active Vision*, vol. 4197. International Society for Optics and Photonics, 2000, pp. 262–269.

[3] G. Newstadt, K. Green, D. Anderson, M. Lang, Y. Morton, and J. McCollum, "Miami redblade III: A GPS-aided autonomous lawnmower," *Positioning*, vol. 8, no. 1, pp. 25–34, 2010.

[4] S. Baichbal, "Mapping algorithm for autonomous navigation of lawn mower using Sick laser," 2012.

[5] Y. Xin, H. Liang, T. Mei, R. Huang, J. Chen, P. Zhao, C. Sun, and Y. Wu, "A new dynamic obstacle collision avoidance system for autonomous vehicles," *International Journal of Robotics & Automation*, vol. 30, no. 3, pp. 278–288, 2015.

[6] J. A. Beno, "CWRU cutter: Design and control of an autonomous lawn mowing robot," Ph.D. dissertation, Case Western Reserve University, 2010.

[7] J. Zhang, B. Yang, N. Geng, and L. Huang, "An obstacle detection system based on monocular vision for apple orchardrobot," *International Journal of Robotics and Automation*, vol. 32, no. 6, 2017.

[8] Z. M. Zin and R. Ibrahim, "Vision-based obstacle recognition system for automated lawn mower robot development," in *Third International Conference on Digital Image Processing (ICDIP 2011)*, vol. 8009. International Society for Optics and Photonics, 2011, p. 80092K.

[9] M. Franzius, M. Dunn, N. Einecke, and R. Dirnberger, "Embedded robust visual obstacle detection on autonomous lawn mowers," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 44–52.

[10] A. Iloie, I. Giosan, and S. Nedevschi, "UV disparity based obstacle detection and pedestrian classification in urban traffic scenarios," in *2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE, 2014, pp. 119–125.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005.

[12] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1491–1498.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[14] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[19] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, no. 1, pp. 98–136, 2015.