Array 18 (2023) 100283

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

Improved VIDAR and machine learning-based road obstacle detection method

Yuqiong Wang, Ruoyu Zhu, Liming Wang, Yi Xu *, Dong Guo, Song Gao

School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, China

ARTICLE INFO

Keywords:
Road obstacle detection
VIDAR
Machine learning
Monocular vision
MSER
Normalized cross-correlation

ABSTRACT

There are various types of obstacles in an emergency, and the traffic environment is complicated. It is critical to detect obstacles accurately and quickly in order to improve traffic safety. The obstacle detection algorithm based on deep learning cannot detect all types of obstacles because it requires pre-training. The VIDAR (Vision-IMU-based Detection and Range method) can detect any three-dimensional obstacles, but at a slow rate. In this paper, an improved VIDAR and machine learning-based obstacle detection method (hereinafter referred to as the IVM) is proposed. In the proposed method, morphological closing operation and normalized cross-correlation are used to improve VIDAR. Then, the improved VIDAR is used to quickly match and remove the detected unknown types of obstacles in the image, and the machine learning algorithm is used to detect specific types of obstacles to increase the speed of detection with the average detection time of 0.316s. Finally, the VIDAR is used to detect regions belonging to unknown types of obstacles in the remaining regions, improving detection performance with the accuracy of 92.7%. The flow of the proposed method is illustrated by the indoor simulation test. Moreover, the results of outdoor real-world vehicle tests demonstrate that the method proposed in this paper can quickly detect obstacles in real-world environments and improve detection accuracy.

1. Introduction

Autonomous driving systems rely on road obstacle detection for obstacle location, tracking, distance, and speed measurement. Despite the existence of detection methods with high detection accuracy and good robustness, such as LiDAR and millimeter wave radar, their application in low-cost vehicles is limited due to their high cost [1–4]. Vision-based obstacle detection has the advantages of rich detection information, low cost, strong scalability, low hardware requirements, and robust programmability [5–7]. Vision-based obstacle detection methods can be classified into morphology-based methods, machine learning-based methods, and motion compensation-based methods. Morphology-based methods are not applied for autonomous driving systems because of the low detection accuracy.

Many machine learning algorithms can identify specific targets such as pedestrians, bicycles, vehicles, and traffic signs for self-driving systems [8–16]. Cheng E J, Prasad M et al. referred to the Deformable Part Model and combined the Adaboost, Haar-like features, and support vector machine to efficiently and accurately detect pedestrians [17]. Yu G, Wang S et al. estimated atmospheric illumination and transmissivity by the transmission network and the airlight network prior to defogging by the refinement network, and a key points-based network will effectively detect the vehicle in the defogged images [18]. Chang S., Zhang Y., et al. proposed an end-to-end training spatial attention fusion with a deep learning detection network and constructed a generation that trained the neural network by converting radar points into images [19].

With the rapid development of computer vision and machine learning, the requirements for accuracy and speed in monocular obstacle detection are also increasing [20–27]. Machine learning improves the classification ability of obstacle detection based on vision, thereby making obstacle detection based on machine learning the mainstream for automatic driving environment perception. Nevertheless, the machine learning-based monocular vision obstacle detection method can only detect trained specific types of obstacles, as shown in Fig. 1. There are often unknown types of obstacles in emergencies, which are likely to have a serious impact on vehicles. Therefore, applying the generalized obstacle detection method capable of detecting any three-dimensional obstacles to the monocular vision obstacle detection system is crucial for enhancing road traffic safety.

Methods based on motion compensation, such as the optical flow

* Corresponding author.
E-mail address: xuyisdut@163.com (Y. Xu).

https://doi.org/10.1016/j.array.2023.100283
Received 29 November 2022; Received in revised form 23 January 2023; Accepted 31 January 2023
Available online 10 February 2023

Fig. 1. Detection result of emergencies with unknown types of obstacles using the YOLO v3.

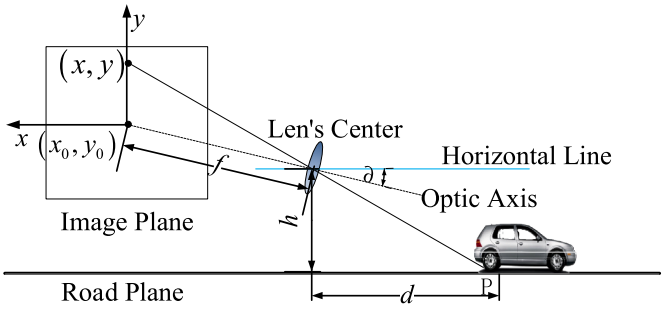

Fig. 2. Schematic diagram of 3D obstacle pinhole imaging.

method, can detect obstacles above the road surface [28–31]. Kaneko A M and Yamamoto K proposed a model based on a flat surface to estimate the height of an obstacle while taking unevenness into account [32,33]. Jung S., Cho Y., et al. extracted feature points using the Harris detector before classifying background and foreground using epipolar geometry [34]. These methods are based on the change of pixel to achieve the detection of the target. While the detection result will be more precise, more feature points will be generated.

Stable regions cover more pixels. Compared to the algorithm for detecting feature points, region detection can detect fewer and more stable numbers, and it is easier to track a target. Xu et al. improved the Maximally Stable Extremal Regions (MSER) method for detecting and matching extreme regions in adjacent frames [35,36] in previous research. In addition, it was proposed that the VIDAR(Vision-IMU-based Detection and Range method), using MSER and pinhole imaging, could effectively detect obstacles higher than the road surface (hereinafter referred to as obstacles of unknown type) [37,38]. However, VIDAR runs longer than methods based on machine learning.

This paper proposes an obstacle detection method that uses improved VIDAR and machine learning to detect obstacles more efficiently for self-driving systems. In the proposed method, the YOLO v3 is used as a machine learning model to detect specific types of targets, while the improved VIDAR rapidly detects obstacles higher than the road surface in the non-target region.

This paper is organized as follows. In Section 2, the principal of the improved VIDAR is presented. In Section 3, a brief overview of the improved VIDAR and machine learning-based obstacle detection method is given. In Section 4, the performance results of the proposed method are presented and analyzed. In Section 5, the conclusion and future work are drawn.

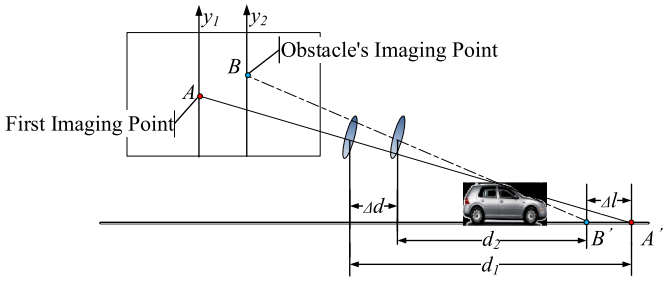

Fig. 3. The static obstacle imaging.

2. Principal of the improved VIDAR

In emergency situations, unknown obstacle detection is a necessary supplement to machine learning-based obstacle detection. For an unknown type of obstacle detection method based on motion compensation, the conventional feature extraction, and matching method is both time-consuming and space-consuming. Although the VIDAR using the MSER-based image region matching method proposed in papers [37,38] can effectively detect unknown types of obstacles, the detection speed can be increased.

2.1. The VIDAR

As shown in Fig. 2, suppose that the effective focal length of the camera is f, the optical axis height of the camera lens from the ground is h, the pixel size is μ, the pitch angle of the camera is σ, the coordinate origin of the image coordinate system (x_0,y_0), and the coordinates of the intersection of the front obstacle and the road plane in the image plane coordinate system (x,y) are known. Thus, the horizontal distance d between the camera and the intersection of the front obstacle and the road plane can be worked out.

$$d=\frac{h}{\tan(\partial+\arctan[(y_0-y)/f])} \tag{1}$$

Suppose that y_1 is the y-axis in the previous image, y_2 is the y-axis in the latter image. The imaging point of the obstacle's top is A in the previous image (see Fig. 3), and the imaging point of the obstacle's top is B in the latter image. Assuming that the obstacle is two-dimensional on the road plane, the corresponding point of A on the path plane is A', and the corresponding point of B on the path plane is B'. Then, the horizontal distance from A' to the camera is d_1, and the horizontal distance from B' to the camera is d_2. d_1 and d_2 can be calculated using equation (1). The

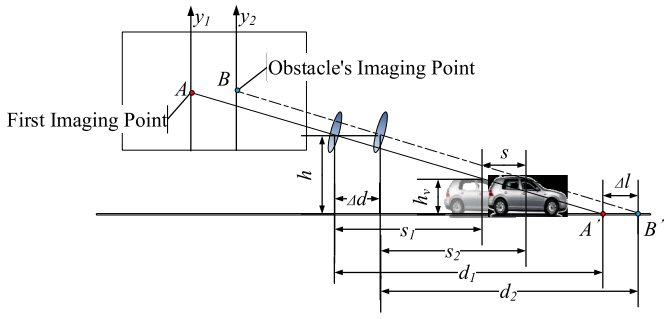

Fig. 4. The moving obstacle imaging.

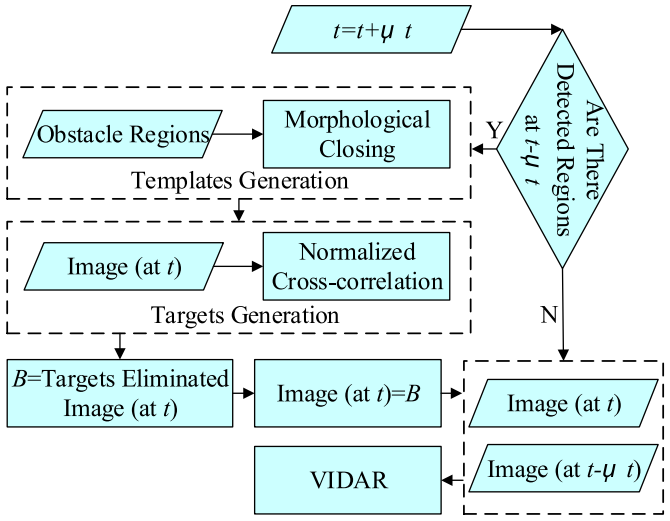

Fig. 5. The improved VIDAR flow.

camera moved a certain distance of Δd during the time between the previous image and the latter image, $d_1 = d_2 + \Delta d$. Actually, the obstacle is three-dimensional, and $d_1 = d_2 + \Delta d + \Delta l$. As a result, A' and B' have height if $d_1 \neq d_2 + \Delta d$. In VIDAR, d is obtained using IMU (Inertial Measurement Unit), and static three-dimensional obstacles can be recognized by Δl.

In VIDAR, feature points are first extracted using the MSER fast image region matching method, and the two before and after image frames are then matched. In the obstacle range, the lowest point of the maximally stable extremal region connected to the detected region is considered the intersection point of the obstacle and road plane, and pinhole imaging is used to calculate the distance between the camera and the obstacle. On this basis, the VIDAR stereoscopic obstacle discrimination principle is applied to eliminate the extracted non-obstacle points, allowing for the direct and rapid detection of image obstacles.

Besides, if the obstacle is moving (see Fig. 4), the Δl can also be used as an obstacle judgment in most environments. The relevant parameter description and certification process are shown in papers [38,39].

2.2. The improved VIDAR

We discovered that every time VIDAR is utilized, every pixel in the image must be processed. The corresponding regions will be processed multiple times despite the fact that the morphological changes of unknown-type obstacles may not be significant. As depicted in Fig. 5, we improved the VIDAR in order to increase the speed of detection.

Before each use of the VIDAR, determine whether unknown types of obstacle regions were detected in the last use of the VIDAR. If the answer is no, the image at t and the image at $t+\Delta t$ will be input into the VIDAR. If the answer is yes, a morphological closing operation is performed on the unknown type of obstacle regions, and the results are used as templates. Calculate the normalized cross-correlation matrix of the image at $t+\Delta t$ and templates, and find the peaks of normalized cross-correlation matrix as unknown type of obstacle target positions. The image at $t+\Delta t$ after removing unknown type of obstacle targets will be regarded as background, and then be input into the VIDAR together with the image at t. Unknown types of obstacle regions detected by the VIDAR will serve as the basis for the subsequent evaluation.

In the improved VIDAR, the morphological closing operation can classify multiple detected regions belonging to each unknown type of obstacle into a single category, thereby decreasing the number of templates. Using the normalized cross-correlation method instead of the feature matching method for template matching can eliminate the need for feature extraction and reduce the matching time. Compared to the VIDAR, the improved VIDAR uses template matching, which reduces the amount of data required for MSER feature extraction and matching and shortens the time required to detect an unknown type of obstacle.

3. Improved VIDAR and machine learning-based obstacle detection method

Obstacle detection methods based on machine learning, such as the YOLO [39], can achieve a speed of more than 40 frames per second. To ensure that all types of obstacles can be detected quickly in emergency situations, a VIDAR- and machine-learning-based method for obstacle detection is proposed. First, the improved VIDAR is used to match detected regions as unknown types of obstacles, then the machine learning framework is used to distinguish specific types of obstacles from background regions, and finally, the VIDAR is used to detect regions belonging to unknown types of obstacles in the remaining regions.

We do not use machine learning frameworks to match detected obstacles because machine learning frameworks require training samples prior to detecting a new type of target, the number of samples in emergency situations is difficult to meet the demand for, and the online training will reduce the speed.

3.1. Process of improved VIDAR and machine learning-based obstacle detection method

The process of improved VIDAR and machine learning-based obstacle detection method is shown in Fig. 6.

3.1.1. Improved VIDAR-based unknown type of obstacles extraction

If no region belonging to unknown types of obstacles was detected when the VIDAR was the last run, the image at t and the image at $t + \Delta t$ are represented by I_{it} and $I_{it+\Delta t}$, which will be processed by machine learning. If regions belonging to unknown types of obstacles were detected when the VIDAR was the last run, detected regions would be divided into several obstacles by morphological closing operation, which are represented by the template set $Te(Te = \{Te_1, Te_2, ..., Te_m\})$. The normalized cross-correlation for finding matches of Te in the image at $t + \Delta t$ $t + \Delta t$ is referred to formula (1) in paper [40]. The found matches $Go(Go = \{Go_1, Go_2, ..., Go_n\})$ are regarded as unknown types of obstacles. The image at $t + \Delta t$ after removing Go will be represented by $I_{it+\Delta t}$, and then be input into the machine learning framework together with the image at t represented by I_{it}.

3.1.2. Machine learning based on specific types of obstacles extraction

$I_{it+\Delta t}$ will be processed by specific types of obstacle samples trained by a machine learning framework. In $I_{it+\Delta t}$, the identified and classified obstacles are specific types of obstacles, which are represented by $So(So = \{So_1, So_2, ..., So_i\})$. After removing specific types of obstacles, I_{it} and $I_{it+\Delta t}$ will be represented by I_{bt} and $I_{bt+\Delta t}$, respectively, and then input into the VIDAR.

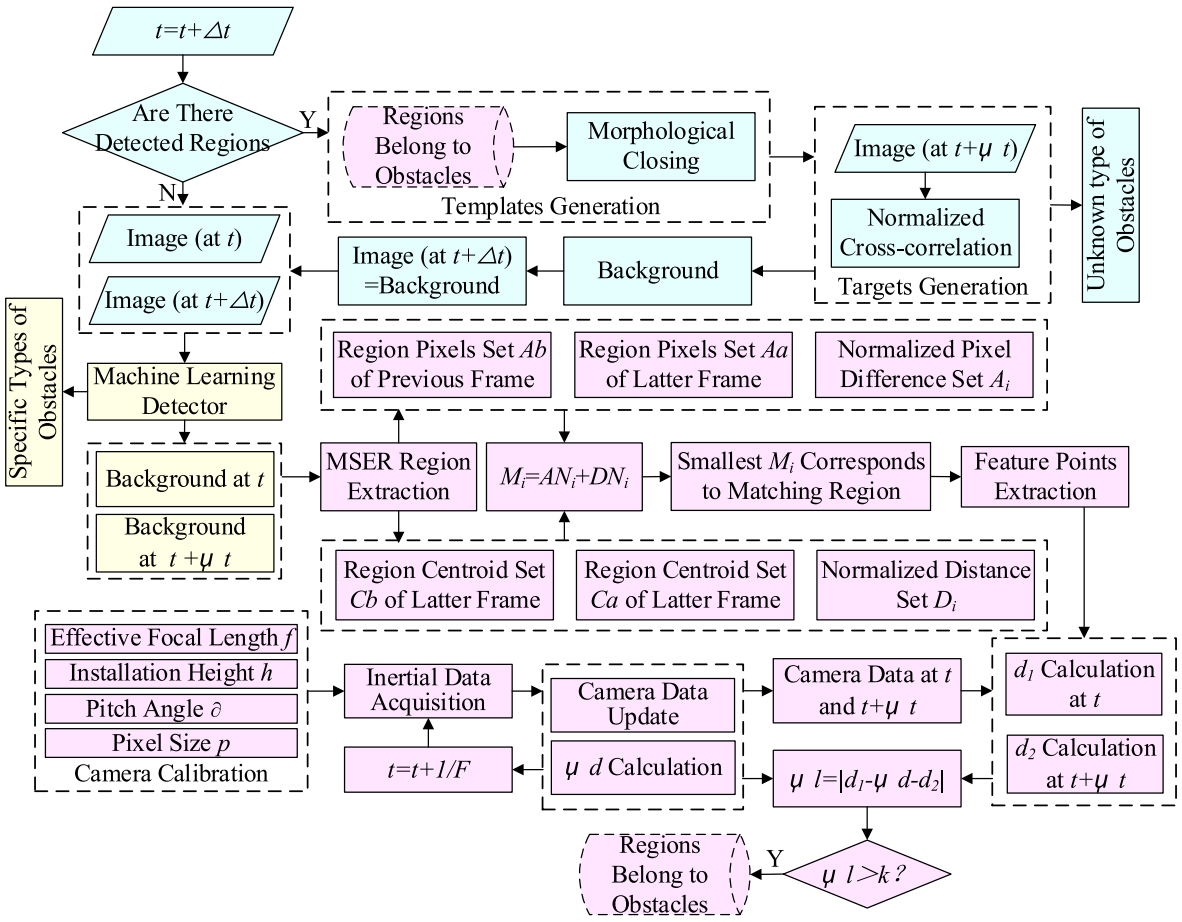

Fig. 6. Improved VIDAR and machine learning-based obstacle detection method(are the improved VIDAR steps, are machine learning steps, are the VIDAR steps).

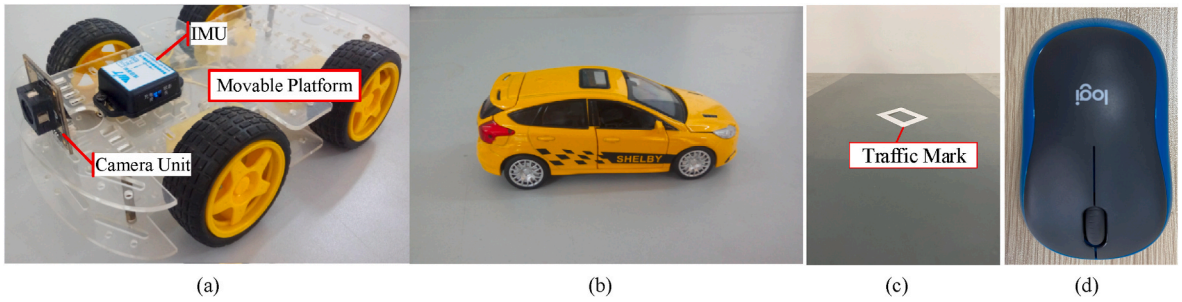

Fig. 7. Simulation test equipment. (a). Mobile platform, IMU, and camera module. (b). Car scaled model (a specific type of obstacle). (c). Road and traffic mark. (d). Mouse (unknown type of obstacle).

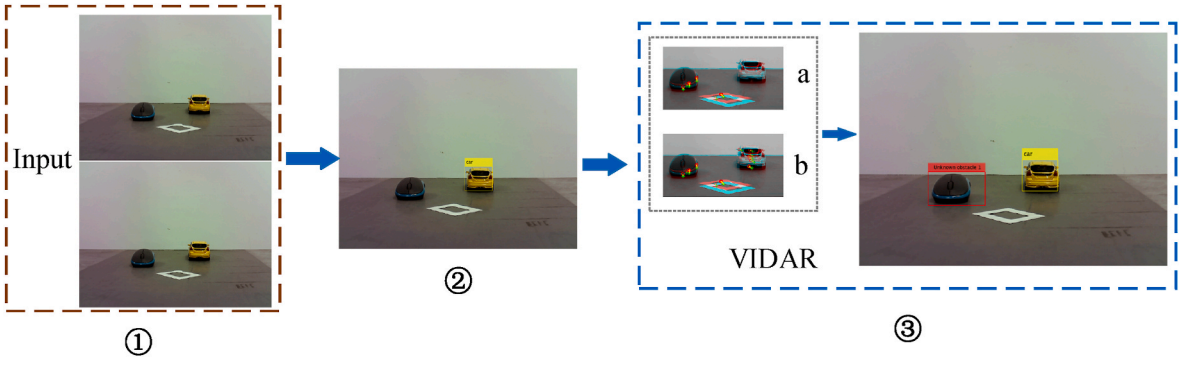

Fig. 8. Detection processing in the first loop.

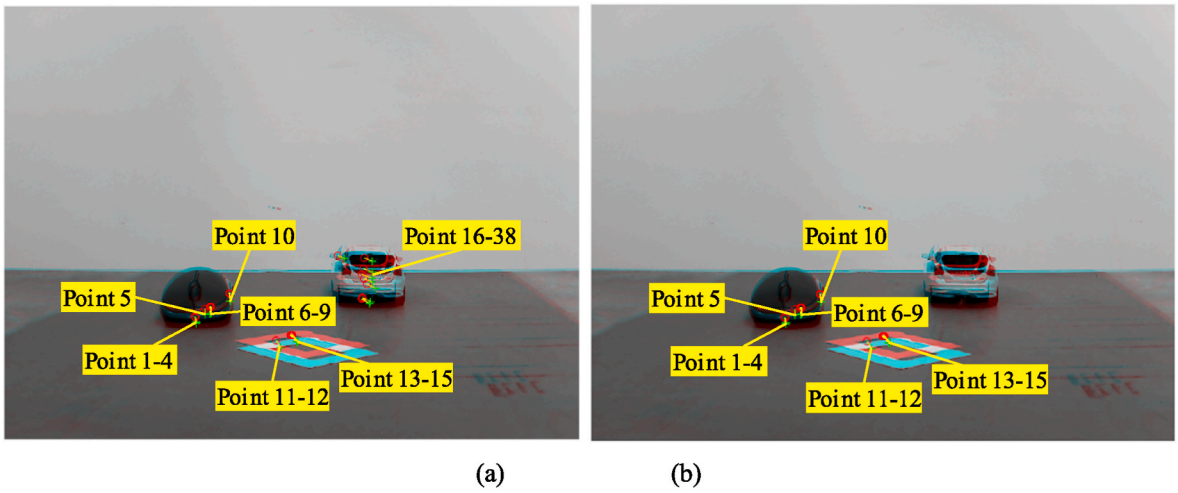

Fig. 9. Feature points extraction. (a). Result of regions matching without removing the car. (b). Result of regions matching after removing the car. The image at $t = 0$ is cyan, and the image at $t = 1$ is red. The o are centroids of maximally stable extremal regions in I_{it}. The + are centroids of maximally stable extremal regions in $I_{it+\Delta t}$.

3.1.3. Detection of regions belonging to unknown types of obstacles

MSER regions in I_{bt} and $I_{bt+\Delta t}$ will be extracted and matched by MSER based on the image region matching method. Then, mass centers of matched regions will be calculated and used as feature points. Assuming that feature points are located on the road plane, and calculated distance (d_1) from the camera to the feature point at t and the distance (d_2) from the camera to the feature point at $t + \Delta tF$; compare the obstacle judgment Δl and the threshold $k(k > 0)$; if $\Delta l \leq k$, the feature point is located on the road plane, so the corresponding region is not an obstacle; otherwise, if $\Delta l > k$, the feature point is not located on the road plane, so the corresponding region is an obstacle. Regions belonging to obstacles will be used as basic data for the improved VIDAR.

3.2. Simulation test

In this paper, simulation tests are conducted to demonstrate the method's flow. The equipment for the simulation test is shown in Fig. 7.

3.2.1. Detection processing in the first loop

Fig. 8 shows the detection processing in the first loop. The details are as follows.

Step 1. Images input.

In this case, the two images I_{it} and $I_{it+\Delta t}$ captured by the camera are taken as input.

Step 2. Machine learning detect specific types of obstacles.

As this is the first loop, no detected region is present. Consequently, machine learning is used to detect specific types of obstacles, as depicted in Fig. 6. The YOLO v3 machine learning algorithm is widely used for obstacle detection due to its high speed and precision in target recognition [41–46]. In the simulation test, as depicted in Fig. 8 ②, YOLO v3 is used to construct a machine learning detector that distinguishes car regions from background regions.

After removing specific types of obstacles (the car), I_{it} and $I_{it+\Delta t}$ are represented by I_{bt} and $I_{bt+\Delta t}$, and then are input into the VIDAR.

Step 3. VIDAR detects unknown types of obstacles.

In the VIDAR, the focal length $f = 6.779$ mm, camera height $h = 5.872$ cm, pixel size $\mu = 1.4\mu$m, and pitch angle $\partial = 0.135$rad. The position data is obtained by the IMU with $F = 100$Hz. The distance $\Delta d = 3.00$cm in period $\Delta t = 2$s is calculated using position data. MSER-based image region matching method is used to process images when $t = 0$ and $t = 1$. As shown in Fig. 9, unlike 38 matched regions obtained without removing the car, 15 matched regions are obtained, and their centroids are calculated as feature points after removing the car. d_1, d_2 and Δl are calculated as shown in Table 1.

Table 1
Units for magnetic properties.

Feature point	d_1/cm	d_2/cm	$\triangle l$/cm
1	49.89	45.21	1.10
2	50.23	45.26	1.37
3	49.57	44.69	1.29
4	50.09	45.08	1.41
5	51.39	46.28	1.50
6	51.94	46.79	1.55
7	51.83	46.73	1.51
8	51.83	46.68	1.53
9	51.73	46.59	1.53
10	58.55	53.36	1.60
11	40.46	36.83	0.53
12	40.36	36.75	0.50
13	41.19	37.60	0.62
14	41.40	37.81	0.68
15	41.24	37.63	0.59

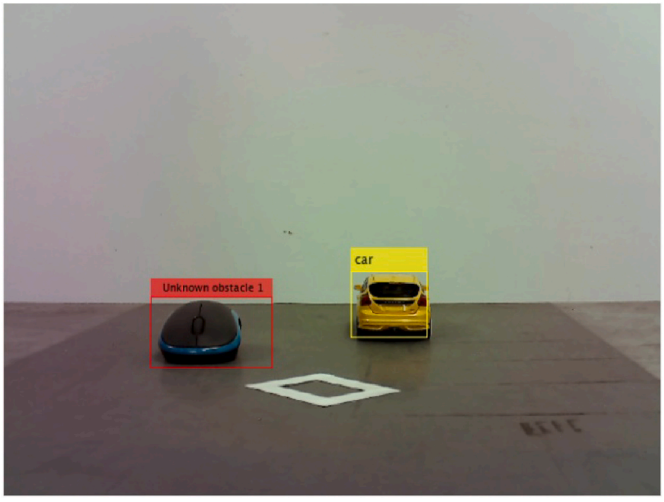

Fig. 10. Detection result.

Given that the threshold $k = 1$cm, if $\Delta l < 1$cm, the feature points are located on the road plane. As shown in Table 1, feature points 1 through

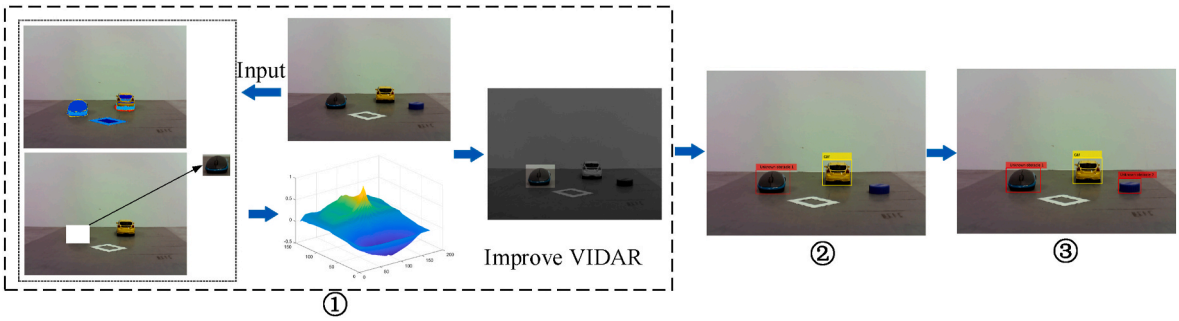

Fig. 11. Detection processing in the second loop.

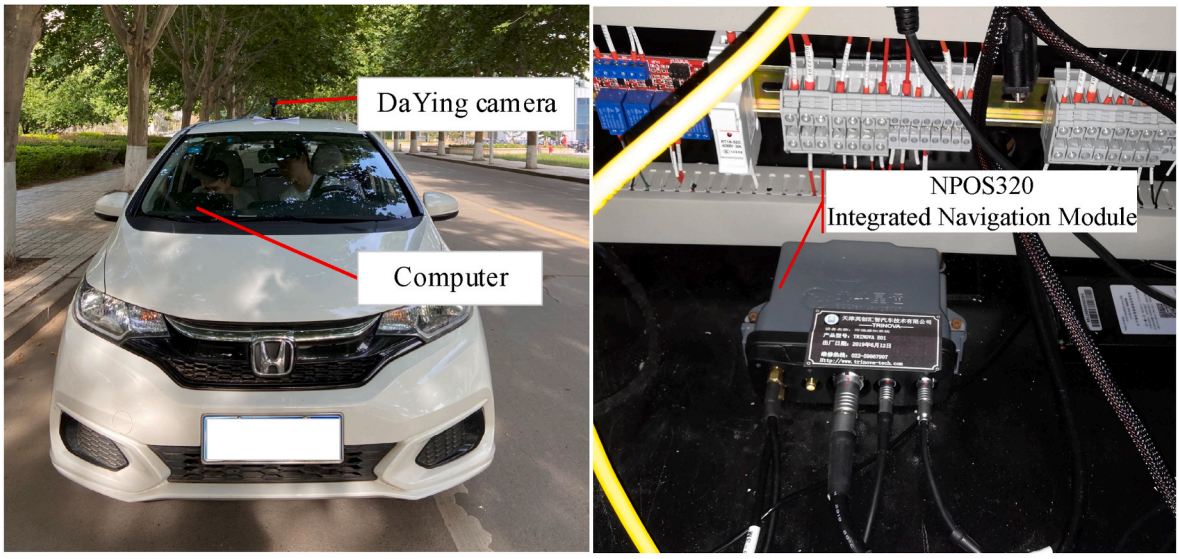

Fig. 12. The mobile platform. (a) The drive recorder on the Honda Fit car. (b) The NPOS320 integrated navigation module.

10 are not on the road plane, and their maximally stable extremal regions are considered to belong to an unknown type of obstacle. Once it is determined that these regions contain an unknown type of obstacle, emergency avoidance can be taken. In the subsequent loop, these detected regions are subdivided into obstacles of unknown type, which are used as templates for normalized cross-correlation matching. The final detection result is shown in Fig. 10. The yellow rectangle represents the target (a car), while the red rectangle represents an unknown obstacle.

3.2.2. *Detection processing in the second loop*

Fig. 11 shows the detection processing in the second loop. There is a new obstacle in the input image. Fig. 11 ① shows the processing of improved VIDAR. In the last loop, the unknown type of obstacle (mouse) that has been detected is extracted by morphological closing operation and is used as template *Te*. Find the Go that matches *Te* in the newly input image and display them together. According to Fig. 6, the final detection result is shown as Fig. 11 ③.

4. Effect analysis of improved VIDAR and machine learning-based obstacle detection method

In our outdoor test, the YOLO v3 serves as both an independent machine learning detection method and a machine learning detector in the IVM. YOLOv3 is limited in its ability to detect targets (The YOLO v3 can detect 80 types of targets if the default configuration is used). To create an emergency in the presence of unknown types of obstacles, we only consider the car to be a known type. Therefore, we modified the YOLO v3 files *yolo. cfg*, *voc_announcement.py*, *coco_class.txt*, and *voc_class. txt*. This enables the YOLO v3 to detect only cars as an obstacle. The mobile platform is a Honda Fit car. The mobile platform used in the outdoor test is shown in Fig. 12. As shown in Fig. 12, traffic images are captured by the Da Ying camera, and the NPOS320 integrated navigation module updates position data. The collected data is processed by the YOLO v3, the VIDAR, and the IVM.

Due to the lack of odometer data in the public data set and the fact that different camera parameters would impact range accuracy, we created an IVM database containing 2800 images. This paper selects five two-lane roads near Shandong University of Technology's east gate for testing. It includes Nanjing Road (1.2 km), Shiji Road (1.2 km), Renmin West Road (1.2 km), Gongqingtuan Road (1.3 km), and Xincun West Road (1.3 km). The particular test roads are shown in Fig. 13, and a portion of the IVM database is shown in Fig. 14.

4.1. *Analysis of detection accuracy*

The YOLO v3 is used as a machine learning detector in the IVM during our outdoor test. In order to verify the detection accuracy of the proposed method, a portion of the test results for the proposed method and YOLO v3 are compared, and the results are shown in Fig. 15.

Fig. 15 demonstrates that YOLO v3 can only detect trained obstacles (in this case, cars), whereas the method proposed in this paper can

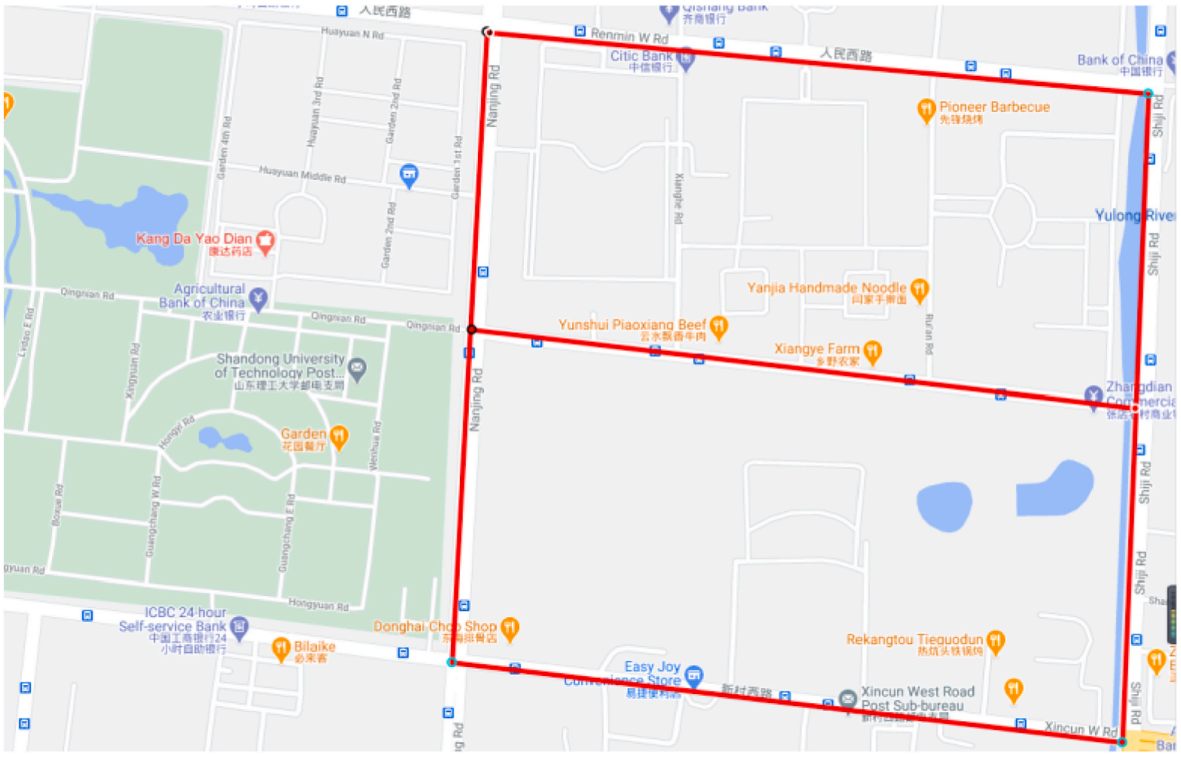

Fig. 13. The specific test roads.

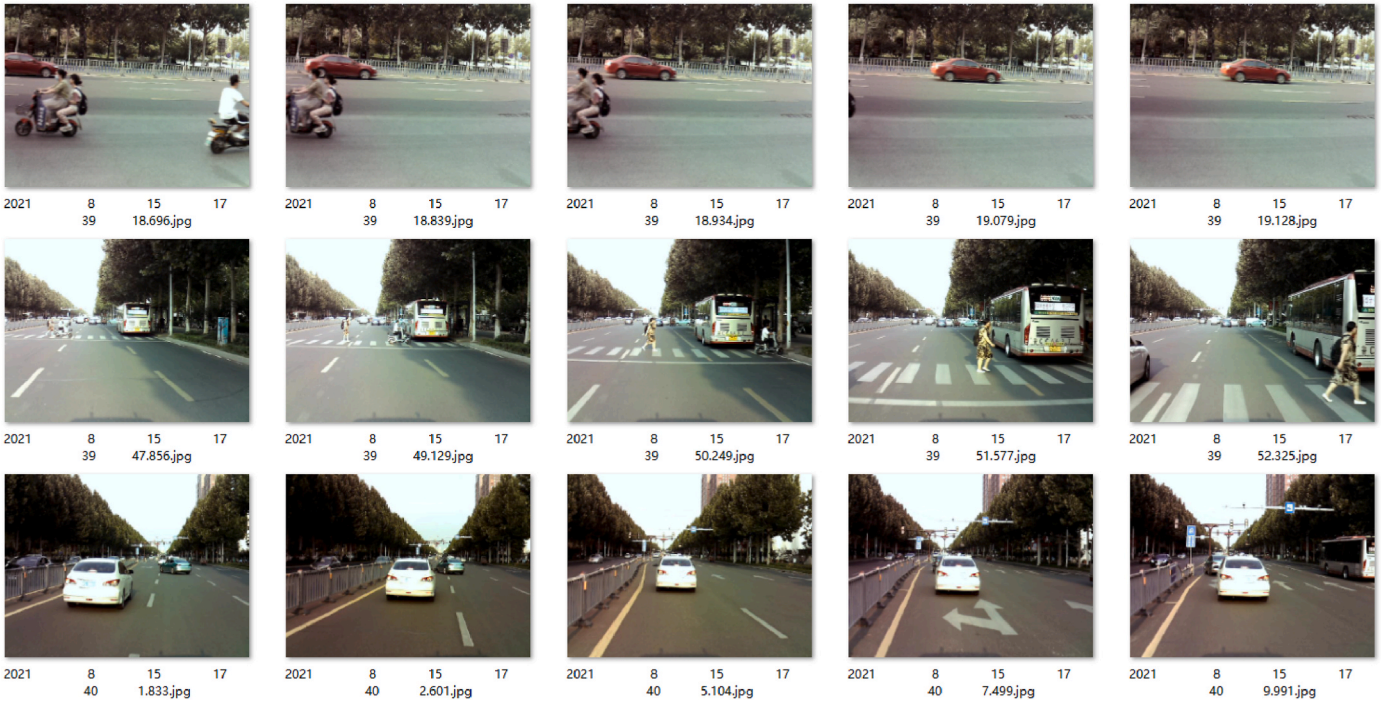

Fig. 14. Part of the IVM database.

detect road cones and other generalized obstacles due to the combination of VIDAR.

As evaluation indexes, we employ A (accuracy), P (precision), R (recall), and F_1 score, with reference to the accuracy analysis method described in the papers [47–49]. Calculate A, P, and R according to the following equations:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$p = \frac{TP}{TP + FP} \tag{3}$$

(a) Obstacles detection results of YOLO v3

(b) Obstacles detection results of IVM

Fig. 15. Comparison of part of the test results of the proposed method and YOLO v3.

Table 2
Detection results of YOLO v3, VIDAR, and IVM.

YOLO v3		Actual	
		Positive	Negative
Detected	Positive	1.268×10^7	2.147×10^7
	Negative	3.519×10^7	2.293×10^8
VIDAR		Actual	
		Positive	Negative
Detected	Positive	3.926×10^7	7.647×10^6
	Negative	8.615×10^6	2.431×10^8
IVM		Actual	
		Positive	Negative
Detected	Positive	3.897×10^7	1.305×10^7
	Negative		

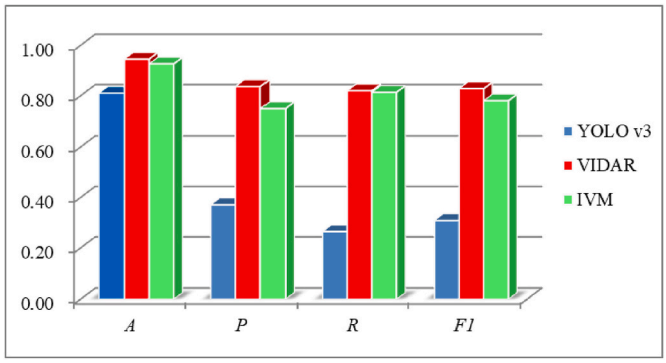

Fig. 16. Histogram of detection accuracy.

Table 3
Comparison of detection accuracy.

	A(%)	P(%)	R(%)	F_1(%)
YOLO v3	81.0	37.1	26.5	30.9
VIDAR	94.6	83.7	82.0	82.8
IVM	92.7	74.9	81.4	78.0

Table 4
Average detection speed.

	YOLO v3	VIDAR	IVM
Detection Time/s	0.082	0.538	0.316
Feature Points	–	121	89

$$R = \frac{TP}{TP + FN} \tag{4}$$

Where TP are true positives, TN are true negatives, FP are false positives, and FN are false negatives. The F_1 can be obtained from the following formula:

$$F_1 = 2\frac{P \cdot R}{P + R} \tag{5}$$

Confusion matrixes of YOLO v3, VIDAR, and IVM detection results are shown in Table 2. Detection accuracy is compared in Table 3 and Fig. 16.

A, P, R, and F_1 of detection results obtained by the YOLO v3 are significantly lower than those of the VIDAR and the IVM, as shown in Table 3 and Fig. 16. Compared to YOLO v3, IVM's detection accuracy has increased by 11.7%. Pedestrians and road cones cannot be detected as obstacles by the YOLO v3 because they are unknown types of targets. It also demonstrates that the VIDAR's accuracy indexes are superior to those of the IVM. That is because the VIDAR can detect all types of obstacles as long as the obstacles are three-dimensional. Nonetheless, it is important to note that the detection based on YOLO v3 is one of the IVM's steps, and that its accuracy will be affected by sample size and the Bayesian classification mechanism. Therefore, the accuracy of the IVM will also be affected by the YOLO v3.

4.2. Analysis of detection speed

Calculate the detection time, including the time required to extract feature points, and compare the detection speeds of the YOLO v3, the VIDAR, and the IVM. The average detection speed is shown in Table 4,

(a) (b)

Fig. 17. Matched feature points. (a). Matched feature points of the VIDAR. (b). Matched feature points of the IVM.

while the matched feature points are shown in Fig. 17.

Table 4 demonstrates that the YOLO v3 is faster than both the VIDAR and the IVM. Feature points are not necessary for the YOLO v3, and images are processed from the bottom up to the 380's pixel to reduce complexity. Thus, YOLO v3's detection speed is the fastest. In addition, the results demonstrate that the average detection time of IVM is 0.222s faster than that of VIDAR. As shown in Fig. 17, the YOLO v3 in the IVM detected specific types of obstacles (cars), the improved VIDAR matched the detected regions, and the extracted background regions are significantly smaller, resulting in the background containing fewer feature points. The fewer the feature points, the less time is required.

5. Conclusions

This paper proposes an improved VIDAR and machine learning based-obstacle detection method. This method employs morphological closing operation and the normalized cross-correlation to match and remove detected unknown types of obstacles, a machine learning framework to detect specific types of obstacles, and VIDAR to detect regions belonging to unknown types of obstacles. In this method, the improved VIDAR and machine learning algorithm can quickly remove detected unknown obstacles and specific obstacles, and reduce the number of feature points to be detected and matched. Regions belonging to unknown types of obstacles will not be missed because of the VIDAR. Therefore, this method can maximize the speed advantage of machine learning and the accuracy advantage of VIDAR. The improved VIDAR and machine learning-based obstacle detection method can detect specific types of obstacles in normal situations and unknown types of obstacles in emergency situations due to its high speed and accuracy. In addition, the choice of the machine learning algorithm is flexible, and future applications may employ more efficient machine learning algorithms. Therefore, we will try to apply the proposed method on intelligent and connected vehicle relying on wireless technologies [50] and X-by-Wire technologies [51–54] in the future. We believe that the method proposed in this paper can not only improve the detection performance of vehicle monocular vision systems, but also provide a means of enhancing the emergency safety of self-driving systems.

Author contribution

Conceptualization, Yuqiong Wang and Yi Xu; Methodology, Ruoyu Zhu; Software, Liming Wang; Validation, Yuqiong Wang and Dong Guo; Formal analysis, Yuqiong Wang; Investigation, Yi Xu; Resources, Song Gao; Data curation, Liming Wang; Writing – original draft preparation, Ruoyu Zhu; Writing – review & editing, Yuiong Wang; Visualization, Ruoyu Zhu; Supervision, Yi Xu; Project administration, Yi Xu and Song Gao; Funding acquisition, Yi Xu and Dong Guo. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

Data availability

No data was used for the research described in the article.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (grant number 51905320), China Postdoctoral Science Foundation (grant number 2018M632696), Natural Science Foundation of Shandong Province (grant number ZR2022MF230 and ZR2021QF039), Shandong Province Major Science and Technology Innovation Project (grant number 2019JZZY010911), Shandong Key R&D Plan Project (grant number 2019GGX104066), Experimental Technology Upgrading Project (grant number 2022003) and Shandong Provincial Program of Introducing and Cultivating Talents of Discipline to Universities (Research and Innovation Team of Intelligent Connected Vehicle Technology).

References

[1] Nabati R, Qi H. Centerfusion: center-based radar and camera fusion for 3d object detection[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 2021:1527–36.
[2] Charouh Z, Ezzouhri A, Ghogho M, et al. A resource-efficient CNN-based method for moving vehicle detection[J]. Sensors 2022;22(3):1193.
[3] Jiang Q, Zhang L, Meng D. Target detection algorithm based on MMW radar and camera fusion[C]//2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE; 2019. p. 1–6.
[4] Payalan YF, Guvensan MA. Towards next-generation vehicles featuring the vehicle intelligence[J]. IEEE Transactions on Intelligent Transportation Systems; 2019.
[5] Zhang Z, Cao Y, Ding M, et al. Monocular vision based obstacle avoidance trajectory planning for Unmanned Aerial Vehicle[J]. Aero Sci Technol 2020;106:106199.
[6] Badrloo S, Varshosaz M, Pirasteh S, et al. A novel region-based expansion rate obstacle detection method for MAVs using a fisheye camera[J]. Int J Appl Earth Obs Geoinf 2022;108:102739.

[7] Shi TW, Chang GM, Qiang JF, et al. Brain computer interface system based on monocular vision and motor imagery for UAV indoor space target searching. J]. Biomedical Signal Processing and Control 2023;79:104114.
[8] Mateus A, Ribeiro D, Miraldo P, et al. Efficient and robust pedestrian detection using deep learning for human-aware navigation[J]. Robot Autonom Syst 2019; 113:23–37.
[9] Yang W, Li Z, Wang C, et al. A multi-task Faster R-CNN method for 3D vehicle detection based on a single image[J]. Appl Soft Comput 2020;95:106533.
[10] Guindel C, Martín D, Armingol JM. Traffic scene awareness for intelligent vehicles using ConvNets and stereo vision[J]. Robot Autonom Syst 2019;112:109–22.
[11] Soetedjo A, Somawirata IK. Improving traffic sign detection by combining MSER and lucas kanade tracking[j]. INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING INFORMATION AND CONTROL 2019;15(2):653–65.
[12] Huang SC, Le TH, Jaw DW. DSNet: joint semantic learning for object detection in inclement weather conditions[J]. IEEE transactions on pattern analysis and machine intelligence; 2020.
[13] Wang Z, Wu Y, Niu Q. Multi-sensor fusion in automated driving: a survey[J]. IEEE Access; 2019.
[14] Choong CS, Nasir AFA, Majeed APPA, et al. Machine learning approach in identifying speed breakers for autonomous driving: an overview[M]//RITA 2018. Singapore: Springer; 2020. p. 409–24.
[15] Nguyen KD, Nguyen K, Le DD, et al. YADA: you always dream again for better object detection[J]. Multimed Tool Appl 2019;78(19):28189–208.
[16] Yin C. Hazard assessment and regionalization of highway flood disasters in China [J]. Nat Hazards 2019:1–16.
[17] Cheng EJ, Prasad M, Yang J, et al. A fast fused part-based model with new deep feature for pedestrian detection and security monitoring. J]. Measurement 2020; 151:107081.
[18] Yu G, Wang S, Li M, et al. Vision-based vehicle detection in foggy days by convolutional neural network[C]//Chinese intelligent systems conference. Singapore: Springer; 2019. p. 334–43.
[19] Chang S, Zhang Y, Zhang F, et al. Spatial attention fusion for obstacle detection using MmWave radar and vision sensor[J]. Sensors 2020;20(4):956.
[20] Galea C, Farrugia RA. Matching software-generated sketches to face photographs with a very deep CNN, morphed faces, and transfer learning[J]. IEEE Trans Inf Forensics Secur 2018;13(6):1421–31.
[21] Wu B, Xu C, Dai X, et al. Visual transformers: token-based image representation and processing for computer vision[J]. 2020. arXiv preprint arXiv:2006.03677.
[22] Issa D, Demirci MF, Yazici A. Speech emotion recognition with deep convolutional neural networks[J]. Biomed Signal Process Control 2020;59:101894.
[23] Guerrero-Ibañez J, Contreras-Castillo J, Zeadally S. Deep learning support for intelligent transportation systems[J]. Transactions on Emerging Telecommunications Technologies 2021;32(3):e4169.
[24] Brock A, Donahue J, Simonyan K. Large scale gan training for high fidelity natural image synthesis[J]. 2018. arXiv preprint arXiv:1809.11096.
[25] Ye J, Sun L, Du B, et al. Co-prediction of multiple transportation demands based on deep spatio-temporal neural network[C]. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2019:305–13.
[26] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]. Proceedings of the IEEE international conference on computer vision 2017:2961–9.
[27] Alp Güler R, Neverova N, Kokkinos I. Densepose: dense human pose estimation in the wild[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018:7297–306.
[28] Pan JS, Bingham N, Chen C, et al. Breaking camouflage and detecting targets require optic flow and image structure information[J]. Appl Opt 2017;56(22): 6410–8.
[29] Zhao T, Zhang X, Zhang J. Research on an improved SURF matching algorithm for door handle images[C]//2022 IEEE 10th joint international information Technology and artificial intelligence conference (ITAIC). IEEE 2022;10:2357–61.
[30] Sun D, Yang X, Liu MY, et al. Models matter, so does training: an empirical study of cnns for optical flow estimation[J]. IEEE Trans Pattern Anal Mach Intell 2019;42 (6):1408–23.
[31] Jiang Z, Zhang Y, Zhai H. Image registration and change detection method based on wavelet transform and SURF algorithm[C]//MIPPR 2019: automatic Target Recognition and Navigation. SPIEL 2020;11429:258–66.
[32] Kaneko AM, Yamamoto K. Monocular height estimation by chronological correction of road unevenness[C]//2016 IEEE/SICE International Symposium on System Integration (SII). IEEE; 2016. p. 31–8.
[33] Kaneko AM, Yamamoto K. Two-view monocular depth estimation by optic-flow-weighted fusion[J]. IEEE Rob Autom Lett 2019;4(2):830–7.
[34] Jung S, Cho Y, Kim D, et al. Moving object detection from moving camera image sequences using an inertial measurement unit sensor[J]. Appl Sci 2020;10(1):268.
[35] Zhou Z, Shi Y, Gao Z, et al. Wildfire smoke detection based on local extremal region segmentation and surveillance[J]. Fire Saf J 2016;85:50–8.
[36] Zulkeflee AN, Yussof WNJW, Umar R, et al. Detection of a new crescent moon using the Maximally Stable Extremal Regions (MSER) technique[J]. Astronomy and Computing 2022:100651.
[37] Yi X, Song G, Derong T, et al. Fast road obstacle detection method based on maximally stable extremal regions[J]. Int J Adv Rob Syst 2018;15(1): 1729881418759118.
[38] Xu Y, Gao S, Li S, et al. Vision-IMU based obstacle detection method[C]// International conference on green intelligent transportation system and safety. Singapore: Springer; 2017. p. 475–87.
[39] Chen B, Yang X. Small obstacles image detection and classification for driver assistance[J]. Multimedia Tools and Applications; 2022. p. 1–11.
[40] Tsai DM, Lin CT. Fast normalized cross correlation for defect detection[J]. Pattern Recogn Lett 2003;24(15):2625–31.
[41] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. 2018. arXiv preprint arXiv:1804.02767.
[42] Ju M, Luo H, Wang Z, et al. The application of improved YOLO V3 in multi-scale target detection[J]. Appl Sci 2019;9(18):3775.
[43] Li J, Gu J, Huang Z, et al. Application research of improved YOLO V3 algorithm in PCB electronic component detection[J]. Appl Sci 2019;9(18):3750.
[44] Lv X, Dai C, Chen L, et al. A robust real-time detecting and tracking framework for multiple kinds of unmarked object[J]. Sensors 2020;20(1):2.
[45] Han G, Su J, Zhang C. A method based on multi-convolution layers joint and generative adversarial networks for vehicle detection[J]. TIIS 2019;13(4): 1795–811.
[46] Kim KJ, Kim PK, Chung YS, et al. Multi-scale detector for accurate vehicle detection in traffic surveillance data[J]. IEEE Access 2019;7:78311–9.
[47] Andrade DC, Bueno F, Franco FR, et al. A novel strategy for road lane detection and tracking based on a vehicle's forward monocular camera[J]. IEEE Trans Intell Transport Syst 2018;20(4):1497–507.
[48] Liu K, Wang W, Tharmarasa R, et al. Dynamic vehicle detection with sparse point clouds based on PE-CPD[J]. IEEE Trans Intell Transport Syst 2018;20(5):1964–77.
[49] Panev S, Vicente F, De la Torre F, et al. Road curb detection and localization with monocular forward-view vehicle camera[J]. IEEE Trans Intell Transport Syst 2018; 20(9):3568–84.
[50] García-García L, Jiménez JM, Abdullah MTA, et al. Wireless technologies for IoT in smart cities[J]. Netw Protoc Algorithm 2018;10(1):23–64.
[51] Wu J, Kong Q, Yang K, et al. Research on the steering torque control for intelligent vehicles co-driving with the penalty factor of human-machine intervention[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2022.
[52] Wu J, Zhang J, Nie B, et al. Adaptive control of PMSM servo system for steering-by-wire system with disturbances observation[J]. IEEE Transactions on Transportation Electrification 2021;8(2):2015–28.
[53] Wu J, Zhang J, Tian Y, et al. A novel adaptive steering torque control approach for human–machine cooperation autonomous vehicles[J]. IEEE Transactions on Transportation Electrification 2021;7(4):2516–29.
[54] Wu J, Tian Y, Walker P, et al. Attenuation reference model based adaptive speed control tactic for automatic steering system[J]. Mech Syst Signal Process 2021;156: 107631.