# A Wide Range Multi-obstacle Detection Method Based on VIDAR and Active Binocular Vision

Ruoyu Zhu, Yi Xu, Liming Wang, Teng Sun, Jinxin Yu, Shaohong Ding, Yuqiong Wang, Dong Guo, Pengwei Wang, Bingzheng Liu

***Abstract*—Binocular vision-based obstacle detection system has a wide range of applications, such as intelligent vehicles, mobile robots and other fields. However, the system has a large blind area because of the fixed camera pose of the traditional binocular vision system. This limits the system to accurately obtain the depth information of obstacles only in the field of view where two cameras overlap, and it is not possible to obtain accurate information otherwise. Aiming at this problem, we propose a wide range multi-target active binocular vision obstacle detection method based on VIDAR (Vision-IMU based detection and range method) for intelligent vehicles. First, two separate cameras are used to detect the three-dimensional obstacles by VIDAR. Second, the optical axis angle of the camera is adjusted in real-time according to the position of the obstacle. This causes the target to be detected such that it always lies in the overlapping field of view of the two cameras, and the obstacle can be actively tracked. Third, the obstacle search strategy is proposed to maximize the number of obstacles detected by the active binocular system. Last, experimental results show that compared with the traditional obstacle detection method based on binocular vision, the proposed method can not only detect unknown types of obstacles, but also have a wider detection range and a guaranteed ranging accuracy.**

***Index Terms*—Active vision, Binocular vision, Morphological operation, VIDAR**

Manuscript received June 26, 2022; revised January 7, 2023.

This work was supported in part by the National Natural Science Foundation of China under Grant 51905320 and 52102465, China Postdoctoral Science Foundation under Grant 2018M642684, Natural Science Foundation of Shandong Province under Grant ZR202102180180 and ZR2021QF039, Shandong Province Major Science and Technology Innovation Project under Grant 2019JZZY010911, Shandong Key R&D Plan Project under Grant 2019GGX104066, and Experiment technology upgrading project under Grant 2022003.

Ruoyu Zhu is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 1339315089@qq.com).

Yi Xu is an associate professor at the Collaborative Innovation Center of New Energy Automotive, Shandong University of Technology, Zibo 255000, China. (corresponding author, e-mail: xuyisdut@163.com).

Liming Wang is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 2463185675@qq.com).

Teng Sun is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 877992783@qq.com).

Jinxin Yu is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 2889717886@qq.com).

Shaohong Ding is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 1846336356@qq.com).

Yuqiong Wang is a PhD student at Shandong University of Technology, Zibo 255000, China.( e-mail: 858151127@qq.com).

Dong Guo is a professor at Shandong University of Technology, Zibo 255000, China. (e-mail: jiaoyun547@163.com).

Pengwei Wang is a lecturer at Shandong University of Technology, Zibo 255000, China. (e-mail: wpwk16@163.com).

Bingzheng Liu is a lecturer at Shandong University of Technology, Zibo 255000, China. (e-mail: lbzheng528@126.com).

## I. INTRODUCTION

Binocular stereo vision is widely used in many fields. For example, a binocular vision system fixed on a welding robot was used by Chen et al. to compute the position information of a spatial seam with an error of less than 0.3 mm [1]. Ling et al. proposed a dual-arm cooperative approach for a tomato harvesting robot using a binocular vision sensor, which could carry out harvesting in an unstructured environment [2]. Jiang et al. proposed a gesture recognition system based on binocular vision using the Boyer-Moore (BM) algorithm for stereo matching [3].

Many researchers have been working on improving the performance of binocular vision systems to obtain higher detection and measurement accuracy. Zhai et al. proposed an accurate stereo matching method to locate the three-dimensional (3D) position of crop rows. This research was based on the rank transformation, Harris detector and random sample consensus methods to reduce the computational complexity and improve the accuracy of image stereo matching [4]. Tippetts et al. provided an example of modifying an existing highly accurate stereo vision algorithm to increase its runtime performance while trying to limit the loss in accuracy [5].

The traditional binocular vision system consists of two cameras. The camera model can be described by a pinhole model when the lens distortion is either corrected or can be ignored without timing [6]. In the traditional binocular vision system, the relative attitude is fixed between two cameras. This feature limits the public field of view and its flexibility in practical application, and also causes inconvenience with respect to maintenance.

Two images taken by a single camera from two viewpoints can also form a binocular stereo vision; however, the relative attitude of the camera from one viewpoint to another is not accurate, which seriously affects the measurement accuracy. In a binocular detection method, different baseline widths have different detection ranges and accuracy values. Previous studies have shown that on the premise of ensuring an effective binocular field, a wide baseline system has a higher detection accuracy than a narrow baseline system [7]-[17]. However, the growth of the baseline in the wide baseline system increases the blind area of the system. Although the research on binocular vision detection method is relatively mature, it still cannot effectively solve the problem of blind area in the binocular vision system.

Biological vision has always been an important source of inspiration for computer vision algorithm design [18]-[22].

Therefore, this paper applies the feature of flexible rotation in the horizontal direction of human eyeballs to the binocular detection system and builds a non-parallel binocular visual model [23]-[25]. The VIDAR is used to initially detect 3D obstacles [26], and subsequently, the position information of the obstacles is used to calculate the camera angle. In this way, the camera optical axis angle can be adjusted in real-time, and the blind area in the traditional binocular vision can be avoided. Consequently, the detection target always lies in the overlapping field of the two cameras. In addition to the aforementioned model, the active optical axis search strategy of the binocular camera is proposed to realize the optimal detection of multiple obstacles.

The rest of the article is structured as follows. Section 2 introduces the VIDAR obstacle detection principle and the binocular vision model. Section 3 describes the active obstacle detection method of binocular camera based on unmatched regions, and the active obstacle tracking method of binocular vision based on matched regions, and presents the design of the corresponding camera rotation algorithm. In Section 4, the simulation and real vehicle experiments are presented, and the experimental results are compared with those from the traditional binocular vision and active binocular vision based on YOLO V5S. The comparison verifies the superiority of the detection accuracy and range of the proposed method. Section 5 concludes this paper.

## II. Principle of VIDAR and binocular vision model

### *A. Principle of VIDAR*

In this paper, VIDAR is used to initially detect the 3D obstacles in front of the vehicle. Based on monocular vision, VIDAR is an effective obstacle detection algorithm with a simple operation and a high detection accuracy. It can accurately obtain the location information of target obstacles and detect unknown obstacles that cannot be identified by machine learning. The research methods developed in this paper are primarily based on VIDAR.

During the process of obstacle detection by VIDAR, the feature points are extracted using the fast image region matching method based on maximally stable extremal regions (MSER), and two frames of images are matched [27]. On this basis, the discrimination principle of VIDAR 3D obstacles is used to eliminate non-obstacle points extracted via the MSER image region matching method, and the obstacles in the image can be detected directly and quickly.

*Calculation of the horizontal distance of obstacles*

The lowest point of MSER connected to the measured area is regarded as the intersection *P* between the obstacle and the road plane. The horizontal distance between the point *P* and the camera can be obtained as follows:

For the convenience of calculation, it is assumed that the optical axis of the camera points at the point *P* exactly, as shown in Fig. 1. The effective focal length of the camera is denoted by *f*, the optical axis height between the camera lens and the ground is represented by *h*, the pixel size is given by $\mu$, and the pitching angle of the camera is equal to $\partial$. The coordinate origin of the plane coordinate system $(x_0, y_0)$, i.e., the intersection between the image plane and the camera optical axis is usually set to (0,0). Given that the coordinates of the intersection point of the obstacle in front and the road plane in the imaging plane are denoted by (*x*,*y*), the horizontal distance *d* between the intersection point and the camera can be obtained from (1).

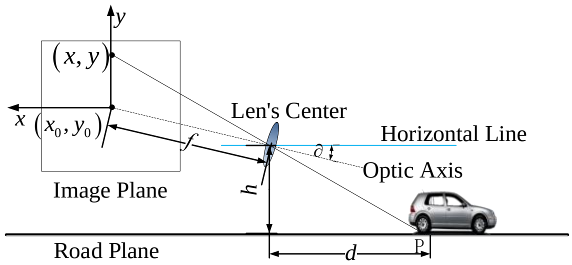



Fig. 1. Schematic diagram of 3D obstacle pinhole imaging.

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

*Discrimination of 3D obstacles*

As Fig. 2 shows, the first imaging point of the obstacle is *A*, the *y* axis is moved from $y_1$ to $y_2$ in the image plane because of the camera's movement, and the imaging point of the obstacle's top is *B*. Suppose there is a two-dimensional (2D) obstacle on the road plane, and *A′* and *B′* are the points corresponding to *A* and *B* on the road plane, respectively. The horizontal distance between the camera to point *A′* and point *B′* are $d_1$ and $d_2$, respectively.

The distances $d_1$ and $d_2$ can be obtained using (1). The camera moves a certain distance $\Delta d$ between two consecutive images, and this distance can be obtained using the IMU. The relationship between $d_1$ and $d_2$ can be expressed as $d_1 = d_2 + \Delta d$ . Actually, the obstacle is 3D and, therefore, $d_1 = d_2 + \Delta d + \Delta l$ . Thus, the target point is not on the road if $d_1 \neq d_2 + \Delta d$ , which can be used as the basis for distinguishing 3D obstacles.

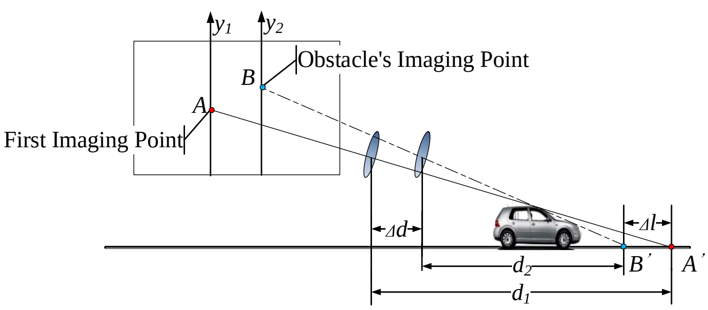



Fig. 2. Schematic diagram of stationary obstacle imaging.

In addition, the movement of an obstacle can also be used for obstacle determination in most environments. Relevant parameter descriptions and certification processes are described in [26] and [28].

*B. Binocular vision model*

In the broad sense, binocular vision can be divided into parallel and non-parallel binocular vision. The binocular vision imaging model is also based on the pinhole imaging model. The 3D spatial coordinates of the points in the world coordinate system can be calculated according to the camera calibration parameters and parallax values of the left and right projection points.

*Parallel binocular vision model*

In the research method presented in this paper, the system can be regarded as a standard parallel binocular vision system at the initial moment, while slight errors are ignored. This scenario is shown in Fig. 3. The system assumes that the left and right cameras (denoted as cameras *l* and *r*, respectively) are in the same horizontal position, the optical axis is parallel and intersect vertically with the imaging plane, and the same object is being captured from different positions to obtain two images with a specific parallax.

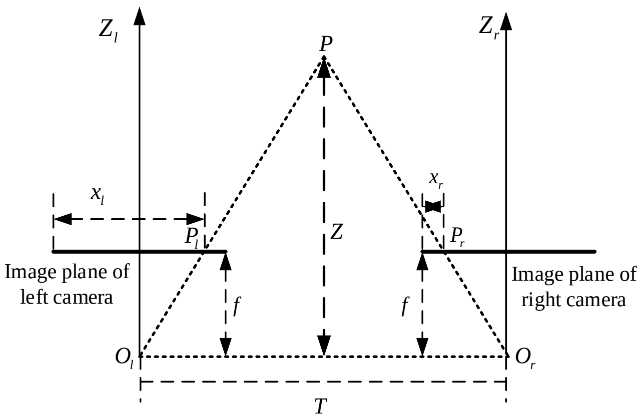



Fig. 3. Parallel binocular vision model.

The point $P$ is a certain target point in space, $p_l$ and $p_r$ are the pixels projected by $P$ on the imaging plane, and $O_l$ and $O_r$ are the optical centers of the cameras *l* and *r*, respectively. The focal length of the camera is *f*, *T* is the baseline, $x_l$ and $x_r$ are the coordinates of $p_l$ and $p_r$ in the pixel coordinate system, respectively, and *Z* is the distance between *P* and the camera's light center in the world coordinate system. The depth value *Z* of *P* can be calculated according to the triangle similarity principle. The corresponding calculation formula is as follows:

$$\frac{T-(x_l-x_r)}{Z-f}=\frac{T}{Z}\Rightarrow Z=\frac{fT}{x_l-x_r} \tag{2}$$

*Non-parallel binocular vision model*

Figure 4 shows that the two cameras have offsets in the *X*-axis, *Y*-axis and *Z*-axis directions, and the optical axis of camera *r* rotates by an angle θ in the counterclockwise direction around the *Y*-axis. Similarly, with camera *l* coordinate system as the world coordinate system, the 3D coordinates of *P* can be calculated by the coordinates of projection point, rotation matrix and translation vector.

In the active vision system studied in this paper, it is assumed that the optical axes of two cameras are coplanar and the visual frames coincide. The optical centers of cameras *l* and *r* are represented by the origins $O_l$ and $O_r$, respectively. In the initial state, *Z* denotes the optical axis direction towards the scene. The *X*-axis and *Y*-axis are the directions corresponding to the horizontal and vertical axes of the image, respectively. The camera rotates actively around the *Y*-axis when the active vision is triggered. Therefore, the non-parallel binocular vision model proposed in this paper only considers the rotation angle of the optical axis, and the other offsets are not considered.

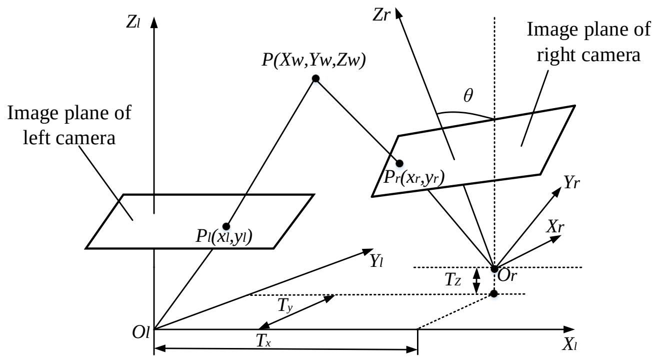



Fig. 4. Non-parallel binocular vision model.

## III. Multi-obstacle detection and tracking based on VIDAR and active binocular vision

*A. Overview of obstacle detection and tracking based on unmatched region*

In this paper, first, the cameras *l* and *r* are used to detect 3D obstacles in front of the vehicle based on VIDAR. Second, feature points are matched between the two images obtained by the cameras *l* and *r* to determine the area where the obstacle is located. Last, the camera angle is calculated, and the camera optical axis is rotated so that most or even all obstacles are in the overlapping field of view of the two cameras. Subsequently, the binocular distance is measured.

After feature point matching, the possible regions of obstacles can be divided into the following five scenarios, as shown in Fig. 5: (1) All obstacles are in the matching region; (2) All obstacles are in the unmatched region on one side; (3) The obstacles are in both the matching and unmatched regions on one side; (4) Obstacles are located in the unmatched and matching regions on both sides; (5) All obstacles are located in the unmatched regions on both sides. The obstacle can be directly tracked by the binocular distance measurement in scenario (1), whereas in scenarios (2) and (3), all the obstacles are in the field of view of one camera and only one camera needs to rotate to construct the binocular vision system. In scenarios (4) and (5), the situation is complicated as all the obstacles are scattered in the field of view of the cameras on both sides. Therefore, it is necessary to successively rotate the cameras on both sides to construct the binocular vision system.

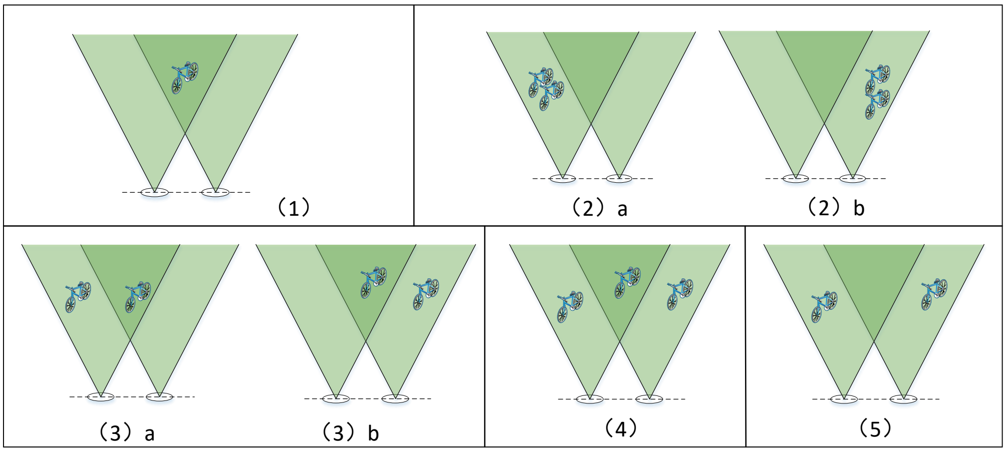



Fig. 5. Classification diagram of the regions where obstacles are located.

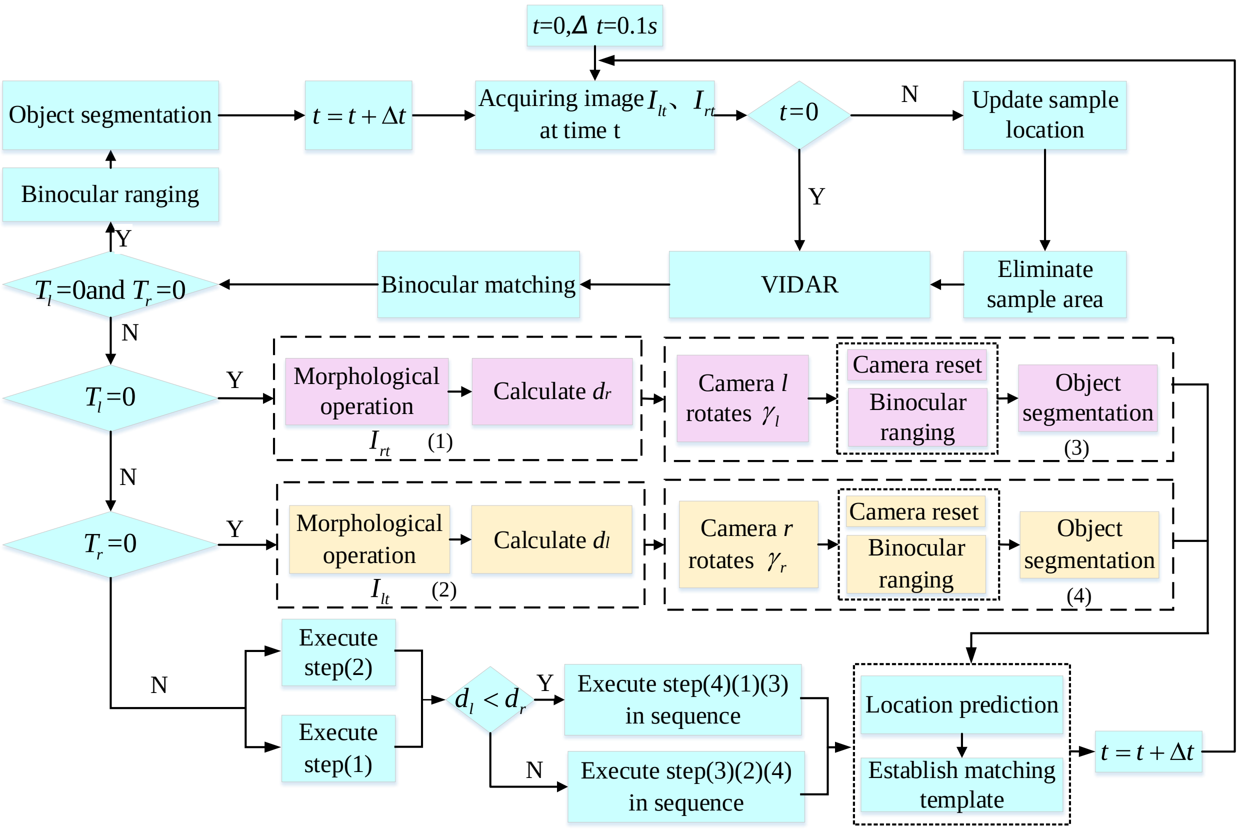



Fig. 6. Flow chart of obstacle detection and tracking.

Figure 6 shows the flow chart of obstacle detection and tracking, which is described as follows:

(1) When $t$=0, cameras $l$ and $r$ collect the environmental information to obtain images $I_{lt}$ and $I_{rt}$, respectively. The VIDAR is used to effectively discriminate between all feature points and initially detect the 3D obstacles.

(2) The obstacle feature points are matched in images $I_{lt}$ and $I_{rt}$ to identify the region where the obstacle is located.

(3) Let $N_l$ and $N_r$ be the number of unmatched points in $I_{lt}$ and $I_{rt}$, respectively, and $N_{lr}$ be the number of matching points in $I_{lt}$ and $I_{rt}$.

If $N_l = 0$ and $N_r$=0, it indicates that all obstacles are within the matching region, and the binocular distance can be measured directly.

If $N_l = 0$ and $N_r \neq 0$, it indicates that the region where the obstacle is located is given by scenarios (2)b and (3)b, as shown in Fig. 5. Morphological operation is performed on the feature points in $I_{rt}$ to obtain the distance of the lowest point of the obstacle $d_r$. In addition, the optical axis angle $\gamma_l$ of camera $l$ is obtained using the camera rotation algorithm.

If $N_l \neq 0$ and $N_r$=0, it indicates that the region where the obstacle is located is given by scenarios (2)a and (3)a as shown in Fig. 5. Morphological operation is performed on the feature points in $I_{lt}$ to obtain the distance of the lowest point of the obstacle $d_l$. Furthermore, the optical axis angle $\gamma_r$ of camera $r$ is obtained using the camera rotation algorithm.

If $N_l \neq 0$ and $N_r \neq 0$, it indicates that scenarios (4) and (5) represent the region where the obstacle is located, as shown in

Fig. 5. Morphological operation is performed on the feature points in $I_{lt}$ and $I_{rt}$ to obtain the distance of the lowest point of the obstacles $d_l$ and $d_r$, respectively. The camera is rotated in turn according to the size relationship between $d_l$ and $d_r$, and the camera rotation algorithm.

(4) After rotating the optical axis of the camera, the obstacle is measured by the binocular ranging and the camera is reset at the same time. The obstacle target is segmented and the points with known distance are extracted as samples to generate the matching template.

(5) As the vehicle moves, the identification and active detection of obstacles enter the next frame. When $t = t + \Delta t$, $I_{lt}$ and $I_{rt}$ are compared with the template to update the sample location. Subsequently, the sample area is eliminated so that the repeated ranging does not affect the detection speed. Steps (2) and (3) shown in Fig. 6 are repeated to actively track samples and detect new obstacles.

*B. Active binocular vision obstacle search strategy based on unmatched region*

*Active binocular obstacle search strategy in scenarios (2) and (3)*

Figure 5 shows that when the region of the obstacle is located according to scenario (3), the distance of the obstacle in the binocular matching region can be accurately measured by the binocular vision system without any camera rotation. On the other hand, the distance of the obstacle in the unilateral unmatched region should be measured by the camera rotation. As scenario (3) is consistent with the search strategy for scenario (2), we shall discuss them together.

When the region of the obstacle is located according to scenarios a or b shown in Fig. 5, the obstacle search strategy and camera rotation algorithm in the two cases are completely identical except for the different rotation directions of the camera optical axis. Therefore, this paper only describes one of the scenarios in detail.

Considering scenario (2)a as an example, the specific camera rotation algorithm is as follows:

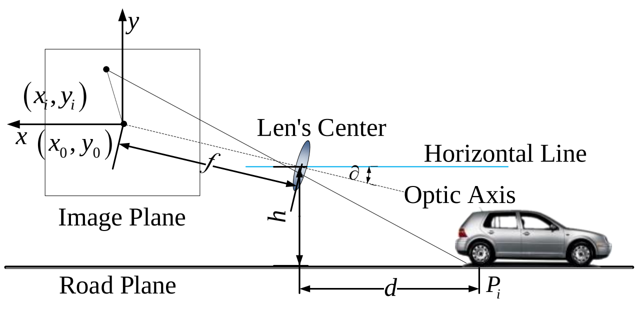



Fig. 7. Schematic diagram of point $P_i$ pinhole imaging.

The camera $l$ first performs a morphological operation on the feature points in image $I_{lt}$ after detecting the obstacle. The lowest point after the morphological operation is regarded as the intersection point $P_i$ between the obstacle and the road plane, where $i$ is the number of point $P$. Figure 7 shows the projection of $P_i$ on the camera $l$ imaging plane, where $(x_i, y_i)$ are the coordinates of $P_i$ in the imaging plane. The distance between the camera $l$ and $P_i$ can be calculated using (1). The horizontal distance between $P_i$ and the optical axis of camera $l$ is denoted by $w_i$, which can be obtained based on the pinhole imaging principle and trigonometry given by (3).

$$w_i = \frac{d_i\left(x_0 - x\right)\mu}{f} = \frac{h\left(x_0 - x\right)\mu}{\tan\left\{\partial + \arctan\left[\left(y_0 - y\right)\cdot \mu / f\right]\right\}} \quad (3)$$

When $i = 1$, the position relationship between the obstacle and the cameras is shown in Fig. 8(a), where $\alpha$ is the field angle of the two cameras. At this time, only the field of view of camera $r$ contains obstacles, therefore, the distance to the obstacles cannot be obtained by the binocular vision.

Let $d' = d_1$ and $w' = w_1$. Subsequently, the rotation angle $\beta_r$ of the optical axis of camera $r$ is

$$\beta_r = 90° - \arctan\frac{d'}{\left(w' + T\right)} \quad (4)$$

Where $T$ is the distance between the optical centers of the two cameras, and $d'$ and $w'$ are the intermediate variables used for calculating the camera angle. The camera $r$ is actively rotated by an angle $\beta_r$ to align its optical axis with point $P_1$ so that the field of view of camera $r$ also contains the obstacle. Figure 8(b) shows the ideal camera rotation results.

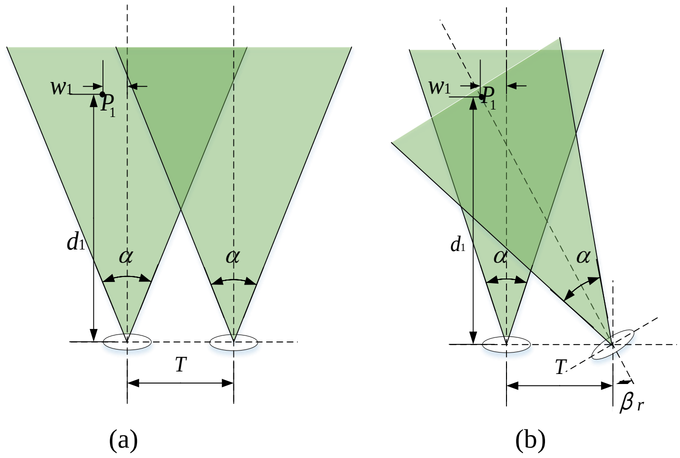



Fig. 8. Schematic diagram of camera before and after rotation when $i$=1.

Figure 9 shows the detection process diagram of the active binocular system based on VIDAR for $i$ = 2.

Generally, there are multiple obstacles on the road. Therefore, it is necessary to accurately rotate the camera so as to maximize the number of obstacles in the overlapping field of view of the two cameras, and obtain an optimal obstacle detection performance. When $i \geq 2$, the middle position of the area where the obstacle is placed is selected to locate the optical axis of the camera $r$. At this point, the values of $d'$ and $w'$ are shown by (5) and (6), respectively. The rotation angle of camera $r$ is still calculated using (4).

$$d' = \frac{\max(d_i) - \min(d_i)}{2} \quad (5)$$

$$w' = \frac{\max(w_i) - \min(w_i)}{2} \quad (6)$$

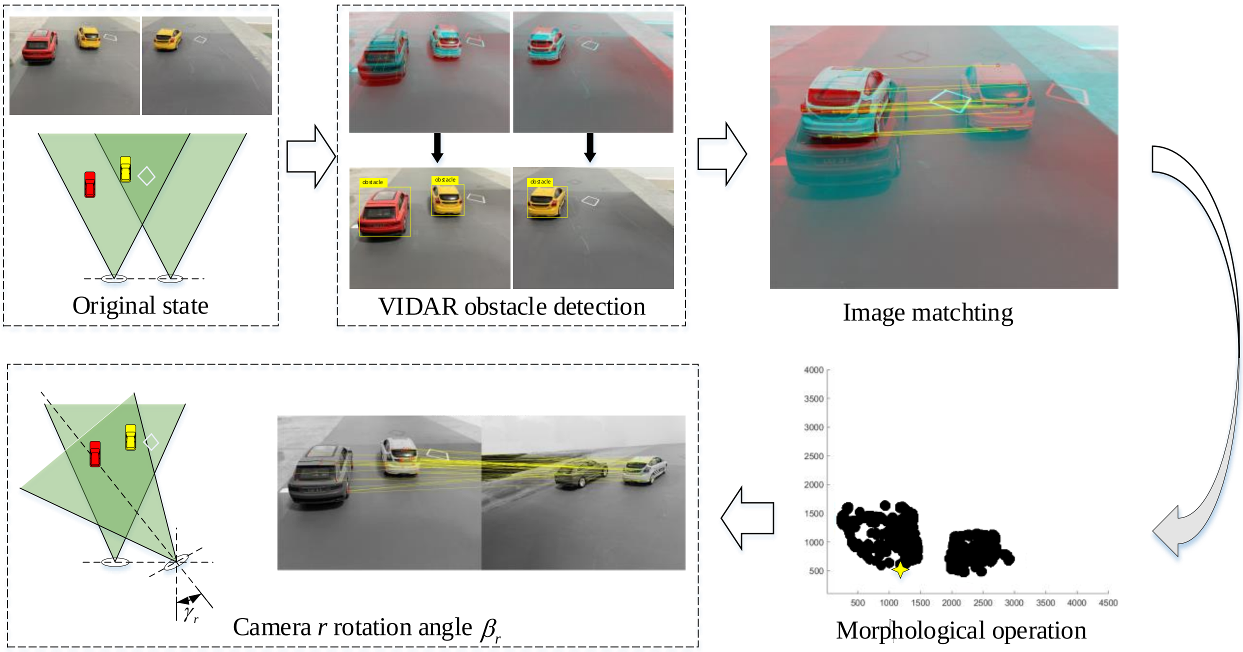
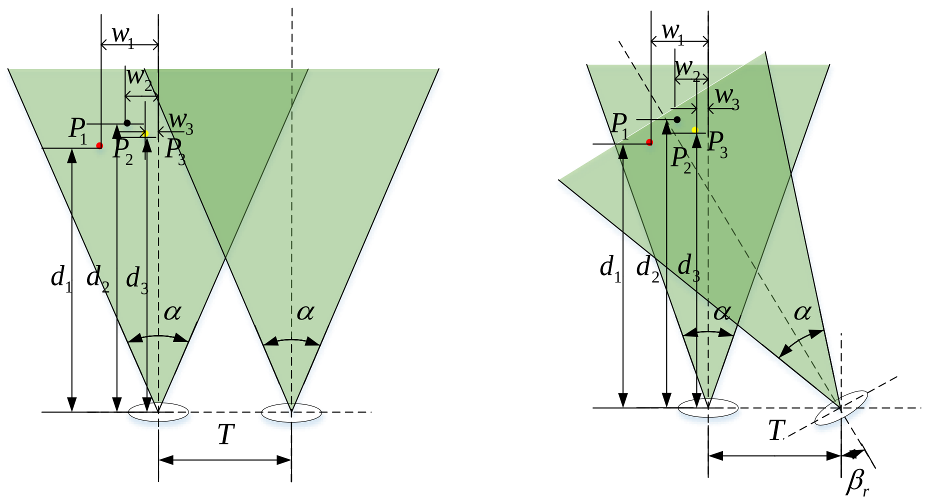



Fig. 9. Detection process diagram of active binocular system based on VIDAR when *i*=2.




(a) (b)

Fig. 10. Schematic diagram of camera before and after rotation when *i*=3.

Figure 10 shows the ideal rotation diagram of the camera *r* when *i* = 3. It can be observed that after the camera angle adjustment, the lowest points $P_1$, $P_2$, $P_3$ of the three obstacles are all in the overlapping area of the fields of view of the two cameras. Thus, the system can measure the binocular distance.

*Active binocular obstacle search strategy in scenarios (4) and (5)*

If all the obstacles are scattered in the field of view of the cameras on both sides, the two cameras should rotate respectively. Considering scenario (5) as an example and assuming *i* = 2, the position relationship between the obstacles and the camera is shown in Fig. 11. In the figure, the two parts of the obstacles are located in the separate fields of view of cameras *l* and *r*, respectively. The intersection points between the lowest point of the obstacle and the road plane are denoted by $P_1$ and $P_2$, respectively, and the horizontal distances between these points and the camera are represented by $d_1$ and $d_2$, respectively. When $d_1 < d_2$, the rotation angle of camera *r* is calculated according to (4) considering $d' = d_1$, and the camera *r* is rotated according to this angle. After the camera *r* is reset, it is considered that $d' = d_2$ so that the camera *l* can rotate and its rotation angle of camera *l* can be obtained.

However, as the intermediate variable $d'$ used to calculate the rotation angle is obtained from the single visual distance, there will be a small error with respect to the real value. Furthermore, in the presence of multiple obstacles, the detection performance may not be optimal because the camera angle is calculated from the middle position of the area where the obstacles are located. Therefore, the camera angle is adjusted by rotating the camera by 10% of $\alpha$ to the left and right to ensure that the overlapping area of the fields of view of the two cameras contain all obstacles to the maximum extent. For example, assume that the number of obstacles in the matching area after the camera *r* rotates by $\beta_r$ is m, and the numbers of obstacles in the matching area after the camera $r$ rotates by 10% of $\alpha$ to the left and right are $n$ and $n'$, respectively. If $\begin{cases} m \geq n \\ and \\ m \geq n' \end{cases}$, the optical axis rotation of camera is kept as $\beta_r$, and if $\begin{cases} m < n \\ or \\ m < n' \end{cases}$, the rotation angle of camera $r$ is finally determined as $\gamma_r$.

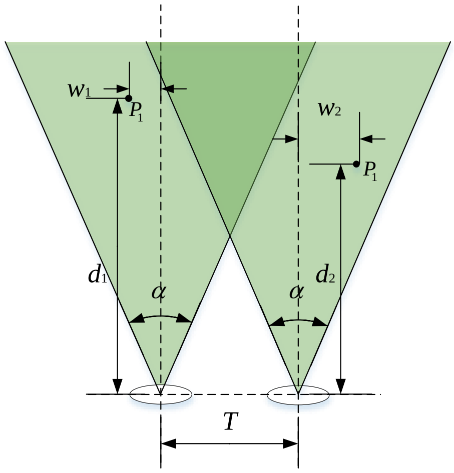



Fig. 11. Diagram of the position relationship between the obstacle and the camera in scenario (5).

After the camera rotates, the binocular vision system can simultaneously find the distance to the obstacles and reset the camera to ensure that the binocular vision system has a larger field of vision. Subsequently, the depth information is used to segment the obstacle target. The obstacles with known distances are considered as samples, and the Kalman filter is used to track them, predict their positions, and establish matching templates. When $t = t + \Delta t$, the target is located in a new frame of image and the sample position is updated. The effectiveness of the system can be ensured by eliminating the ranging area, i.e., the sample area. This elimination is used to judge whether there is a new 3D obstacle in the unmatched area in the next frame, which is used to decide whether to rotate the camera or not. Steps (2) and (3) shown in Fig. 6 are repeated to update the matching template in order to actively track the samples and detect the new obstacles.

*C. Binocular vision active obstacle tracking based on unmatched region*

If all the obstacles are in the binocular matching region during the obstacle detection process, the binocular distance is directly measured for all the obstacles, and subsequently it is segmented. In the next frame, when no obstacle appears in the unmatched regions, the Kalman filter is used to continuously track the obstacle. During the obstacle tracking process, the distance between the obstacles and the camera reduces continuously. This phenomenon either reduces the number of obstacles in the matching area or even makes them disappear.

Therefore, during the tracking process, it is necessary to actively rotate the camera inward so that the overlapping area of the fields of view of cameras $l$ and $r$ can contain all the obstacles. It is assumed that there are multiple obstacles in the binocular matching area. The farther the obstacle is from the camera, the wider the overlapping range of the two cameras. In addition, for the consideration of driving safety, the autonomous vehicle can adopt the obstacle avoidance strategy in time. Therefore, the position of the lowest point $P_Q$ of the obstacle $Q$ closest to the autonomous vehicle in the matching area is considered as the basis of active tracking trigger.

As Fig. 12 shows, active tracking is started when the distance $d_Q$ between the vehicle and $P_Q$ is detected to be less than $a$ ($a > h$). Figure 13 shows the tracking flow chart.

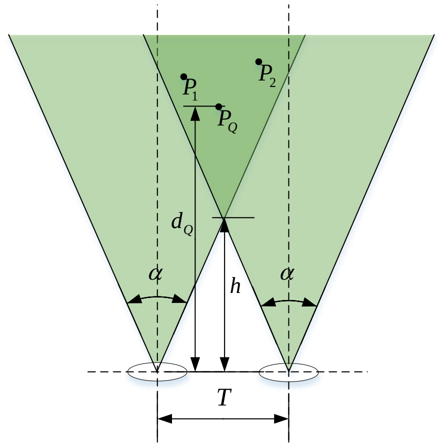



Fig. 12. Schematic diagram of active trace trigger condition.

According to Fig. 12, the conditions for obstacles in the overlapping field of view of two cameras are as follows:

$$\frac{d}{w_l} \geq \tan(\frac{\alpha}{2}) \text{and} \frac{d}{w_r} \geq \tan(\frac{\alpha}{2}) \tag{7}$$

Substituting (1) and (3) into (7) and after simplification, we obtain:

$$\frac{(x_{ol} - x_l)\mu}{f} \geq \tan(\frac{\alpha}{2}) \text{and} \frac{(x_{or} - x_r)}{f} \geq \tan(\frac{\alpha}{2}) \tag{8}$$

Where $\omega_l$ and $\omega_r$ represent the horizontal distances between the lowest point $P_Q$ of the obstacle and the optical axes of cameras $l$ and $r$, respectively. The abscissas of the origin in the image coordinate system of cameras $l$ and $r$ are denoted by $x_{ol}$ and $x_{or}$, respectively, and $x_l$ and $x_r$ represent the abscissas of the obstacle in the imaging planes of cameras $l$ and $r$, respectively.

The minimum rotation angles $\eta_l$ and $\eta_r$ of cameras $l$ and $r$, respectively, are calculated after the vehicle moves forward a distance $s$. This calculation is such that the overlapping area of the fields of view of the two cameras can contain all obstacles. Using (8), the camera angle is calculated based on the critical condition that the obstacle exists in the overlapping area of the two cameras' fields of view. The calculations are carried out according to (9) and (10) as follows:

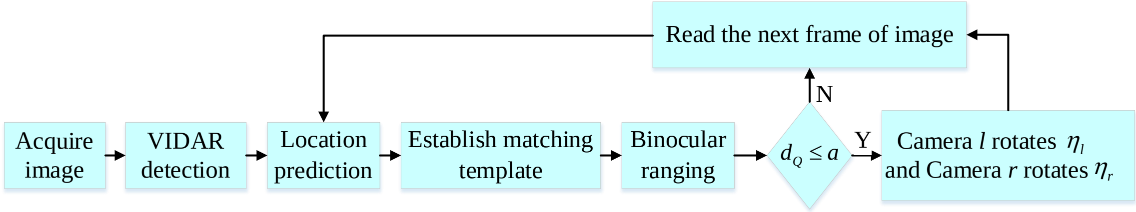



Fig. 13. Flow chart of active tracking.

$$\eta_l = 2\arctan(\frac{d}{w_l}) - \alpha \tag{9}$$

$$\eta_r = 2\arctan(\frac{d}{w_r}) - \alpha \tag{10}$$

Figure 14 shows the schematic diagram of binocular active tracking.

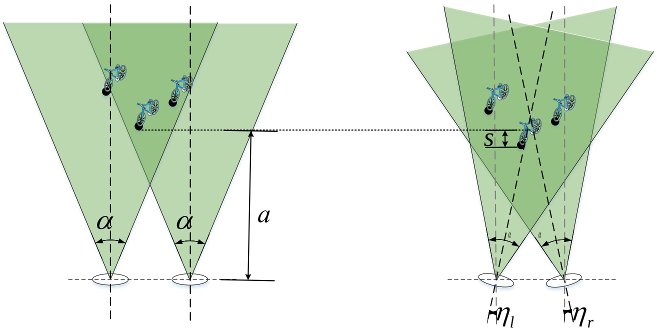



Fig. 14. Schematic diagram of binocular active tracking.

## IV. Experiments and Result Analysis

The experiments consist of two parts: simulation experiments in a controllable scene and real vehicle experiments. According to the experimental results, the proposed obstacle detection method can improve the detection range and reduce the miss ratio.

### A. Experiment preparation

In this paper, experimental equipment is used to carry out simulation experiments in controllable scenes to verify the obstacle detection and tracking performance using the VIDAR and active binocular vision. The experimental equipment mainly includes a pure electric vehicle, two Da Ying cameras, two DS3115MG steering engines, a computer, a CRS07-11 IMU and a STM32 microcontroller.

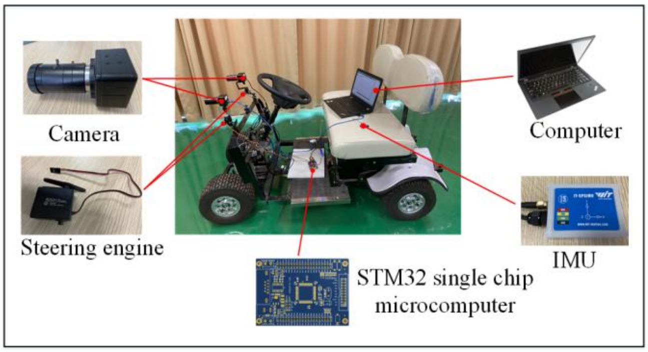



Fig. 15. Diagram of laboratory equipment.

As Fig. 15 shows, the cameras used to collect the front image are installed in the front of the vehicle according to the ideal binocular system standard. The installation heights and the baseline lengths of the cameras are 1 m and 60 mm, respectively. The IMU is installed at the bottom of the vehicle and can locate itself in real-time, and obtain the attitude, speed and displacement information of the vehicle to provide information for the VIDAR. The computer runs a series of codes such as image acquisition and image processing to obtain detection results, and uses the RS232 protocol to transmit the signal to the STM32 microcontroller. In this paper, the PWM control strategy is used, the STM32 single chip microcomputer is utilized as the control core, and the output voltage is controlled using the RS232 data analysis. The steering gear controls the camera rotation angle based on the output voltage.

It is necessary to calibrate the cameras in order to obtain their internal and external parameters, and subsequently obtain the conversion relationship between them. The camera calibration includes single target calibration and binocular stereo calibration, where the former is the basis of binocular calibration. In this paper, Zhengyou Zhang calibration method is used to calibrate the two cameras. Parts of the checkerboard calibration images are shown in Fig. 16, and the calibration results are shown in Fig. 17.

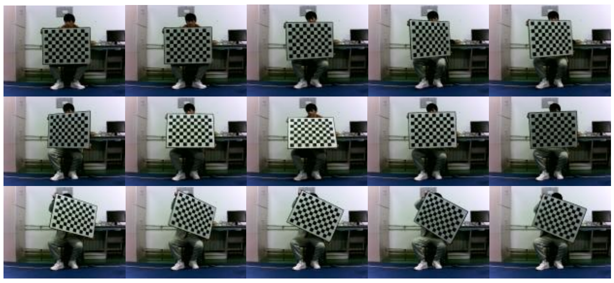


Fig. 16. Parts of the checkerboard calibration images.

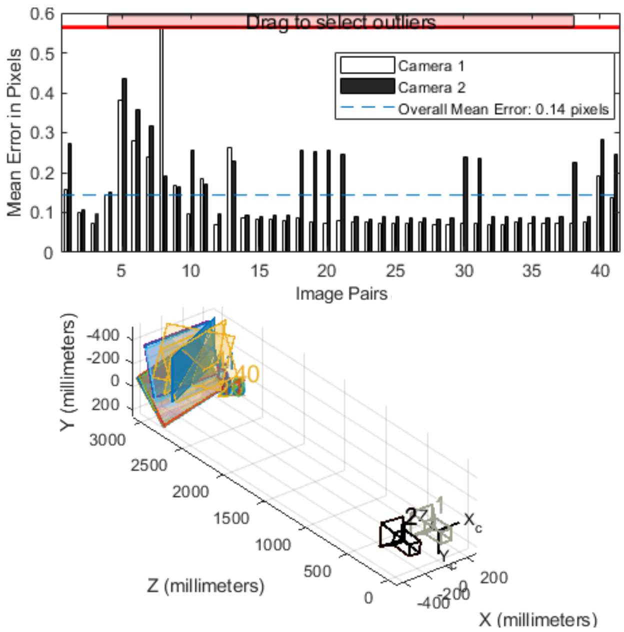



Fig. 17. Calibration results.

*B. Simulation static experiments*

*Verification of ranging accuracy and detection range*

A total of five obstacles are selected in this experiment, as shown in Fig. 18. The obstacles 1, 4 and 5 are known types of obstacles, 2 is a generalized obstacle, and 3 is a pseudo-obstacle without any height. Three groups of experiments are conducted respectively by placing obstacles in different positions and distances. Figure 18 shows the first group of experiments.

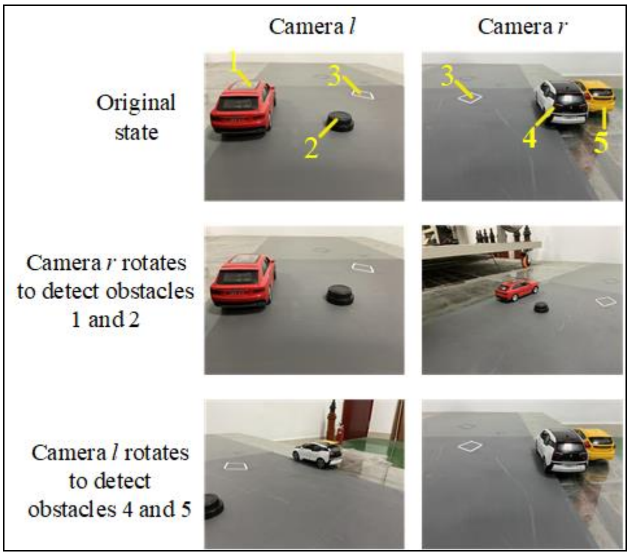



Fig. 18. Images of the first group of static experiments.

Figure 18 shows that the field of view of camera *l* contains obstacles 1, 2 and 3, and that of camera *r* contains obstacles 3, 4 and 5 in the initial state. At this time, only the pseudo-obstacle 3 is in the overlapping area of the fields of view of cameras *l* and *r*. The system cannot carry out binocular distance measurement for the other 3D obstacles. After the camera *r* rotates to the left according to the camera rotation algorithm, obstacles 1 and 2 also appear in its field of view. Subsequently, the camera *l* rotates to the right based on the camera rotation algorithm, and obstacles 4 and 5 also appear in its field of view. Therefore, utilizing the rotation of cameras on both sides, the distances to obstacles 1, 2, 3 and 5 can be measured by the binocular distance, which increases the detection range.

Table 1 shows the results of the three groups of static experiments. The serial number 1-1 represents obstacle 1 in the first group of experiments, and the actual value of the obstacle distance is measured manually. In the three groups of experiments, the experimental vehicle is in the same position and kept stationary. In the second and third groups of experiments, the distances between obstacles 1, 2 and the experimental vehicle remain constant in the *Z*-axis direction, and the distances between obstacles 4, 5 and the experimental vehicle change in both the *X*-axis and *Z*-axis directions.

Consider the first group as an example. Initially, as the overlapping area of the fields of view of cameras *l* and *r* only contains obstacle 3, the distance of other obstacles cannot be obtained by the traditional binocular system and is replaced by "/". However, the active binocular system proposed in this paper has the advantage of eliminating false obstacles. Therefore, obstacle 3 will not be within the range of the active binocular system. In this case, "/" is also used. Table 1 shows that the ranging accuracy of the active binocular system proposed in this paper has no large error compared with the traditional binocular ranging. This is because the relative posture of the camera from one viewpoint to another is known, which can ensure the ranging accuracy. In addition, the proposed method significantly improves the detection range owing to the rotation of the cameras, and can detect obstacles only existing in the field of view of monocular camera by stereo vision.

TABLE I.
RESULTS OF THREE GROUPS OF EXPERIMENTS

| Serial number | Actual value (cm) | Traditional binocular vision (cm) | Error (%) | Proposed method (cm) | Error (%) |
|---|---|---|---|---|---|
| 1-1 | 80.98 | / | / | 82.05 | 1.32 |
| 1-2 | 68.71 | / | / | 69.14 | 0.63 |
| 1-3 | 79.65 | 80.12 | 0.59 | / | / |
| 1-4 | 87.18 | / | / | 87.74 | 0.64 |
| 1-5 | 93.45 | / | / | 94.64 | 1.27 |
| 2-1 | 80.98 | / | / | 82.14 | 1.41 |
| 2-2 | 68.71 | / | / | 69.23 | 0.76 |
| 2-3 | 79.65 | 80.12 | 0.59 | / | / |
| 2-4 | 67.14 | / | / | 67.73 | 0.88 |
| 2-5 | 63.65 | / | / | 64.11 | 0.72 |
| 3-1 | 80.98 | 80.76 | 0.27 | 81.65 | 0.83 |
| 3-2 | 68.71 | 69.09 | 0.55 | 69.11 | 0.58 |
| 3-3 | 79.65 | 80.12 | 0.59 | / | / |
| 3-4 | 66.53 | 66.24 | 0.44 | 66.96 | 0.65 |
| 3-5 | 73.45 | / | / | 74.02 | 0.78 |

*Analysis of distance-measuring error*

Table 1 shows that on the one hand, the distance-measurement error increases with the increase of obstacle depth. On the other hand, in the three groups of experiments, the distances of obstacles 1 and 2 remain unchanged with respect to the experimental vehicle, but the experimental error of the second group is significantly higher than that of the other two groups. This is because the camera rotation angle of the second group is larger than those of the other two groups, and it is difficult to ensure that the rotation axis of the camera passes through the projection center or optical axis of the camera. In other words, the rotation axis of the camera deviates from its optical axis, causing errors in the baseline *T* that ultimately leads to ranging errors. In addition, in the three groups of experiments, the measured distance obtained by the proposed method is larger than the actual distance, which is due to the larger baseline *T* caused by the camera's inward rotation.

*C. Simulation dynamic experiments*

The experiments select a section of 100 meters of flat road in the laboratory building No. 6 of Shandong University of Technology, China. The locations of the obstacles are fixed. The total number of known types of obstacles is 100, and the total number of generalized obstacles is 10. The experimental vehicle moves at a speed of 2 m/s. The research method proposed in this paper considers the detection of obstacles by the active binocular vision system as the standard, and compares the detection results of the four detection methods, which are shown in Table 2. A comparison of the detection results of the four methods shows that the stability of the

TABLE II.
RESULTS OF THREE GROUPS OF EXPERIMENTS

| Detection method | Number of known types of obstacles | Number of generalized obstacles | TP | TN | FP | FN |
|---|---|---|---|---|---|---|
| Traditional binocular vision based on YOLO v5s | 100 | 10 | 83 | 4 | 6 | 17 |
| Traditional binocular vision based on VIDAR | 100 | 10 | 90 | 8 | 2 | 10 |
| Active binocular vision based on YOLO v5s | 100 | 10 | 88 | 3 | 7 | 12 |
| Active binocular vision based on VIDAR | 100 | 10 | 96 | 9 | 1 | 4 |

proposed method is higher than that of the other three methods. The YOLO v5s lacks training with respect to unknown types of obstacles. As a result, it cannot detect generalized obstacles and will consequently offer reduced safety when used in realistic vehicle situations. The VIDAR method does not require training and can detect all generalized obstacles.

However, the optical axis of the camera in the aforementioned two methods is fixed in the traditional binocular system, which limits the field of view of the camera and causes missed detection of obstacles. Therefore, the traditional binocular vision detection method based on YOLO v5s has the largest number of errors and missed detections. The proposed active binocular vision detection method has the advantage of VIDAR in detecting generalized obstacles. Furthermore, it can enlarge the camera's field of vision and reduce the missed detection of obstacles, thus ensuring the effectiveness of obstacle detection results on the road.

In the analysis of results, accuracy (A), precision (P) and missing rate (M) are used as evaluation indices for the four obstacle detection methods, as shown in Table 3. The results in Tables 2 and 3 show that misjudgment or misdetection may occur during vehicle movement due to vehicle fluctuation and other factors. Compared with YOLO v5s, relying on the ability of VIDAR to eliminate the false obstacles on road surface, the proposed method in this paper is mainly advantageous in terms of detection accuracy. Compared with traditional binocular vision, the proposed method mainly improves the missing rate of the detection method. Compared with the traditional binocular vision detection method based on YOLO v5s, the accuracy, precisions of obstacle detection method proposed in this paper improve by 16.4%, 5.7%, respectively, and the missing rate reduce by 10.9%.

TABLE III.
EVALUATION INDICES OF FOUR DETECTION METHODS

| Detection method | A (%) | P (%) | M (%) |
|---|---|---|---|
| Traditional binocular vision based on YOLO v5s | 79.09 | 93.26 | 15.45 |
| Traditional binocular vision based on VIDAR | 90.07 | 97.83 | 9.09 |
| Active binocular vision based on YOLO v5s | 82.72 | 92.63 | 10.91 |
| Active binocular vision based on VIDAR | 95.45 | 98.97 | 4.55 |

*D.Real vehicle experiments*

In order to verify the reliability and feasibility of practical applications of the distance measurement method proposed in this paper, a group of outdoor obstacle detection experiments are carried out for real vehicles. A pure electric vehicle with a laser radar is used as the experimental vehicle. A total of five two-lane roads near the east gate of Shandong University of Technology are selected for testing. They include: 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 obstacles on the test roads include pedestrians, cars, bicycles, buses and other generalized obstacles. Figure 19 shows the experimental vehicle and roadmap.

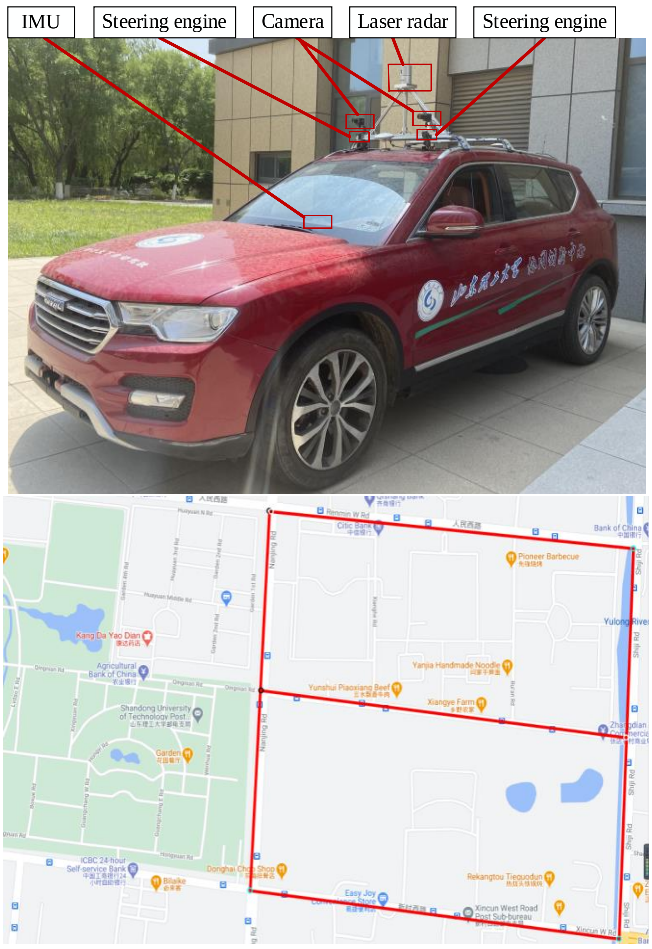



Fig. 19. Experimental vehicle and roadmap.

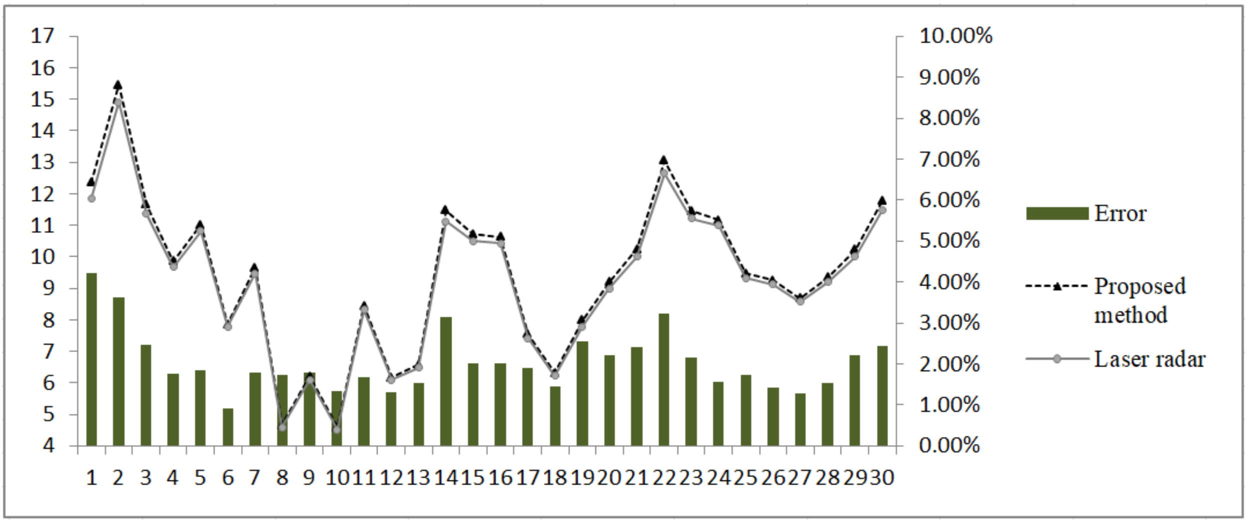



Fig. 20. Parts of the detection results.

Parts of the detection results are selected and compared with the laser radar ranging results, as shown in Fig. 20.

Experimental results show that the error of the method proposed in this paper is less than 5% at a short distance (<20 m). According to [29-30], this measurement error meets the existing requirements of vision-based ranging. Therefore, these results show that the proposed active binocular distance measurement algorithm based on VIDAR meets the requirements of measurement accuracy and can achieve accurate distance measurement of obstacles.

## V. CONCLUSIONS

The main contribution of this paper was the development of new obstacle detection method based on VIDAR and active binocular vision. In this paper, the VIDAR detection method was used to initially detect 3D obstacles, avoiding the shortcomings of the machine learning methods that could only detect known types of obstacles. In order to accurately obtain the driving information of the vehicle in VIDAR, other sensors could also be combined with IMU and the choice of sensors was relatively flexible. For example, if the combination of wheel encoder and IMU was used, the accuracy of self-propelled vehicle velocity and moving distance estimation would be improved. An obstacle search strategy based on active binocular vision was proposed, which solved the problem of a large blind area in the traditional binocular vision and made it possible to measure the distance of the obstacle in the non-matching area. The effectiveness of the proposed method was verified by the simulation experiment under controllable scenes and outdoor real vehicle experiment. Experimental results showed that the proposed method could improve the flexibility of binocular vision system and significantly increase the detection range. Compared with the traditional obstacle detection method based on binocular vision, the missing rate of the proposed method was reduced by 10.9% and the ranging accuracy could also be ensured.

The method proposed in this paper provided a way to improve the integration of artificial intelligence and vehicles. This method would help the driver to judge the obstacle location more accurately in order to take the next avoidance measures, which could considerably improve the vehicle driving safety. In the future, accurate location of the camera pose and improvement of the ranging accuracy will be considered.

## REFERENCES

[1] Chen XZ, Huang YM, and Chen SB, "Model analysis and experimental technique on computing accuracy of seam spa-tial position information based on stereo vision for welding robot," Industrial Robot: An International Journal, vol. 39, no.4, pp349–356, 2012.

[2] Ling, X., Zhao, Y., Gong, L., Liu, C., and Wang, T., "Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision," Robotics and Autonomous Systems, vol. 114, pp134-143, 2019.

[3] Du Jiang, Zujia Zheng, Gongfa Li, Ying Sun, Jianyi Kong, Guozhang Jiang, Hegen Xiong, Bo Tao, Shuang Xu, Hui Yu, Honghai Liu, and Zhaojie Ju, "Gesture recognition based on binocular vision," Cluster Computing, vol. 22, no.6, pp13261-13271, 2019.

[4] Zhiqiang Zhai, Zhongxiang Zhu,Yuefeng Du, Zhenghe Song, and Enrong Mao, "Multi-crop-row detection algorithm based on binocular vision," Biosystems Engineering, vol. 150, pp89-103, 2016.

[5] Beau Tippetts, Dah Jye Lee, Kirt Lillywhite, and James Archibald, "Efficient stereo vision algorithms for resource-limited systems," Journal of Real-Time Image Processing, vol. 10, no.1, pp163–174, 2015.

[6] Rigoberto Juarez-Salazar, Juan Zheng, and Victor H. Diaz-Ramirez, "Distorted pinhole camera modeling and calibration," Applied Optics, vol. 59, no.36, pp11310-11318, 2020.

[7] Seung-Hwan Baek, and Min H. Kim, "Stereo fusion: Combining refractive and binocular disparity," Computer Vision and Image Understanding, vol. 146, pp52-66, 2016.

[8] Imran, S., Khan, M. U. K., Mukaram, S., and Kyung, C. M., "Unsupervised Monocular Depth Estimation with Multi-Baseline Stereo," BMVC, 2020.

[9] Mansour, Mostafa, Pavel Davidson, Oleg Stepanov, and Robert Piché, "Relative importance of binocular disparity and motion parallax for depth estimation: a computer vision approach," Remote Sensing, vol. 11, no.17, 2019.

[10] Hongtao, Y., Haishuang, H., Li, L., Rongrong, Z., and Yu, Z, "Precision analysis of binocular stereo vision measurement system," Transducer and Microsystem Technologies, vol. 39, 2020.

[11] Hedman, Peter, and Johannes Kopf, "Instant 3d photography," ACM Transactions on Graphics (TOG), vol. 37, no.4, pp1-12, 2018.

[12] Kim M H., "Foundations and Applications of 3D Imaging," Theory and Applications of Smart Cameras. Springer, Dordrecht, pp63-86, 2016.

[13] Yang, S., Gao, Y., Liu, Z., and Zhang, G., "A calibration method for binocular stereo vision sensor with short-baseline based on 3D flexible control field," Optics and Lasers in Engineering, vol. 124, 2020.

[14] Fangyan Nie, and Jianqi Li, "Image Segmentation with Thresholding based on Relative Arithmetic-Geometric Divergence," IAENG International Journal of Computer Science, vol. 49, no.3, pp848-855, 2022.

[15] J. Wu, Q. Kong, K. Yang, Y. Liu, D. Cao and Z. Li, "Research on the Steering Torque Control for Intelligent Vehicles Co-Driving With the Penalty Factor of Human-Machine Intervention," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, no.1, 2022.
[16] J. Wu, J. Zhang, B. Nie, Y. Liu, and X. He, "Adaptive Control of PMSM Servo System for Steering-by-Wire System With Disturbances Observation," IEEE Transactions on Transportation Electrification, vol. 8, no.2, pp2015-2028, 2022.
[17] J. Wu, J. Zhang, Y. Tian, and L. Li, "A novel adaptive steering torque control approach for human–machine cooperation autonomous vehicles," IEEE Transactions on Transportation Electrification, vol. 7, no.4, pp2516–2529, 2021.
[18] Cheng, L., Li, T., Zha, S., Wei, W., and Gu, J., "Multichannel Saliency Detection Based on Visual Bionics," Applied Bionics and Biomechanics, vol. 2020, 2020.
[19] Ren, H., Liu, W., Shi, T., and Li, F., "A biomimetic adaptive fuzzy edge detection method based on visual features," 2016 35th Chinese Control Conference (CCC). IEEE, pp3902-3906, 2016.
[20] Chunxia Qi, and Jiandong Diao, "A Bio-Inspired Algorithm for Maximum Matching in Bipartite Graphs," IAENG International Journal of Computer Science, vol. 47, no.1, pp56-60, 2020.
[21] Wu, J., Tian, Y., Walker, P., and Li, Y., "Attenuation reference model based adaptive speed control tactic for automatic steering system," Mechanical Systems and Signal Processing, vol. 156, 2021.
[22] Wu, J., Zhang, J., Nie, B., Liu, Y., and He, X., "Adaptive control of PMSM servo system for steering-by-wire system with disturbances observation," IEEE Transactions on Transportation Electrification, vol. 8, no.2, pp2015-2028, 2021.
[23] Li Q, Chang T, and Jiao X., "Error analysis in parallel-axes binocular stereoscopic measurement system caused by the calibrated parameters," 2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS) Proceedings. IEEE, pp163-166, 2012.
[24] Jiang, J., Liu, L., Fu, R., Yan, Y., and Shao, W., "Non-horizontal binocular vision ranging method based on pixels," Optical and Quantum Electronics, vol. 52, no.4, pp1-10, 2020.
[25] Ling F, Jimenez-Rodriguez A, and Prescott T J., "Obstacle Avoidance Using Stereo Vision and Deep Reinforcement Learning in an Animal-like Robot," 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, pp71-76, 2019.
[26] Xu, Y., Gao, S., Li, S., Tan, D., Guo, D., Wang, Y., and Chen, Q., "Vision-IMU Based Obstacle Detection Method," International Conference on Green Intelligent Transportation System and Safety. Springer, Singapore, vol. 503, pp475-487, 2017.
[27] Gamma Kosala, Agus Harjoko, and Sri Hartati, "Robust License Plate Detection in Complex Scene using MSER-Dominant Vertical Sobel," IAENG International Journal of Computer Science, vol. 47, no.2, pp214-222, 2020
[28] Yi, X., Song, G., Derong, T., Dong, G., Liang, S., and Yuqiong, W., "Fast road obstacle detection method based on maximally stable extremal regions," International Journal of Advanced Robotic Systems, vol. 15, no.1, 2018.
[29] Noa Garnett, Shai Silberstein, Shaul Oron, Ethan Fetaya, Uri Verner, Ariel Ayash, Vlad Goldner, Rafi Cohen, Kobi Horn, and Dan Levi, "Real-time category-based and general obstacle detection for autonomous driving," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp198-205, 2017.
[30] Ya Zhang, Fei Yu, Yanyan Wang, and Kai Wang, "Performance Evaluation of Feature Detection Methods for Visual Measurements," Engineering Letters, vol. 27, no.2, pp320-327, 2019.