An Improved Obstacle Detection and Segmentation Method Based on VIDAR and Binocular Vision

Liming Wang, Yi Xu, Ruoyu Zhu, Shaohong Ding, Jinxin Yu, Teng Sun, Guoxin Jiang, Shanshang Gao, Xiaotong Gong, Yuqiong Wang, Dong Guo, Pengwei Wang, Bingzheng Liu

***Abstract*—Obstacle detection is one of the key steps of intelligent driving. The traditional stereo vision obstacle detection method has the advantages of low cost and direct ranging, but it still has the disadvantages of slow detection speed, small detection range and low detection accuracy. In order to overcome these problems, an obstacle detection and segmentation method based on VIDAR and binocular vision is proposed in this paper. First, the MSER algorithm extracts the obstacle feature area in the image, and regards the centroid of the feature area as the corresponding feature point. VIDAR judges whether the feature point has a height, and removes the pseudo-obstacle feature points and retains the real obstacle feature points. Then, for the obstacle feature points in the binocular field of vision, three-dimensional coordinate point K-means clustering and binocular vision technology are used to segment the feature points. For the obstacle feature points in the monocular field of vision, two-dimensional morphological closed operation and monocular vision technology are used to segment the feature points. Based on the characteristics that the height of the same feature point on the same obstacle remains unchanged at the different time, FFT template matching, DBSCAN clustering, threshold judgment and other methods are used to locate, match, track, calculate the height and verify the threshold results of the template points, so as to realize the real-time detection and accurate segmentation of obstacles. Finally, the sparse optical flow method is used to track the segmentation results, and the detected obstacles are continuously segmented and optimized to achieve accurate detection of obstacles. The experimental results show that the proposed method can effectively improve the detection distance, detection accuracy and segmentation accuracy of obstacles, provide a new idea for the accurate detection of multiple overlapping obstacles, and further improve the intelligence, safety and reliability of driverless vehicles.**

***Index Terms*—Binocular Vision, Morphological Closed Operation, *K*-means Clustering, FFT Template Matching, DBSCAN Clustering**

Manuscript received June 24, 2022; revised January 6, 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.

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

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

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

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

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

Guoxin Jiang is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 2250534973@qq.com).

Shanshang Gao is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 2823862098@qq.com).

Xiaotong Gong is a postgraduate student at Shandong University of Technology, Zibo 255000, China. (e-mail: 1412354020@qq.com).

Yuqiong Wang is a PhD student at Shandong University of Technology, Zibo 255000, China. (e-mail: wangyuqiong@sdut.edu.cn).

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

Compared with monocular vision, binocular vision is affected by more external factors when obtaining distance information on a spatial three-dimensional (3D) scene [1]-[7]. Previous studies have shown that a wide baseline can provide higher accuracy than a narrow baseline [8]-[13]. Namely, when measuring the same obstacle, the error of binocular ranging based on a wide baseline is smaller than that based on a narrow baseline [14]-[17]. Olmos R. et al. [18] introduced a convolutional neural network model into the field of binocular detection, calculated a parallax map by constructing a low-cost symmetric dual-camera system, and used this information to improve the candidate regions of input frames, thus improving the accuracy of binocular detection. Baek E. T. and Ho Y. S. [19] proposed a detection method for stereo matching and hole filling occlusion based on dynamic programming, which can improve the accuracy of depth estimation compared with the traditional binocular depth map generation method. Sayyedbarzani S. A. and Emam S. M. [20] reduced the influence of uncertainties on the 3D reconstruction by establishing a mathematical relationship model between the convergence angles of two cameras. Priya L. et al. [21] used cameras to capture multiple views of objects to overcome the self-occlusion problem of objects and employed binocular parallax, depth estimation, and area extraction to address occlusion between objects. They also used improved geometric mapping techniques to realize the 3D reconstruction of a target, achieving the detection of specific targets with an occlusion degree not exceeding 80%. Although wide-baseline settings based on stereo vision can provide larger and more reliable measurement results of different images and ensure more

accurate reconstruction, there is an inherent problem of wide baselines, the occlusion problem, which cannot be effectively solved [22]-[27]. In addition, although the obstacle detection methods based on stereo vision have the advantages of low cost and direct ranging, they still require extracting and matching a large number of pixels or feature points to generate a depth map for judging obstacles, resulting in a slow speed [28]-[32]. These methods also have a small detection range and a low parallax calculation accuracy under a small baseline. To solve these problems, this paper proposes an obstacle detection and segmentation method based on VIDAR and binocular vision. The proposed method mainly consists of two steps, namely, feature region generation and feature region verification, optimization, and tracking. Compared with the traditional obstacle detection methods, this method has obvious advantages in the detection range, detection accuracy and segmentation accuracy of obstacles, which can provide a basis for the precise control of driverless vehicles in the later stage and improve the safety of vehicle driving [33]-[37].

In this paper, the MSER algorithm is used to perform binocular matching on the feature points detected by VIDAR. According to whether the matching is successful or not, the feature points are divided into matching points and mismatching points, and the corresponding areas are recorded as matching and mismatching regions, respectively. Comparing the corresponding fields of vision of the single and binocular cameras, it can be seen that the matching region belongs to the common detection area of the two cameras; in the binocular camera, the mismatching region is the remaining detection area other than this area, as shown in Fig. 1, where the orange area denotes the matched region, and the blue area is the mismatching region.

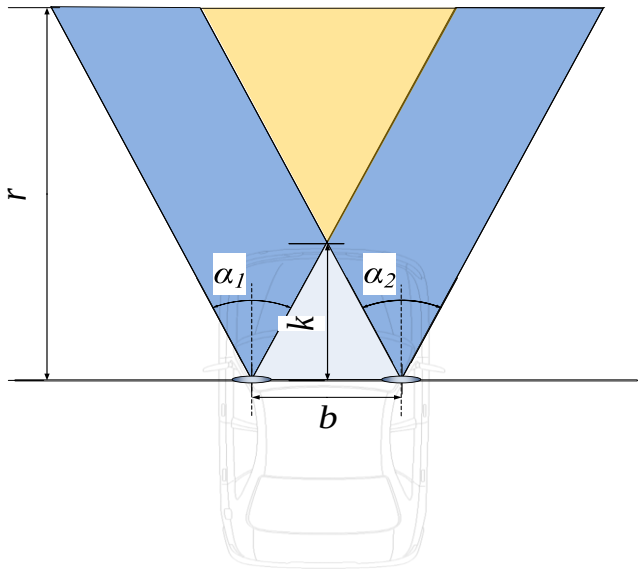


Fig. 1. Illustration of matched and mismatching regions.

II. Obstacle Detection and Segmentation Algorithm

To detect and segment obstacles more quickly and accurately, this paper proposes an obstacle detection and segmentation method based on VIDAR and binocular vision, whose block diagram is presented in Fig. 2.

The overall steps of the proposed method are as follows.

(1) At time $T1$, a vehicular binocular camera acquires an image of the front field of vision and performs binarization processing. The MSER algorithm is used to obtain MSER regions (denoted as MSERs), and the centroid of each MSERs is taken as the corresponding feature point to establish a one-to-one correspondence between the MSERs and the feature points. An obstacle detection method based on VIDAR is used to determine the height of all feature points in the extracted image. If feature points belong to the point on an obstacle, they are denoted as $C\left(C=\{C_1, C_2, \ldots, C_n\}\right)$; otherwise, they are denoted as $C^*\ \left(C^*=\{C_1^*, C_2^*, \ldots, C_n^*\}\right)$.

(2) At time $T1$, binocular matching feature point C, according to the matching results, the camera's field of view is divided into matching and mismatching regions. If the feature point C can be successfully matched, it belongs to the matching point and is denoted as $Ca(Ca=\{Ca_1, Ca_2, ..., Ca_n\})$ (the area composed of matching points represents a matching region); otherwise, it belongs to the mismatching point and is denoted as $Cb(Cb=\{Cb_1, Cb_2, ..., Cb_n\})$ (the area composed of mismatching points represents a mismatching region).

(3) Binocular ranging and *K*-means clustering are used to segment obstacles in the matching region. Assume the pixel coordinates of Ca are expressed as $Ba=(X_i, Y_i)(i=1,2,...,n)$, the distance between different matching points and the binocular camera is $Da(Da=\{Da_1, Da_2, ..., Da_n\})$ and let $Z=Da(Z=Z_1, Z_2, ..., Z_n\})$; then, a three-dimensional coordinate point of the matching point is denoted by $A=(X_i, Y_i, Z_i)(i=1,2,...,n)$. Assume that the optical center of a camera is the coordinate origin $(0,0)$; the line between the two cameras denotes the X-axis, the Y-axis is perpendicular to the X-axis, and the Z-axis is perpendicular to the XOY plane in the direction toward the obstacle. The *K*-means clustering is applied to different matching points, and each matching point is segmented into obstacles of different depths according to the segmentation results.

(4) Monocular ranging and morphological operation are used to segment obstacles in the mismatching region. Let the pixel coordinates of mismatching point Cb be $Bb=(X_i, Y_i)(i=1,2,...,n)$ and perform the closed morphological operation on the pixel point coordinate Bb. Assume that the lowest point of the obstacle area after two-dimensional morphological processing is the contact point between the obstacle and the road surface, denoted as $P(P=\{P_1, P_2, ..., P_n\})$. Using monocular vision measures, distances from different obstacle areas to the camera are obtained and denoted by $Db(Db=\{Db_1, Db_2, ..., Db_n\})$. The feature point Cb is divided into obstacles in different planes according to different distances from Db.

(5) The accuracy of obstacle segmentation is verified by analyzing the height of the same feature point on an obstacle. For the matching region, the horizontal distances from the camera to the landmark feature point at times $T1$ and $d_1=d_2+\Delta d+\Delta l$ are measured by binocular ranging and

recorded as S_{1i} and $S_{2i}(i=1,2,...,\mathrm{n})$, respectively. For the mismatching region, the lowest point of an obstacle where the feature point is located in step (4) is taken as a contact point P with the road surface, and the distance from the obstacle to the monocular camera after the measurement is Db; S_{1i} and S_{2i} are distances from the corresponding feature point to the monocular camera. The installation height of the camera is h; the camera moves at a distance of Δd at Δt time; the actual height of the matching point Ca is denoted as $h_{vi}(i=1,2,...,\mathrm{n})$; the detection distance is obtained using the pinhole imaging principle (i.e., the horizontal distance between the line between the camera projection center and the feature point and the road intersection) are denoted as d_1 and d_2. It is assumed that values of the height of the same feature point on an obstacle from the ground at times $T1$ and $(T1+\Delta t)$ are h_{v1} and h_{v2}; if $h_{v1}=h_{v2}$, the feature point belongs to the obstacle, and obstacle segmentation results obtained in steps (3) and (4) are correct; otherwise, the obstacle segmentation results (3) and (4) are wrong.

(6) According to the obstacle segmentation result (5), the height $h_{v1}(i=1,2,...,\mathrm{n})$ of different feature points at different times is obtained by (9). The matching point at time $T1$ is used as a template point, and the fast Fourier transform (FFT) template matching method is used to locate and track the template point at different times. The height threshold α is conditionally defined; the DBSCAN clustering is performed to obtain the effective threshold range; the height calculation and threshold result verification of the template point are conducted by the threshold judgment method.

(7) For the feature points with segmentation errors, if they belong to the matching region, they are directly taken as target points to be tracked; if they belong to a mismatching region, the method of dividing obstacles is used in (4) to perform re-segmentation. For a single or a few discrete feature points that cannot be segmented, the segmentation is no longer performed, and they are directly taken as target points to be tracked.

(8) Select feature points in re-segmented obstacles and discrete feature points that cannot be segmented. A sparse optical flow algorithm is adopted to take an image at time $T1$ as a reference image, and a set of pixel points of the corresponding feature points on the image plane is denoted as $M(M=\{M_1,M_2,...,M_\mathrm{n}\})$. The image at time $(T1+\Delta t)$ is taken as the current image, and a set of pixel points on the image plane corresponding to the feature points is denoted as $N(N=\{N_1,N_2,...,N_\mathrm{n}\})$; the set of pixels on the two frames of images at the previous and next moments are matched to track the obstacles.

A. Feature region generation

The feature region is not only the result of obstacle segmentation in the early stage but also the object of segmentation verification in the later stage. Accurate and fast segmentation of obstacles is the main goal of this stage. The specific steps of the feature region generation are presented in Fig. 3 and explained in the following.

VIDAR-based feature point detection

During feature region generation, binocular cameras are treated as two separate cameras. The obstacle detection algorithm based on VIDAR is used to make a high degree of discrimination on all feature points extracted by the MSER algorithm. Retain feature points that belong to obstacles; Eliminate feature points that are pseudo-obstacles. The VIDAR is an obstacle detection method based on a visual inertial measurement unit (IMU). It has a good detection effect on static and dynamic obstacles, achieving fast detection speed and high detection accuracy at a low cost. The detection principle of VIDAR is as follows.

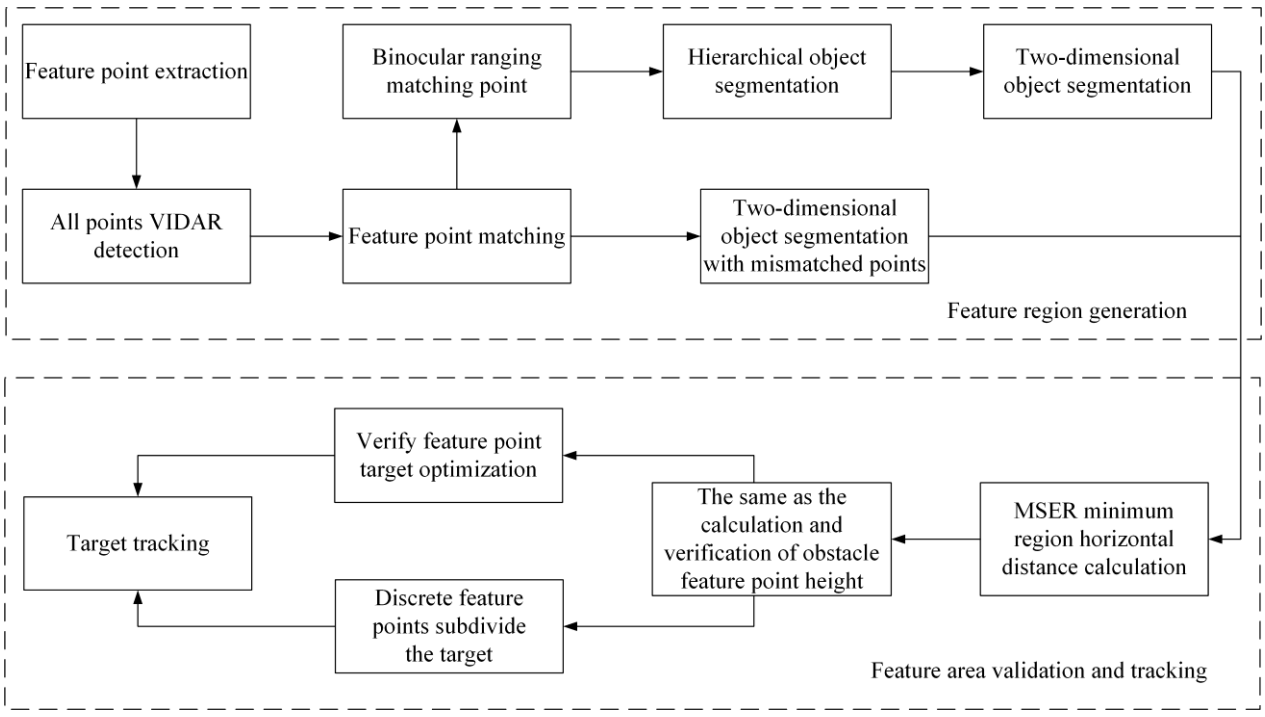


Fig. 2. Schematic diagram of the detection and segmentation process of the proposed method.

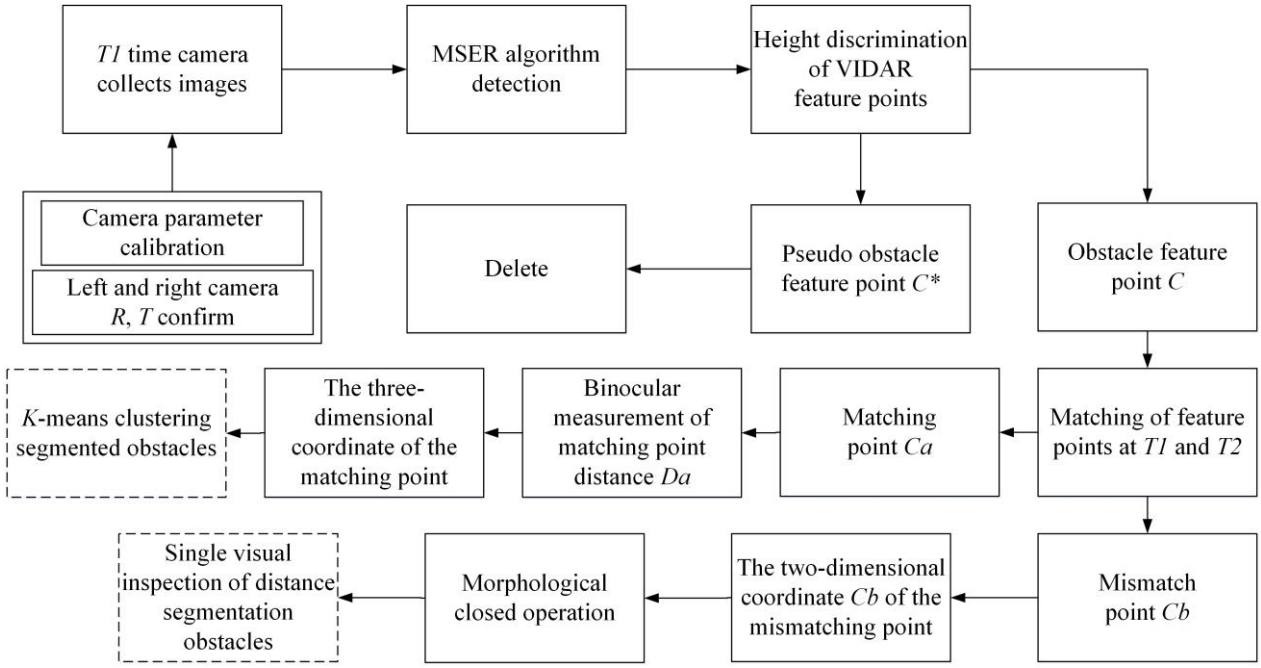


Fig. 3. The flowchart for the feature region generation process.

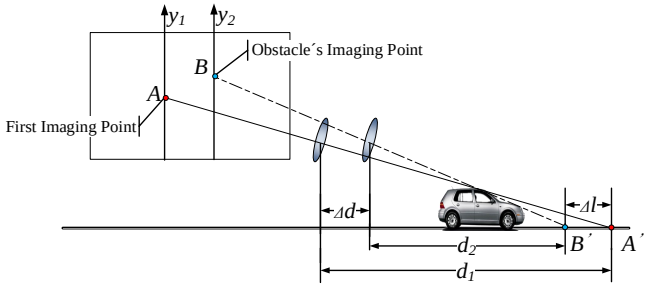


Fig. 4. Static obstacle detection based on VIDAR.

When VIDAR is used to detect feature points of static obstacles, as shown in Fig. 4, it is assumed that the top of an obstacle is the detected feature point; the imaging point of the feature point in the camera at $T1$ is A, the corresponding point on the road plane is A', and the horizontal distance from A' to the camera is d_1. At time ($T1+\Delta t$), the moving distance of the camera is Δd. The axis on the image plane changes from y_1 to y_2, the imaging point of the feature point in the camera changes to B; the imaging point on the corresponding road plane is B'; the horizontal distance from the camera to B' is d_2. Then, the relationship between distances is expressed as follows:

$$d_1 = d_2 + \Delta d + \Delta l \tag{1}$$

Comparison of the magnitudes of Δl ($\Delta l = |d_1 - \Delta d - d_2|$) and K (K is threshold, $K > 0$) shows that when $\Delta l \le k$, the feature points are points on the road surface, and the corresponding MSER region does not belong to obstacles; therefore, the corresponding feature points are removed. When $\Delta l > k$, the feature point is not a point on the road surface, and the corresponding MSER region belongs to the obstacle, so this point is retained. The detection of dynamic obstacles by VIDAR is similar to static obstacle detection, which has been described in detail in [38].

In a binocular detection method, a different baseline width results in different values of detection range and detection accuracy. Under the premise of ensuring an effective binocular field, a wide-baseline system has higher detection accuracy than a narrow-baseline system. This study employs a dual-VIDAR wide-baseline obstacle detection method a dual-VIDAR wide-baseline obstacle detection method under the condition of guaranteeing binocular precision. Compared to the traditional wide baseline obstacle detection methods based on binocular stereo vision, this method can provide a wider testing range and higher accuracy. The comparison between the dual-VIDAR detection range and the traditional binocular detection range is shown in Fig. 5.

Assume two camera optical axis in space gathering point D, the depth distance detected by the camera is r, and the horizontal distance between the projection center of the camera and the depth r detected on the camera side is e_1 and e_2; the field of views of the cameras are α_1 and α_2; the vertical distance between the camera baseline and the intersection D of the optical axes of the two cameras is denoted by K; the detection area of dual VIDAR is S_V, the detection area of the binocular camera is S_S; the horizontal distance b_1 from the intersection point D of the optical axes of the two cameras to the projection center of the camera on the left is regarded as the vertical distance K from the camera baseline b to point D remains constant, and it can be calculated by:

$$\begin{aligned}
\tan\left(\frac{\pi-\alpha_1}{2}\right) &= \frac{k}{b_1} \\
\tan\left(\frac{\pi-\alpha_2}{2}\right) &= \frac{k}{b-b_1} \\
\tan\left(\frac{\pi-\alpha_1}{2}\right) &= \frac{r}{e_1} \\
\tan\left(\frac{\pi-\alpha_2}{2}\right) &= \frac{r}{e_2}
\end{aligned} \tag{2}$$

Equation can be obtained after arrangement as follows:

$$\begin{cases} k = \dfrac{b\tan\left(\dfrac{\pi-\alpha_1}{2}\right)\tan\left(\dfrac{\pi-\alpha_2}{2}\right)}{\tan\left(\dfrac{\pi-\alpha_1}{2}\right)+\tan\left(\dfrac{\pi-\alpha_2}{2}\right)} \\ e_1 = \dfrac{r}{\tan\left(\dfrac{\pi-\alpha_1}{2}\right)} \\ e_2 = \dfrac{r}{\tan\left(\dfrac{\pi-\alpha_2}{2}\right)} \end{cases} \tag{3}$$

If $\alpha_1 = \alpha_2 = \alpha$, since $\tan\left(\dfrac{\pi-\alpha}{2}\right) = \dfrac{1}{\tan\dfrac{\alpha}{2}}$, it holds that:

$$\begin{cases} S_V = rb + r^2\tan\dfrac{\alpha}{2} - \dfrac{b^2}{4\tan\dfrac{\alpha}{2}} \\ S_S = \left(r - \dfrac{b}{2\tan\dfrac{\alpha}{2}}\right)^2 \tan\dfrac{\alpha}{2} \end{cases} \tag{4}$$

After the arrangement, we have:

$$\begin{cases} S_V = rb + r^2\tan\dfrac{\alpha}{2} - \dfrac{b^2}{4\tan\dfrac{\alpha}{2}} \\ S_S = -rb + r^2\tan\dfrac{\alpha}{2} + \dfrac{b^2}{4\tan\dfrac{\alpha}{2}} \end{cases} \tag{5}$$

When $S_V < S_S$, the detection range of dual-VIDAR is smaller than that of binocular vision, and the baseline width is $b > 4r\tan\dfrac{\alpha}{2}$. Assume the camera detection depth is $r \in [10,150]$, camera field angle is $\alpha \in [30,114]$ and let $g = 4r\tan\dfrac{\alpha}{2}$, then $g_{\min} = 10.72$, indicating that $S_V < S_S$ is valid only when the baseline width is $b > 10.72$ m. Since the vehicle width is less than 10.72 m under normal conditions, the baseline distance of the vehicle wide baseline binocular vision cannot meet $b > 10.72$ m. This indicates that the proposed dual-VIDAR obstacle detection method can provide a larger detection range than the traditional binocular vision obstacle detection method.

Obstacle feature point segmentation

When the points with a height determined by VIDAR are in the matching region, binocular ranging divides the extracted matching points into different planes according to their vertical distance from the binocular camera. Then, the corresponding MSERs are obtained according to the position of matching points. Finally, the *K*-means clustering method is used to segment the extracted MSERs into obstacles belonging to different depths. The *K*-means clustering segmentation algorithm based on binocular distance measurement is adopted for feature points of the matching region, which avoids the need to measure multiple dense points to form a region like conventional binocular ranging, and improves the accuracy and speed of segmentation [39]-[42].

When the points with a height determined by VIDAR are in the mismatching region, the morphological closing algorithm is used, and the feature points are divided into obstacles in different areas by the expansion and erosion algorithms. Then, the lowest point of the segmented obstacle is regarded as a contact point *P* of the obstacle and the road surface, and the distance *Db* between the obstacle and the camera is obtained based on the monocular measurement of distance. Finally, the obstacles are divided into different planes according to the different *Db* values.

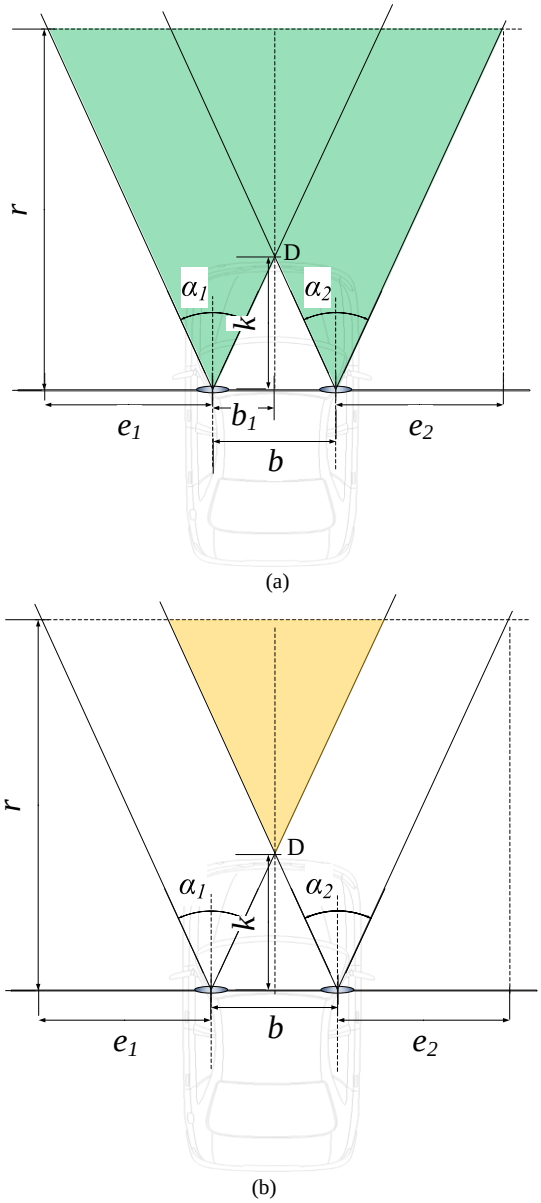


Fig. 5. Detection range: (a) dual-VIDAR; (b) traditional binocular vision.

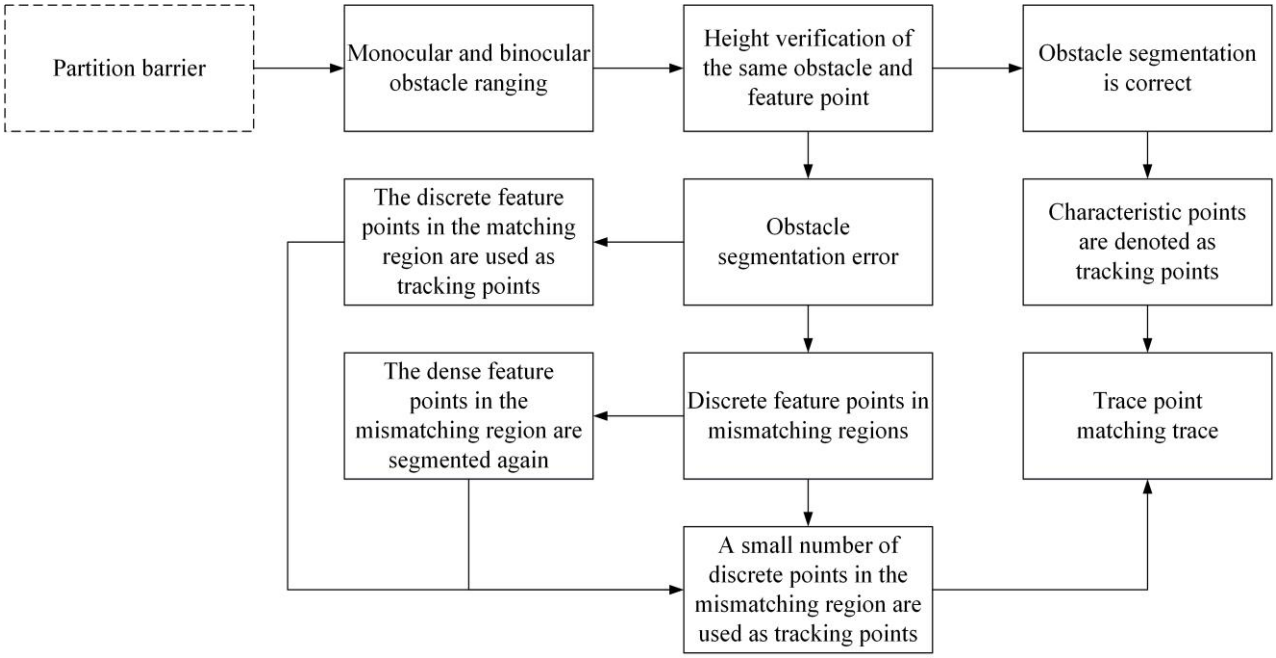


Fig. 6. Flowchart of the feature region verification, optimization, and tracking.

B. Feature region verification and target tracking

To ensure the accuracy of the obstacle segmentation and target tracking, a new obstacle feature point height verification method based on monocular vision and pinhole imaging is proposed. The height verification of the same obstacle and feature point after segmentation is performed based on the principle of single visual distance and pinhole imaging. Namely, it is determined whether the height difference of the same feature point of the same obstacle at the front and the back two moments meets the threshold requirements. Also, the accuracy of the obstacle segmentation in the early stage is verified. The flowchart of the feature region verification, optimization, and target tracking is presented in Fig. 6.

Height verification of feature points of the same obstacle

Image information collection is a process of converting 3D objects into 2D planes, which can be simplified into a pinhole camera model, as shown in Fig. 7. Assume that the effective focal length of a camera is *f*, the mounting height of the camera is *h*, and the pitching angle of the camera is ∂; the origin of the plane coordinate system (x_0, y_0), which is an intersection of the image plane and the camera optical axis, has been commonly set to (0, 0). The intersection point of the obstacle in front and the road plane is *P*, and the coordinates of point *P* in the image plane coordinate system are (*x*, *y*). The horizontal distance *d* between point *P* and the camera is obtained by the monocular measurement of distance, and it is calculated by (6).

$$d = \frac{h}{\tan\left(\partial + \arctan\left[\left(y_0 - y\right)\mu / f\right]\right)} \tag{6}$$

The lowest point of the segmented obstacle is regarded as point *P* that intersects with the ground, and the horizontal distances S_1 and S_2 from the obstacle to the camera are measured by monocular measurement of distance, as shown in Fig. 8. The height of the camera *h* is known, and the moving distance Δd of the camera at time Δt is obtained by the speed odometer; virtual distances d_1 and d_2 are obtained by the pinhole imaging. Then, the corresponding heights h_{v1} and h_{v2} of the same feature point at times $T1$ and $T1+\Delta t$ satisfy relations (7).

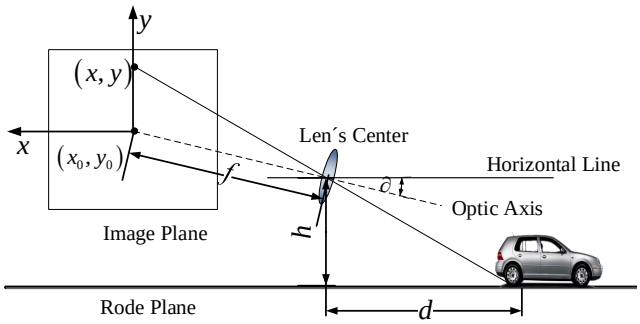


Fig. 7. Monocular camera ranging model.

$$\begin{cases} \dfrac{h_{v1}}{h} = \dfrac{d_1 - s_1}{d_1} \\ \dfrac{h_{v2}}{h} = \dfrac{d_2 - s_2}{d_2} \end{cases} \tag{7}$$

If the same feature point is on the same obstacle at two moments, the corresponding heights must be equal, $h_{v1} = h_{v2}$. That is, if $\frac{d_1 - s_1}{d_1} = \frac{d_2 - s_2}{d_2}$ is satisfied, feature points belong to the same obstacle; otherwise, feature points do not belong to the same obstacle. For feature points that do not belong to the same obstacle, optimization processing is required.

Template matching and height difference threshold verification

Taking the matching point at time $T1$ as a template point, and using the FFT template matching algorithm to locate,

track, and match the feature points at time $T1+N\Delta t\left(N=1,2...,n\right)$. By calculating the height difference between the template point and the feature point, it is determined whether the tracking target is the same feature point. Considering the accuracy of template selection and template matching, the template feature point and the actual matching feature point are not the same feature point, which can cause that after template matching, for the same obstacles, the height of the feature points is different in different times. Using the method of limiting conditions and setting thresholds, the height calculation and threshold result verification of the same feature point at different times are carried out. If the difference between the feature point and the template point is within the preset effective threshold interval, these points are considered to be the same feature point, indicating that the obstacle segmentation in the early stage is correct; otherwise, the early obstacle segmentation is wrong.

During the experiment, for the same feature point of the same obstacle at times $T1$ and $T1+\Delta t$, the absolute value of its height difference was calculated by $\left|\mathrm{h}_{v1}\text{-}\mathrm{h}_{v2}\right|=\alpha$, as shown in Fig. 8. The formula can be simplified as follows:

$$\left|\frac{d_1S_2-d_2S_1}{d_1d_2}\right|=\alpha \tag{8}$$

where indicates the threshold; for S_1, S_2, d_1 and d_2, it holds that $S_1<d_1$ and $S_2<d_2$.

The height h_{v1} of a feature point at time $T1$ is calculated by:

$$h_{v1}=\frac{(d_1-s_1)\ h}{d_1} \tag{9}$$

According to (8), factors related to threshold α include S_1, S_2, d_1, and d_2. To ensure the precision of template matching and verify the accuracy of the experiment with the same obstacle and feature point height, the value ranges of S_1, S_2, d_1, and d_2 are defined by experimental equipment to obtain the value of threshold α; the effective range of threshold α is obtained by the DBSCAN clustering. The range of threshold α is used as a basis for the verification of the feature points in the later experiment. In the experiment, the RENVISION camera was used; the size of the target plane was 6.18 mm × 5.85 mm; the focal length range was f = 5 mm–50 mm; the image resolution was 3264 × 2448; the frame number was 15 frames per second; the minimum distance between the height of an obstacle on the target plane was the distance length of a single pixel, and the maximum distance was half of the height (the camera target surface height of an obstacle was denoted by H). It was assumed that the camera's internal and external parameters were fixed; vehicle speed was 10 m/s–20 m/s; the camera installation height was h=1.7 m; the camera elevation angle was ∂ = 0°; the origin of the plane coordinate system was set to (x_0, y_0) = (0, 0); the position of a template point was (x_i, y_i) (i = 1, 2,…, n). The calculation started at a distance of zero from the coordinate position of the camera; the visual field distance of the camera at time $T1$ was S_1, and that at time $T2$ was S_2. Based on the visual field of the camera at time $T1$, the minimum and maximum values of S_2, were obtained according to the vehicle speed and the number of frames of the camera. The specific parameters' values are shown in Table I.

TABLE I
CAMERA PARAMETERS AND THE CORRESPONDING FIELD OF VIEW

Symbol	Quantity(When the camera focal length is 5 mm)	Quantity(When the camera focal length is 50 mm)
H	(0.002389 mm, 2.925 mm)	(0.002389 mm, 2.925 mm)
S_1	(5 m, 100 m)	(5 m, 100 m)
S_2	(6 m, 102 m)	(6 m, 102 m)
d_1	(0.523797 m, 62.064187 m)	(0.523797 m, 62.064187 m)
d_2	(0.523797 m, 62.064187 m)	(0.523797 m, 62.064187 m)

Considering the field of vision of the later experiment, it was assumed that S_1=[0 m, 20 m], S_2=[1 m, 22 m], and d_1=d_2=[1.62387584 m, 124.178232 m]. The results of the DBSCAN clustering corresponding to the height threshold α at a camera focal length of f= 5 mm are shown in Fig. 9.

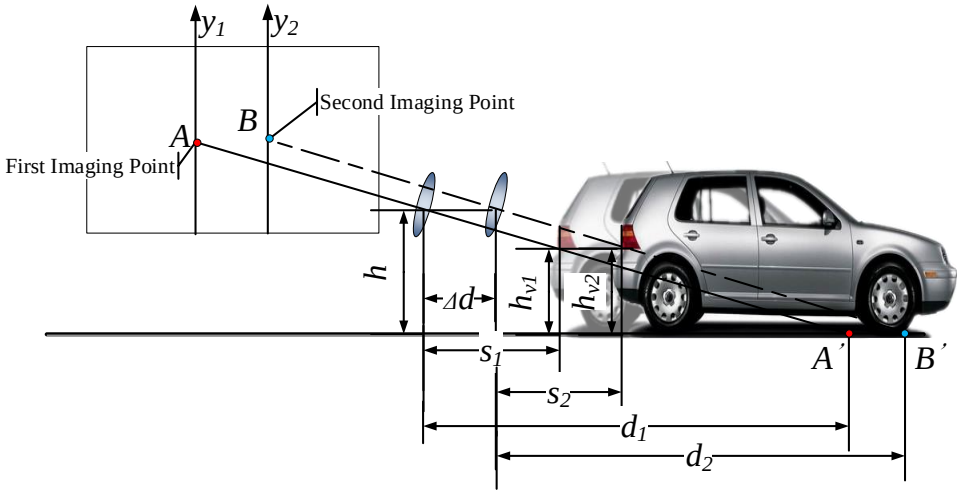


Fig. 8. Calculation principle of the feature point height.

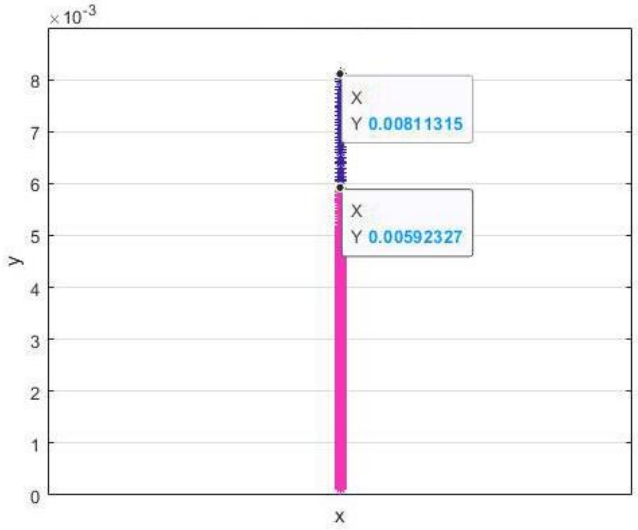


Fig. 9. The DBSCAN clustering result of the height threshold.

The DBSCAN density clustering results showed that the height threshold was divided into two intervals (0, 0.00592327 m) and (0.00592327 m, 0.00811315 m). The maximum height threshold α was 0.00811315 m, but most of the threshold values were in the interval of (0, 0.00592327 m), and only a small part of them was in the interval of (0.00592327 m, 0.00811315 m). Therefore, the effective threshold interval was set as (0, 0.00592327 m). The upper limit of the valid threshold was 0.00592327 m, and its value was rounded at 0.006 m. In the later experiment, if the height difference between the template point and a feature point at different times was larger than 0.006 m, the template point was selected again after removing the feature point; if the height difference between the template point and a feature point at different times was less than 0.006 m, but the height difference of a certain frame was larger than 0.006 m, the framed picture was removed, and the feature points at the remaining moments are matched and tracked again with the FFT template matching to ensure the accuracy of feature point tracking when the same obstacle is highly verified.

Obstacle feature point optimization and target tracking

After the height verification with obstacle feature points, if feature points with segmentation errors appear in the matching region, they are regarded as target points to be tracked; if there are discrete feature points with segmentation errors in the mismatching region, the closed morphological algorithm is used for subsequent segmentation; feature points that cannot be subdivided are directly taken as target points to be tracked.

The sparse optical flow method is used to track stable feature points in the obstacle after re-segmentation and feature points that cannot be re-segmented. Set an image at time $T1$ as a reference image; a set of pixels of the corresponding feature point on the image plane is M ($M = \{M_1, M_2, \ldots, M_n\}$). Set the image at time ($T1+\Delta t$) as the current image; the corresponding feature point on the image plane includes a collection of pixels N ($N = \{N_1, N_2, \ldots, N_n\}$). By matching the feature points on the image at times $T1$ and ($T1+\Delta t$), the successfully matched feature point indicates that the tracking of the obstacle is realized, and then time ($T1+\Delta t$) is set as the previous time, and time ($T1+2\Delta t$) is set as the next time; the feature points on the image in the two frames, are matched again, and the above operations are repeated to achieve continuous tracking of the target.

III. Experimental Result Analysis

In the indoor experiment, two RENVISION cameras fixed on the camera bracket were placed on the road plane with a baseline width of 10 cm. The road line stickers simulated the road environment in reality and acted as false obstacles. Four model vehicles were used to simulate obstacles in front of the vehicle under test; the lid acted as a generalized barrier. Zhang Zhengyou calibration method is used to calibrate the two cameras, the calibration results are shown in Fig. 10. According to the results of the binocular camera calibration, the internal and external parameters of the camera, including the rotation variable R and the translation variable T between the cameras, were adjusted; the camera installation height was 8 cm; the camera effective focal length was 5 mm; the camera pitch angle was 0.132 rad, the test data acquisition and analysis process is as follows.

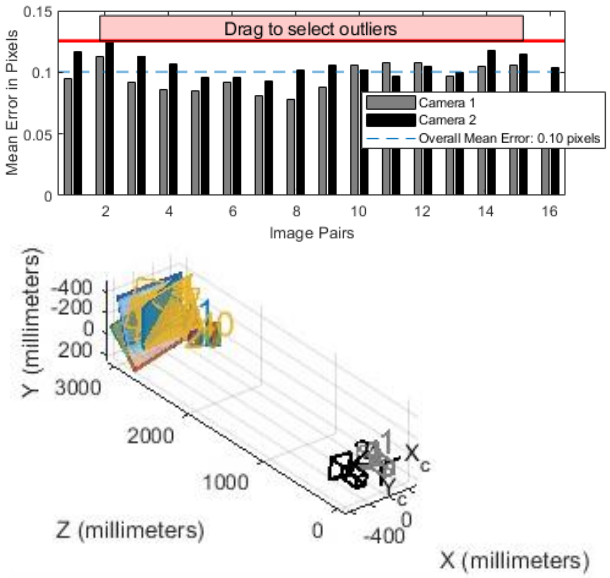
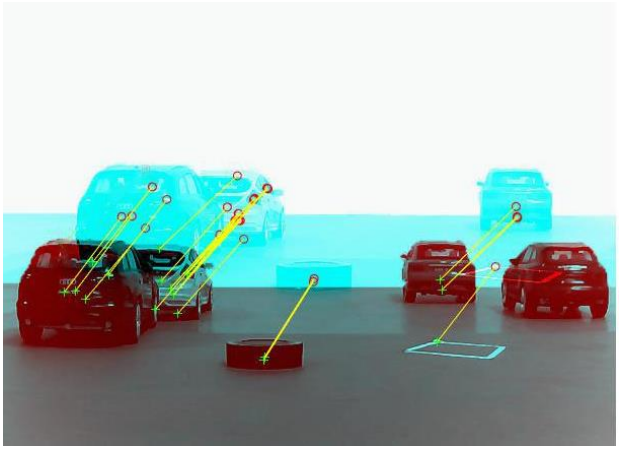


Fig. 10. Calibration results.

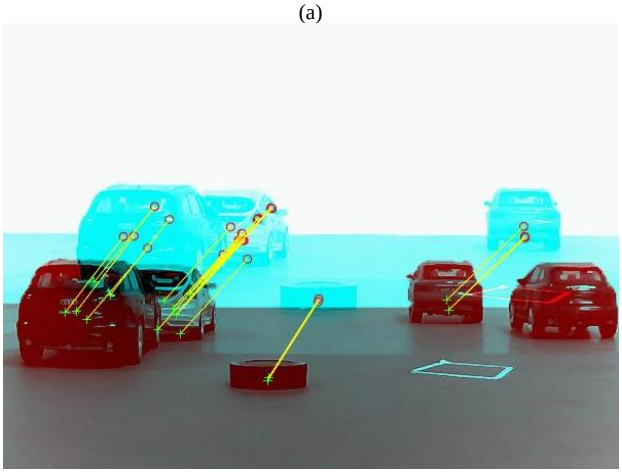
(a)

(b)

Fig. 11. Collected images: (a) left camera; (b) right camera.



(a)



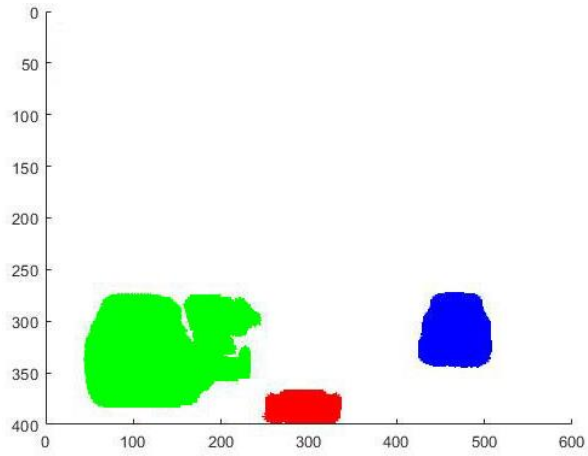
(b)

Fig. 12. Matching graphs of feature points: (a) without VIDAR detection; (b) after VIDAR detection.

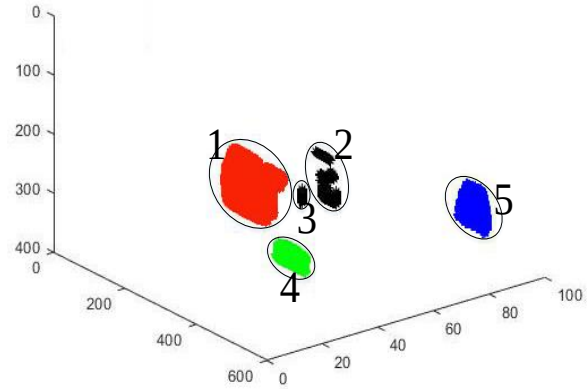
Taking the optical center of the camera as the coordinate origin, the line between the two cameras was the X-axis; the Y-axis was perpendicular to the X-axis; the direction perpendicular to the XOY plane in the direction toward an obstacle was the Z-axis, where $Z=Da$ ($Da = \{Da_1, Da_2, \ldots, Da_n\}$). The Z value of each matching point was taken as the Z value of its corresponding MSERs, and the ranging accuracy of a millimeter-wave radar was taken as the Z value reference. By adjusting the internal and external parameters of the binocular camera, the ranging accuracy of the binocular camera was corrected. Table II shows the 3D coordinates of the obtained matching points at time $T1$.

A. Segmentation accuracy analysis

At time $T1$, a three-dimensional coordinate point A (X_i, Y_i, Z_i) of the matching point was processed by the K-means clustering to segment it into obstacles of different depths.



(a)



(b)

Fig. 13. The K-means clustering results: (a) two-dimensional matching points; (b) three-dimensional coordinate points.

TABLE II
THE 3D COORDINATES OF THE $T1$ TIME MATCHING POINT

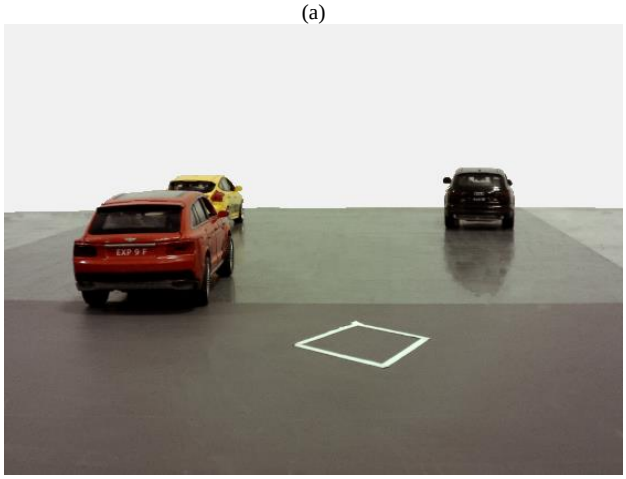
Serial number	Three-dimensional coordinate (X, Y, Z)	Serial number	Three-dimensional coordinate (X, Y, Z)
Point 1	(124.185, 294.329, 53.256)	Point 9	(203.826, 326.783, 71.582)
Point 2	(139.594, 307.344, 55.242)	Point 10	(186.207, 340.914, 67.479)
Point 3	(96.2857, 323.81, 55.3196)	Point 11	(208.278, 344.829, 72.4812)
Point 4	(106.682, 323.364, 55.3185)	Point 12	(293.75, 387.999, 48.1661)
Point 5	(116.6, 331.371, 55.175)	Point 13	(292.888, 391.004 48.1456)
Point 6	(217.393, 305.857, 72.165)	Point 14	(466.916, 311.76, 91.4511)
Point 7	(191.711, 313.105, 71.822)	Point 15	(468.859, 321.686, 91.6722)
Point 8	(200.625, 320.281, 71.579)	Point 16	(464.333, 372.762, 68.458)

The *K*-means clustering results of three-dimensional coordinate points show that the MSERs of matching points in the matching region were divided into four categories, representing four obstacles in total. Among them, "1" represented obstacle A, "2" and "3" represented obstacle B, "4" represented obstacle C, and "5" represented obstacle D. In fact, in the real obstacle situation, there were three obstacles, namely obstacles A, B, and D, and a generalized obstacle was obstacle C. Although the *K*-means clustering segmentation result was the same as the total number of obstacles in the matching region, the *K*-means clustering result had the following problems. Due to the occlusion of obstacle A, and the close distance between the tail of obstacle B and the front of obstacle A, part of obstacle A was mistakenly classified as obstacle B during the *K*-means clustering; that is, in a real environment, after the *K*-means clustering, "3" in obstacle B would represent a part of obstacle A.

By comparing the *K*-means clustering results of three-dimensional coordinate points and two-dimensional matching points, it can be seen that the three-dimensional coordinate points can more intuitively show the relative position relationship between the feature points and better reflect the true state of the obstacles than the two-dimensional matching points.

At time $T2$, with the movement of vehicles, the distance between vehicles increased, and the vehicle information after the *K*-means clustering segmentation became more prominent.

(a)



(b)

Fig. 14. Collected images: (a) left camera; (b) right camera.

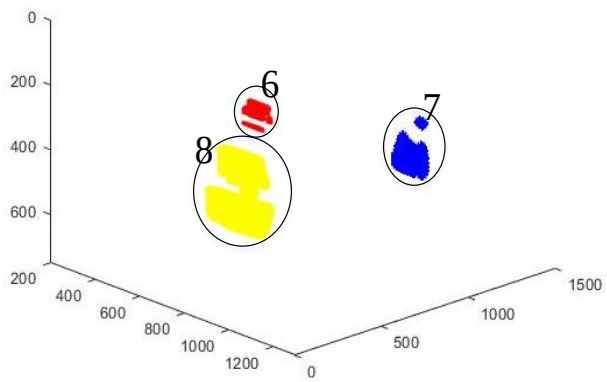


Fig. 15. The *K*-means clustering method of three-dimensional coordinate points.

The *K*-means clustering results show that the matching region included three obstacles. Among them, "6" represented obstacle E, "7" represented obstacle F, and "8" represented obstacle G; the result of obstacle division was consistent with the real obstacle situation. The *K*-means clustering of the three-dimensional coordinate points made the information on the obstacle segmentation more 3D, and the segmentation result of the obstacle was closer to the real number and contour of the obstacle in the camera's field of vision.

The binocular ranging and *K*-mean clustering were performed on the matching image on the left side of Fig. 12 that had not been detected by VIDAR, as shown in Fig. 16. Obstacles were detected using the Fast-RCNN, and the results are shown in Fig. 17.

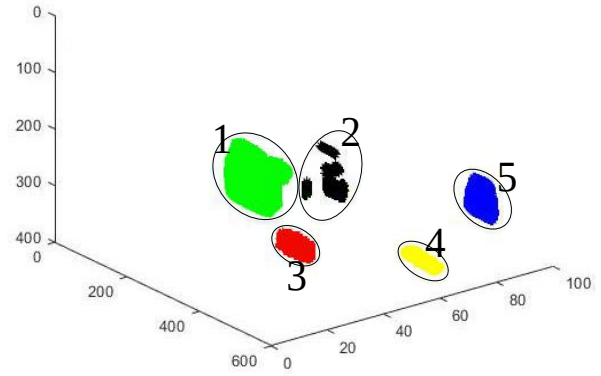


Fig. 16. The *K*-means clustering results of three-dimensional coordinate points.

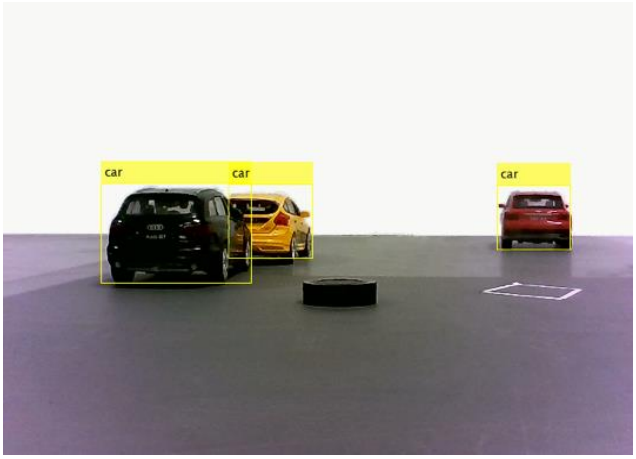


Fig. 17. The results of Fast-RCNN test.

The *K*-means clustering results showed that the matching region included five obstacles. Without considering the segmentation error of the *K*-means clustering in Fig. 13, the feature points that had not been detected by VIDAR were directly matched, resulting in one more obstacle being segmented in Fig. 16. After VIDAR detection, the matching region at times $T1$ and $T2$ was finally segmented into fine obstacles; the reason for the segmentation error was that the MSERs ("4" in Fig. 16) corresponding to the matching point of the road marking were divided into one obstacle. Compared with the Fast-RCNN and other obstacle detection methods, which could detect three obstacles (Fig. 17), the proposed method also had a good detection effect on generalized obstacles. Therefore, the feature points detected by VIDAR could reduce the error rate of obstacle segmentation and improve the accuracy of obstacle segmentation.

The true number of obstacles (NRO), the total number of segmented obstacles (OT), the number of correctly segmented obstacles (OC), the number of incorrectly segmented obstacles (OF), the number of missed segmented obstacles (OM), false detection rate (FDR), missed detection rate (MDR), correct detection rate (CDR), and correct segmentation rate (CSR) were obtained, and FDR = OF / NRO, MDR = OM / NRO, CDR = OC / NRO, and CSR = OT / NRO (CSR less than 1 indicates that the number of obstacles after segmentation is less than the actual number of obstacles, there are multiple obstacles mistakenly divided into an obstacle; CSR greater than 1 indicates that the number of divided obstacles is greater than the actual number of obstacles, and that there is a split of an obstacle into multiple obstacles). Table III compares the segmentation accuracy of different methods in the matching region of the real vehicle experiment.

TABLE III
OBSTACLE SEGMENTATION ACCURACY IN THE MATCHING REGION

Symbol	Traditional binocular vision	Faster-RCNN	VIDAR+ Binocular vision
NRO	1246	1246	1246
OT	1325	1217	1263
OC	1201	1187	1229
OF	101	37	12
OM	22	39	5
FDR (%)	8.11	2.97	0.96
MDR (%)	1.77	3.13	0.4
CDR (%)	96.39	95.26	98.64
CSR (%)	106	97.67	101

It can be seen from Table III that the detection accuracy of the obstacle detection and segmentation method based on VIDAR and binocular vision reached 98.64 %, which was significantly higher than 96.39 % of the traditional binocular vision obstacle detection method and 95.26 % of the Faster-RCNN obstacle detection method. In terms of obstacle false detection rate and missed detection rate, compared with the other two methods, the method proposed in this paper has also been significantly reduced. In terms of correct segmentation rate, the method proposed in this paper makes the correctly segmented obstacles closer to the true number of obstacles by continuously segmenting obstacles in the camera 's field of view at different times. On the basis of VIDAR effectively eliminating pseudo-obstacles on the road surface, the obstacle feature point detection and segmentation method based on VIDAR and binocular vision can effectively detect more generalized obstacles, while the traditional binocular vision obstacle detection method cannot eliminate pseudo-obstacles, and the Faster-RCNN obstacle detection method cannot detect generalized obstacles outside the training set due to the influence of its training samples.

B. Detection range analysis

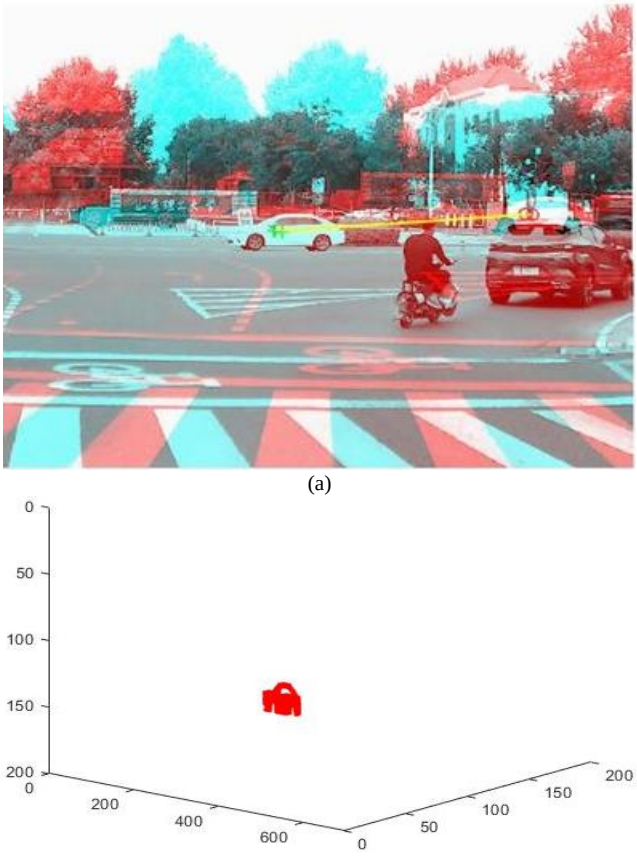
According to (2)–(5), the proposed dual-VIDAR obstacle detection method has a larger detection range and a higher detection accuracy than the traditional binocular vision detection method. In the matching area, the feature points are divided into obstacles in different planes by using the *K*-means clustering results of three-dimensional coordinate points combined with binocular ranging. In the mismatched region, assuming an obstruction after the two-dimensional morphological closed operation is the contact point *P* between the obstacle and the road surface, the distance *Db* of an obstacle to the monocular camera was used to divide into different planes of obstacles.

The results in Figs. 18–21 showed that in the matching region, the dual-VIDAR detection method detected only one obstacle, while the traditional binocular vision detection method detected three obstacles, including one real obstacle and two false obstacles. In the mismatched region, the proposed dual-VIDAR detection method detected three obstacles. Table IV shows the number of obstacles detected by different detection methods in the outdoor experiment.

(a)

(b)

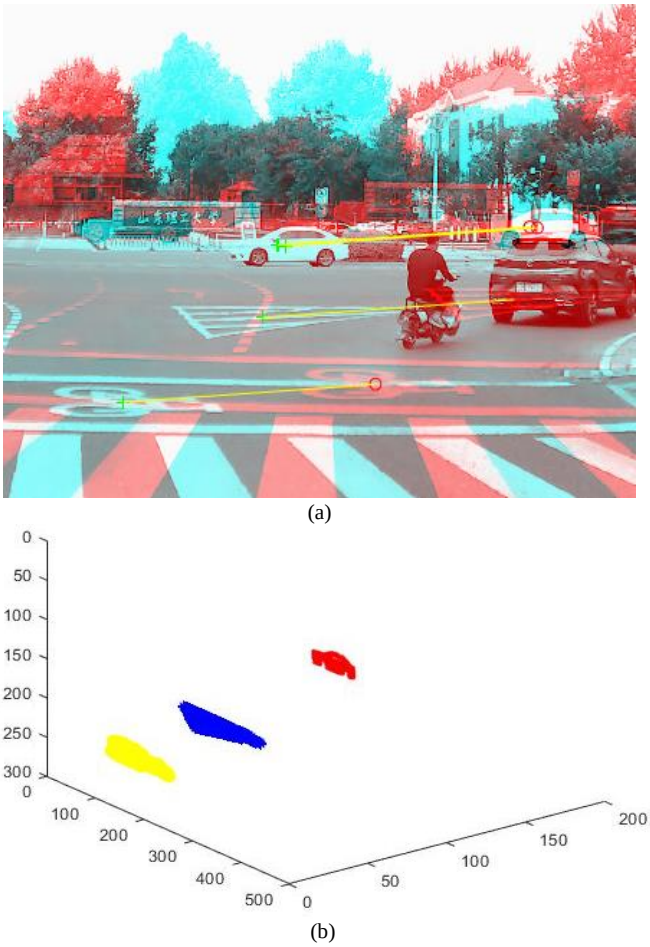
Figure 18: Collected images. (a) left camera; (b) right camera.



(a)

(b)

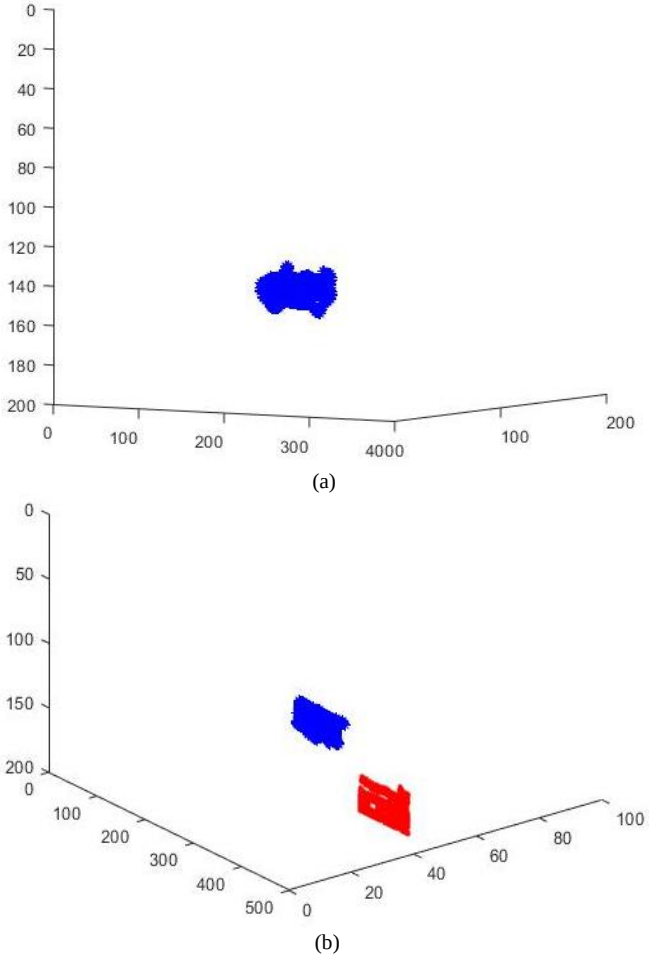
Figure 19: Detection results of the dual-VIDAR in the matching region.



(a)

(b)

Figure 20: Detection results of the traditional binocular vision in the matched region.



(a)

(b)

Fig. 21. Obstacle segmentation results: (a) left camera's mismatched region; (b) right camera's mismatched region.

TABLE IV
NUMBER OF OBSTACLES DETECTED BY DIFFERENT METHODS

Symbol	Traditional binocular vision	YOLOv3	Dual-VIDAR
Detection field of vision	Matching region	Matching region	Matching region + Mismatched region
FDR (%)	7.25	2.15	0.88
MDR (%)	1.64	3.01	0.37
CDR (%)	95.28	94.86	98.93
CSR (%)	104	97.78	101

It can be seen from Table IV that in the range of camera detection, compared with the traditional binocular vision and YOLOv3 obstacle detection method, the dual-VIDAR obstacle detection method has a wider detection range, lower false detection rate and missed detection rate, higher detection accuracy and segmentation accuracy. This is because the dual-VIDAR detection method can detect non-overlapping areas other than monocular camera overlapping scenes. In addition, the method proposed in this paper can detect more generalized obstacles and pseudo-obstacles on the basis of providing a wider detection range, and use VIDAR to eliminate the pseudo-obstacles. The low accuracy of traditional binocular vision detection is not only affected by the detection range of the camera, but also by the type of obstacles detected. The low detection accuracy of YOLOv3 is not only affected by the camera detection area, but also by its training samples.

C. Verification analysis of match point height

According to (7), the height of the same obstacle and the same feature point is constant over time. At time $T1$, the matching point was used as a template point in the matching region, as shown in Fig. 22. The height comparison of template points at times $T1$ and $T1+\Delta t$ is presented in Fig. 23.

As shown in Fig. 23, within the height threshold of 0.006 m, among all points tracked and matched by templates at time $T1$ and $T1+\Delta t$, only the height difference of template point 10 exceeded a value of 0.006 m, and the height of point 10 at time $T1$ was smaller than that at time $T1+\Delta t$. At time $T1$, according to the information on matching points in the experiment, point 10 should belong to obstacle A; however, because these two points were too close to obstacle B, they were divided into obstacle B after the *K*-means clustering, and the calculation results of these two points were wrong when the height of matching points was calculated at that time because S_1 and d_1 in (9) were replaced by S_2 and d_2 at time $T1+\Delta t$, the calculation result of height h was wrong. At time $T1+\Delta t$, when the distance between obstacles A and B increased, the *K*-means clustering segmentation result became correct, and point 10 was divided into obstacles. At this point, the height value of the obstacle matching point satisfied (9), and the calculated result of the actual height value of point 10 was correct. The experimental result showed that the value of point 10 at time $T1$ was less than that at time $T1+\Delta t$, which was consistent with the actual situation. Table V shows the tracking data of matching points at different times in the real vehicle experiment, where d_1 is the height of the starting template matching point, and $\Delta d = \max\{|d_i - d_1|\}$ $(i = 2, 3, ..., 8)$.

The height threshold of the outdoor experiment is 0.02 m. It can be seen from Table V that the $\triangle$d of Trace Points 4 and Trace Points 10 exceeds the threshold range and does not meet the conditions, indicating that the obstacles segmented by these two matching points at different times are wrong. The reason is that the distance between the obstacles is too close during the tracking process, resulting in the feature points belonging to the two obstacles being segmented into one obstacle. In order to ensure the accuracy of target tracking in the later stage, the tracking points exceeding the height threshold are eliminated. In addition, the experimental results show that the obstacle detection and segmentation method based on VIDAR and binocular vision can effectively solve the obstacle occlusion problem that traditional binocular detection is difficult to deal with by continuously segmenting the detected obstacle feature points in the camera field of view, and improve the accuracy of detection and segmentation in the state of multiple overlapping obstacles.

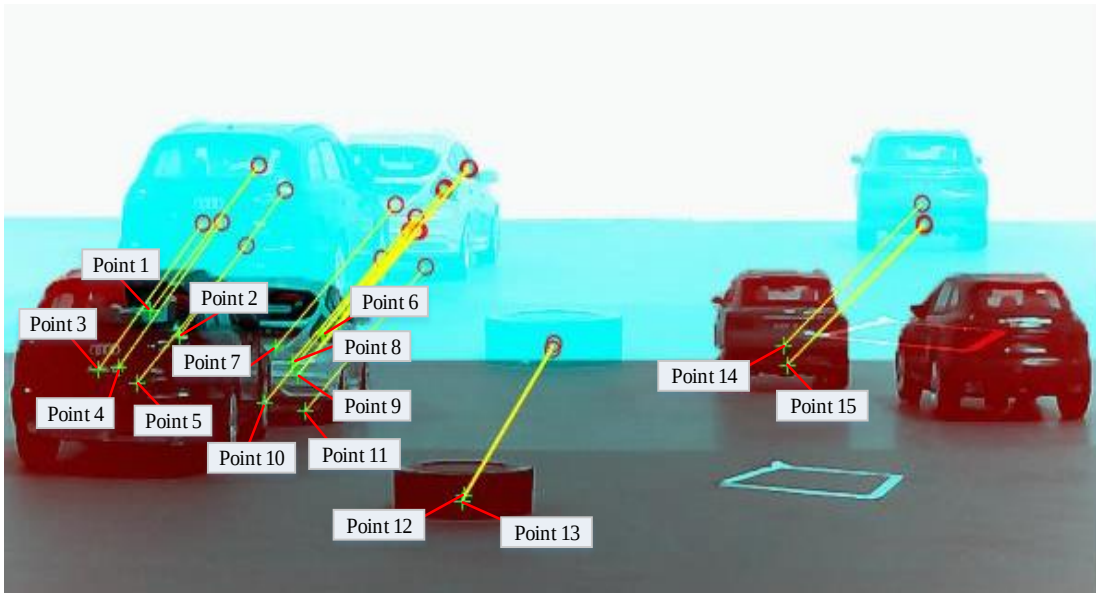


Fig. 22. Tracing number of matching points at time.

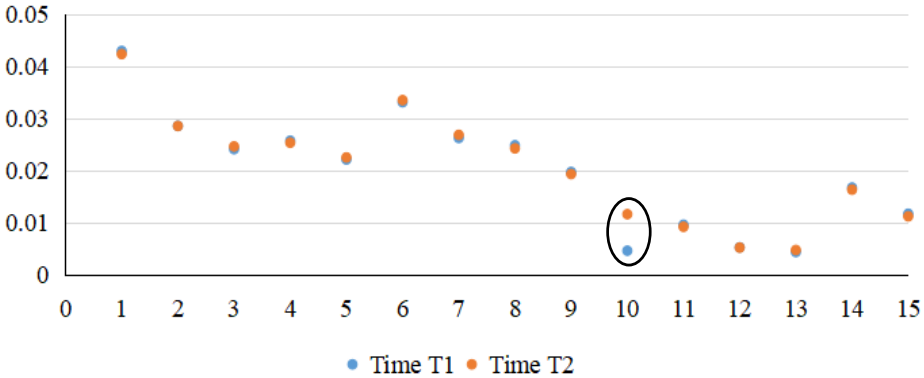


Fig. 23. Comparison of the template point height at times and for points 1–15.

TABLE V
HEIGHT COMPARISON OF MATCHING POINTS AT DIFFERENT TIMES

Trace Points	d_1/m	d_2/m	d_3/m	d_4/m	d_5/m	d_6/m	d_7/m	d_8/m	△d/m
1	0.7834	0.7821	0.7897	0.7801	0.7925	0.7914	0.7887	0.7704	0.013
2	0.4749	0.4722	0.4678	0.4615	0.4774	0.4847	0.4774	0.4819	0.0134
3	0.8631	0.8751	0.8725	0.8601	0.8549	0.8578	0.8619	0.8546	0.012
4	1.1801	1.1237	1.1258	1.1303	1.1274	1.1098	1.1156	1.1214	0.0703
5	1.0747	1.0652	1.0698	1.0735	1.0758	1.0734	1.0841	1.0835	0.0095
6	1.2316	1.2399	1.2301	1.2307	1.2485	1.2479	1.2424	1.2248	0.0109
7	1.4144	1.4041	1.4236	1.4058	1.4201	1.4015	1.4089	1.4155	0.0129
8	1.3425	1.3467	1.3328	1.3574	1.3533	1.3299	1.3388	1.3354	0.0149
9	1.4597	1.4547	1.4658	1.4662	1.4551	1.4532	1.4529	1.4588	0.0068
10	0.6788	0.9366	0.9544	0.9485	0.9358	0.9874	0.9917	0.9325	0.3129
11	0.8724	0.8689	0.8735	0.8718	0.8615	0.8628	0.8883	0.8847	0.0123
12	1.3779	1.3748	1.3708	1.3735	1.3722	1.3702	1.3751	1.3766	0.0077
13	1.1622	1.1748	1.1602	1.1635	1.1688	1.1659	1.1694	1.1598	0.0072
14	0.9918	1.0013	0.9959	0.9934	1.0105	0.9825	0.9849	0.9979	0.0187

IV. CONCLUSION

This paper presents an obstacle detection and segmentation method based on VIDAR and binocular vision. In the aspect of obstacle detection, VIDAR is used to discriminate the height of a small number of stable feature points extracted by MSER algorithm, eliminate the pseudo-obstacle feature points, retain the real obstacle feature points, improve the accuracy of obstacle detection, and reduce the complexity of post-processing of feature points. In the aspect of obstacle segmentation, according to the camera vision area where different feature points are located, monocular vision, binocular vision, morphological closed operation, *K*-means clustering and other methods are used to segment obstacles belonging to different regions, which improves the accuracy of obstacle segmentation. An obstacle detection method based on double VIDAR is proposed. On the basis of eliminating the detected pseudo-obstacle points, the detection range of traditional binocular vision is increased and the accuracy of obstacle detection is improved. The accuracy of the previous obstacle segmentation results is verified by using the same feature point height of the same obstacle at different times. Using FFT template matching, DBSCAN density clustering, height threshold determination and other methods to track and match the segmented obstacle feature points, the accuracy of obstacle feature point segmentation and the accuracy of subsequent vehicle tracking are improved.

Through simulation experiment and real vehicle test, the process of obstacle detection and segmentation based on VIDAR and binocular vision is described. The traditional binocular vision detection method, Faster-RCNN detection method and obstacle detection and segmentation method based on VIDAR and binocular vision are applied to indoor and outdoor comparative experiments, and FDR, MDR, CDR and CSR are used as evaluation indicators to compare the test results. The results show that the method proposed in this paper can effectively solve the problem of occlusion between obstacles on the basis of eliminating pseudo-obstacles. It has high detection accuracy and segmentation accuracy, and analyzes the reasons for high accuracy. Combined with the real vehicle test, the traditional binocular obstacle detection range, the obstacle detection range based on YOLOv3 and the obstacle detection range based on double VIDAR proposed in this paper are compared, and the detection accuracy and segmentation accuracy of the three detection methods are compared with the corresponding indicators. The results show that the proposed method has a larger detection range, higher detection accuracy and segmentation accuracy. The reasons for the large detection range, high detection accuracy and high segmentation accuracy are analyzed. Taking the experimental equipment such as camera as the constraint condition, the adjustment range of the height threshold of the same feature point at the same time is established, and the height information of different tracking points detected by the camera at different times is analyzed. Based on the value of △d, the tracking points that do not meet the requirements are eliminated, and the accuracy of target segmentation is improved. The method proposed in this paper improves the intelligence of the vehicle, provides a new idea for accurate detection under multiple overlapping obstacles, and further improves the safety and reliability of the autonomous vehicle. Accurate vehicle control will be its next research direction.

REFERENCES

[1] Hibbard P B, Haines A E, Hornsey R L. Magnitude, precision, and realism of depth perception in stereoscopic vision[J]. Cognitive Research: Principles and Implications, 2017, 2(1): 1-11.

[2] Horna L, Fisher R B. Plane labeling trinocular stereo matching with baseline recovery[C]//2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA). IEEE, 2017: 17-20.

[3] Keefe B D, Watt S J. Viewing geometry determines the contribution of binocular vision to the online control of grasping[J]. Experimental brain research, 2017, 235(12): 3631-3643.

[4] Saleem N H, Klette R. Accuracy of free-space detection: monocular versus binocular vision[C]//2016 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, 2016: 1-6.

[5] Harms H, Beck J, Ziegler J, et al. Accuracy analysis of surface normal reconstruction in stereo vision[C]//2014 IEEE intelligent vehicles symposium proceedings. IEEE, 2014: 730-736.

[6] Jin D, Yang Y. Sensitivity analysis of the error factors in the binocular vision measurement system[J]. Optical Engineering, 2018, 57(10): 104109.

[7] Zhang J, Du R, Gao R. Passive 3D reconstruction based on binocular vision[C]//Tenth International Conference on Graphics and Image Processing (ICGIP 2018). International Society for Optics and Photonics, 2019, 11069: 110690Y.

[8] Baek S H, Kim M H. Stereo fusion: Combining refractive and binocular disparity[J]. Computer Vision and Image Understanding, 2016, 146: 52-66.

[9] Imran S, Khan M U K, Mukaram S, et al. Unsupervised Monocular Depth Estimation with Multi-Baseline Stereo[C]//BMVC. 2020.
[10] Mansour M, Davidson P, Stepanov O, et al. Relative importance of binocular disparity and motion parallax for depth estimation: a computer vision approach[J]. Remote Sensing, 2019, 11(17): 1990.
[11] Yang Hongtao, He Haishuang, LI Li, et al. Precision analysis of binocular stereo vision measurement system [J]. Sensor and Microsystem, 2020.
[12] Hedman P, Kopf J. Instant 3d photography[J]. ACM Transactions on Graphics (TOG), 2018, 37(4): 1-12.
[13] Kim M H. Foundations and Applications of 3D Imaging[M]//Theory and Applications of Smart Cameras. Springer, Dordrecht, 2016: 63-86.
[14] Pellicano N, Aldea E, Le Hégarat-Mascle S. Robust wide baseline pose estimation from video[C]//2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016: 3820-3825.
[15] Leistner T, Schilling H, Mackowiak R, et al. Learning to think outside the box: Wide-baseline light field depth estimation with EPI-shift[C]//2019 International Conference on 3D Vision (3DV). IEEE, 2019: 249-257.
[16] Mustafa A, Kim H, Imre E, et al. Segmentation based features for wide-baseline multi-view reconstruction[C]//2015 International Conference on 3D Vision. IEEE, 2015: 282-290.
[17] Li J, Bian J, Wang L, et al. Small baseline stereo matching method based on maximum likelihood estimation[J]. Journal of Networks, 2014, 9(8): 2066.
[18] Olmos R, Tabik S, Lamas A, et al. A binocular image fusion approach for minimizing false positives in handgun detection with deep learning[J]. Information Fusion, 2019, 49: 271-280.
[19] Baek E T, Ho Y S. Occlusion and error detection for stereo matching and hole-filling using dynamic programming[J]. Electronic Imaging, 2016, 2016(5): 1-6
[20] Sayyedbarzani S A, Emam S M. 3D reconstruction accuracy improvement of a stereo vision system by changing the convergence angle of the cameras[J]. Journal of Optical Technology, 2021, 88(10): 574-578.
[21] Priya L, Anand S. Object recognition and 3D reconstruction of occluded objects using binocular stereo[J]. Cluster Computing, 2018, 21(1): 29-38.
[22] Chen K, Snavely N, Makadia A. Wide-baseline relative camera pose estimation with directional learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3258-3268.
[23] Penner E, Zhang L. Soft 3D reconstruction for view synthesis[J]. ACM Transactions on Graphics (TOG), 2017, 36(6): 1-11.
[24] Schönberger J L, Zheng E, Frahm J M, et al. Pixelwise view selection for unstructured multi-view stereo[C]//European Conference on Computer Vision. Springer, Cham, 2016: 501-518.
[25] Richardt C, Kim H, Valgaerts L, et al. Dense wide-baseline scene flow from two handheld video cameras[C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 276-285.
[26] Galliani S, Lasinger K, Schindler K. Massively parallel multiview stereopsis by surface normal diffusion[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 873-881.
[27] Islam A, Asikuzzaman M, Khyam M O, et al. Stereo vision-based 3D positioning and tracking[J]. IEEE Access, 2020, 8: 138771-138787.
[28] Ma W P, Li W X, Cao P X. Binocular vision object positioning method for robots based on coarse-fine stereo matching[J]. International Journal of Automation and Computing, 2020, 17(4): 562-571.
[29] Ueareeworakul P, Saiyod S. Obstacle detection algorithm for unmanned aerial vehicles using binocular stereoscopic vision[C]//2017 9th International Conference on Knowledge and Smart Technology (KST). IEEE, 2017: 332-337.
[30] Haifeng D, Jun Y. Research on robot binocular ranging based on SIFT feature extraction algorithm[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1607(1): 012015.
[31] Aissaoui A, Martinet J, Djeraba C. Rapid and accurate face depth estimation in passive stereo systems[J]. Multimedia tools and applications, 2014, 72(3): 2413-2438.
[32] Jellal R A, Lange M, Wassermann B, et al. LS-ELAS: Line segment based efficient large scale stereo matching[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 146-152.
[33] 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): 2015-2028.
[34] 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.
[35] Liu Ronghui, and Wei Xinhong, "Application of Improved Convolutional Neural Network in Text Classification," IAENG International Journal of Computer Science, vol. 49, no.3, pp762-767, 2022.
[36] Christine Dewi, and Rung-Ching Chen, "Deep Learning for Advanced Similar Musical Instrument Detection and Recognition," IAENG International Journal of Computer Science, vol. 49, no.3, pp880-891, 2022.
[37] Wu J, Tian Y, Walker P, et al. Attenuation reference model based adaptive speed control tactic for automatic steering system[J]. Mechanical Systems and Signal Processing, 2021, 156: 107631.
[38] Yi X, Song G, Derong T, et al. Fast road obstacle detection method based on maximally stable extremal regions[J]. International Journal of Advanced Robotic Systems, 2018, 15(1): 1729881418759118.
[39] Yanfei Lv, "Cloud Computation-Based Clustering Method for Nonlinear Complex Attribute Big Data," IAENG International Journal of Computer Science, vol. 49, no.3, pp736-744, 2022.
[40] Cochard G M, Hifi M, Samod E S, et al. A Population-Based Algorithm for the k-Clustering Minimum Biclique Completion Problem[C]//2022 8th International Conference on Control, Decision and Information Technologies (CoDIT). IEEE, 2022, 1: 1461-1466.
[41] Guan Wang, Cheng Xing, Jie-Sheng Wang, Hong-Yu Wang, and Jia-Xu Liu, "Clustering Validity Evaluation Method Based on Two Typical Clustering Algorithms," IAENG International Journal of Computer Science, vol. 49, no.3, pp871-879, 2022.
[42] Almanza M, Epasto A, Panciones A, et al. k-Clustering with Fair Outliers[C]//Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022: 5-15.