OBSTACLE DETECTION METHOD BASED ON VIDAR AND THERMAL IMAGING

Derong TAN[1], Xiaotong GONG[1], Shanshang GAO[1], Guoxin JIANG[1], Ruoyu ZHU[1], Liming WANG[1], Xiaoqing SANG[1], Yi XU[1, 2*], Yuqiong WANG[1], Jun ZHAO[3]

An obstacle detection method of thermal image based on VIDAR (Vision-IMU (Inertial Measurement Unit) based detection and ranging method) is proposed. We first perform image processing on the thermal image to obtain the initial obstacle, then extract and match feature points of the obstacles, and calculate the height of matched feature points in the target thermal obstacle according to the pinhole imaging technology and eliminate the pseudo-obstacles. Finally, the target image is tracked based on the feature points to realize the accurate detection of thermal obstacles. In this paper, the latest target detection framework based on deep learning technology, YOLOv5, is used to detect specific targets after thermal imaging image acquisition and processing, and the detection speed and accuracy are compared with the proposed method. Experiments show that this method has higher detection accuracy and more stable detection speed than YOLOv5 in specific environment and is not easy to be interfered by pseudo-obstacles. In a word, this method can meet the requirements of accuracy and speed of obstacle detection in low illumination environment and has important application value for recognition of thermal obstacles at night.

Keywords: VIDAR; Thermal imaging; Thermal obstacle detection; YOLO v5

1. Introduction

According to statistics, 60% of the world's traffic accidents occur at night, accounting for 50% of the total number of traffic accident deaths [1]. It is obvious that it is particularly important to improve the safety performance of car driving, especially in a low illumination driving environment at night.

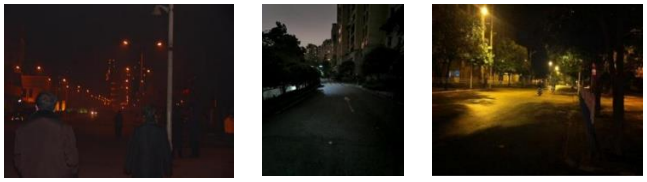
Fig. 1 shows the driving environment in residential areas under low illumination environment, which shows some problems in that environment. These make the traditional radar and visual sensors and other driving ADAS

[1] School of Transportation and Vehicle Engineering, Shandong University of Technology, 255000, Zibo, China, e-mail: 1412354020@qq.com

[2] Collaborative Innovation Center of New Energy Automotive, Shandong University of Technology, 255000, Zibo, China, e-mail: xuyisdut@163.com

[3] Shandong Tongli Construction Project Management Co., Ltd., Zibo 255000, China, e-mail: 465086269@qq.com

(Advanced Driving Assistance System) need higher precision and faster processing speed.



(a) ***(b)*** ***(c)***

Fig.1. Night scene pictures of different road sections

(a) Road sections under low illumination, ***(b)*** Residential section without streetlight, ***(c)*** Road section with streetlight but blocked.

In this paper, a thermal imaging obstacle detection method based on VIDAR is proposed to detect three-dimensional thermal obstacles. The application environment is in residential areas and other high-frequency pedestrian areas. Based on the colour feature extraction of the target frame, and with the help of pinhole imaging principle, the height of feature points is calculated to eliminate the feature points that are not on obstacles, so as to eliminate pseudo-obstacles, and improve the accuracy and speed of obstacle detection. This method not only retains the ability of thermal imaging to find thermal obstacles, but also combines with VIDAR technology to eliminate pseudo-obstacles with simple operation. In the scene of automatic driving, the rapid elimination of pseudo-obstacles in the process of driving makes the car avoid unnecessary driving action, which has an important application for driving safety.

2. Related works

The traditional detection method based on visible image is easy to fail in the area with low illumination conditions, which cannot meet the needs of all-weather positioning of unmanned driving. Based on its detection principle, thermal imaging technology is very suitable for environments with low illumination [2], complex light and heat sources. Therefore, this article combines thermal imaging technology with VIDAR (Vision-IMU based detection and ranging method) to detect thermal obstacles in relatively simple steps.

Previous improvements of thermal imaging technology usually start from the following aspects: region of interest (ROI), image enhancement, pedestrian detection algorithm and so on. In order to improve the real-time performance of the detection system, ROI are usually extracted first [3]. Without considering the use of other sensors, the main methods for ROI extraction are threshold-based segmentation method and infrared pedestrian characteristics-based segmentation method [4]. The pedestrian area can be segmented from the background by using

the brightness threshold [5]. But for the extraction of ROI, too low threshold will integrate the background into the selected obstacles; too high threshold will make the thermal obstacles segmentation incomplete. In recent years, how to improve the accuracy of pedestrian segmentation has become the focus of attention. In one study [6], Markov random field theory is used to introduce symmetry information, and an improved fuzzy clustering algorithm based on geometric symmetry information is proposed C-means clustering method is applied to infrared pedestrian segmentation and obtains better segmentation results compared with other state-of-the-art methods. Based on the c-GAN (Conditional Generative Adversarial Nets) framework [7], a thermal infrared pedestrian segmentation algorithm based on conditional generation countermeasure network (IPS c-GAN) is proposed. Although there are many methods to segment ROIs, their applications have corresponding preconditions.

From the principle of thermal imaging, we know that pedestrian heat source is obviously different from other objects without heat in the thermal image[8], but the edge contour of pedestrian and other obstacles in the preliminary thermal image is not clear. Further processing such as image enhancement technology is needed to obtain more accurate information [9]. This shows that obtaining accurate detection results requires high-quality thermal images, such as more scene information and clearer obstacle contour.

Before extraction and recognition, because of the low resolution and noise characteristics of thermal infrared image, it is necessary to enhance the thermal image [10], which directly affects the accuracy of obstacle recognition. In reference [11], low-light level visible light and infrared images are fused into a single output, and the final fused image is obtained by preserving the image information and enhancing the initial fused image. The recognition of infrared pedestrian target is usually based on machine learning classification. For example, the feature descriptors used for infrared pedestrian detection include the hog feature [12], LBP (Local Binary Pattern) feature [13] and their improved fusion features [14] mentioned above, as well as the Intensity Self-Similarity (ISS) feature [15]. In addition, deep convolution neural network also has applications in infrared pedestrian detection. Rodger et al [16] used CNN to classify pedestrians, vehicles and UAVs in long-wave infrared images, and Akula et al [17] used CNN to detect pedestrians, cars and trucks. The recognition accuracy reached about 90%, which is better than traditional detection methods. In reference [18], a multi-resolution convolution neural network is proposed, which is constructed by a multi-scale convolution operator. This method will calculate the features at different scales. In one study [19], a network structure is designed, which consists of a contraction path to capture feature content and a symmetrical expansion path to accurately locate the feature content.

In this paper [20], a complete and original model is proposed to use infrared monocular scene perception alone to evaluate whether pedestrians participate in the road in a low visibility environment. Through the time series analysis of pedestrian movement, the location of pedestrians relative to the driving area and the distance relative to the vehicle, the street intersection is evaluated. And introduce CROSSIR data set to validate the proposed method. This paper [21] proposes a method to improve the performance of thermal image pedestrian detection algorithm by fusing the features in saliency map to enrich the thermal image features. A modified version of the latest target detection network is used, and the hot images and their saliency maps are fed back to two parallel networks. Experimental results on five different datasets show that the proposed method has better performance in detecting pedestrians in thermal images than vanilla versions and baseline models. Thermal sensors play an important role in intelligent transportation systems. This paper [22] studies the RGB and thermal (RGBT) tracking problem in challenging situations by using multimodal data. An RGBT target tracking method is proposed in the correlation filtering tracking framework based on short-term historical information. Given the initial object bounding box, hierarchical convolutional neural network (CNN) is used for feature extraction. Experiments on three RGBT datasets show that the proposed method achieves comparable results with existing methods. Machine learning classification methods usually need many samples for training, and in order to ensure the accuracy of detection, a comprehensive model base is needed to avoid the problem that unknown obstacles cannot be detected.

The structure of this paper is as follows. The second part introduces the vehicle detection method proposed in this paper and explains its process in detail. The third part tests the method, analyses and compare the results, and draw a conclusion. The fourth part reviews the full paper and gives a summary.

3. Thermal imaging obstacle detection method based on VIDAR

From the principle of thermal imaging and VIDAR, we proposed a method combining infrared thermal imaging technique with VIDAR, which can quickly and accurately detect thermal objects such as pedestrians and pets in low illumination environment. This method can rapidly determine and eliminate pseudo-obstacles and complete the target obstacle detection. Fig. 2 shows the specific process.

Step1: Image preprocessing. The thermal images collected by the thermal imager at time t and the next time t+1 are respectively marked as thermal images R_1 and Z_1.

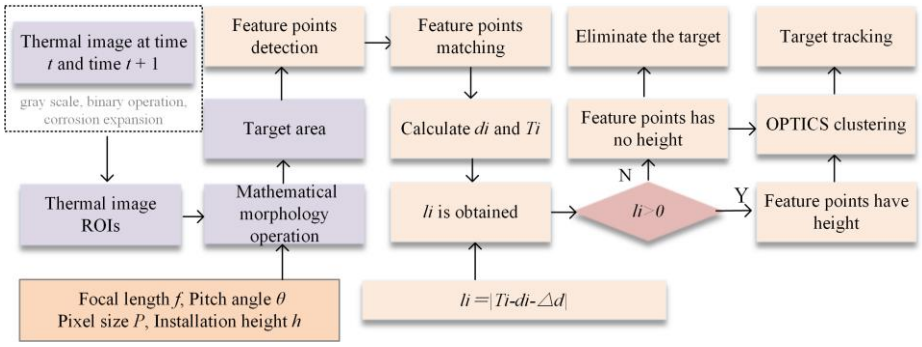


Fig.2. Thermal imaging obstacle detection method based on VIDAR

For each pixel in the thermal images R_1 and Z_1, take $\mathrm{Avg}(R,G,B)$ as the gray value of pixel according to Formula 1.

$$Gray(i,j)=(R(i,j)+G(i,j)+B(i,j)/3) \tag{1}$$

The average value method [21] is used to perform image gray-scale processing on the thermal images R_1 and Z_1 respectively, and then the obtained gray-scale image is binarized. First, the threshold T is set to 0.5, and then divide each pixel data $P(n,m)$ into the object pixel data G_1 and the background pixel data G_2 according to the threshold, and find the average values of G_1 and G_2 to be m_1 and m_2, so the new threshold $T=(m_1+m_2)/2$ is got; Each pixel data is divided into target pixel data and background pixel data with new threshold, and the average value of the sum is calculated to get the new threshold. The process is repeated until the new threshold is equal to the previous threshold. According to the above processing, we obtain the thermal images R_2 and Z_2. Median filtering is a non-linear smoothing technology, which is used to denoise the thermal image. This method sets the gray value of each pixel in R_2 and Z_2 of the thermal image as the median value of all the gray values of the pixels in a neighborhood window of the point to achieve the goal. The processed thermal images R_3, Z_3 are obtained and proceed to step 2.

Step2: ROI (region of interest) extracting. In the structured road environment, according to the various scenes that may appear in the road image and the characteristics of its spatial layout, the general road image can be divided into three different regions [22]:

- Lane area;
- Background area on both sides;
- Sky area.

Firstly, according to the prior knowledge, the ROI is set below the vanishing point of the road, and the ROI is determined by segmenting the image.

Finally, the ROI images of R_3 and Z_3 are obtained called R_{roi} and Z_{roi}, proceed to step 3.

Step3: Image open operation. After obtaining the images R_{roi} and Z_{roi} processed in step 2, there are still isolated small points and burrs. Before framing the target, the geometric operation is performed on the images to further process the ROI of the images. The ROI R_{roi} and Z_{roi} are first corroded and then expanded to get R'_{roi} and Z'_{roi}.

Step4: Target framing. The Two-Pass method [23] is used to label the connected domain of the output R'_{roi}, Z'_{roi} in step 3. In the Two-Pass connected domain labeling, the first pass is scanned from left to right and from top to bottom and set each effective pixel to a label value. The judgment rules are as follows.

- When the left and upper neighboring pixels of the target pixel are invalid values, we set a new label value to the target pixel;
- When one of the left or upper neighboring pixels of the target pixel is a valid value, we assign the label of the valid value pixel to the target pixel;
- When the left and upper neighboring pixels of the pixel are valid values, we select the smaller label value to assign to the target pixel.

The connected domains are saved with the union-find structure, the label values belonging to the same connected domain are stored in the same tree structure, and the marked connected domains are marked with a rectangular box. The frame of the target area in the image is completed, and the target is set $X = \{X_1, X_2, \cdots X_m\}, m \in N$. The image after completing the target frame is marked as R''_{roi}, and Z''_{roi} can be obtained in the same way. R''_{roi} and Z''_{roi} are transferred to the VIDAR module to perform step 5.

Step5: Feature points extraction. The target detection box set by R''_{roi} is superimposed on the image R_{roi}. The feature points outside the box are regarded as feature points without critical information, which were omitted. Harris corner detection is used to extract feature points in the target box of the image R_{roi}. Its expression is:

$$E(u,v) = \sum_{x,y} \underbrace{w(x,y)}_{\text{window function}} [\underbrace{I(x+u, y+v)}_{\text{shifted intensity}} - \underbrace{I(x,y)}_{\text{intensity}}]^2 \tag{2}$$

The meaning of Formula 2 is the self-similarity of image I after translation (u,v) at point (x,y). $w(x,y)$ is the weighted window function of point (x,y), which can be a constant. If the center point of the window is a corner point, the gray level change of the point before and after moving is the most dramatic, so the

weight coefficient of this point should be set to a larger value; while for the points far away from the center (corner) of the window, the gray change of these points is almost smooth, so these points can be set a small weight coefficient. The window function is represented by binary Gaussian function. There are two forms of expression, as shown in Fig.3.

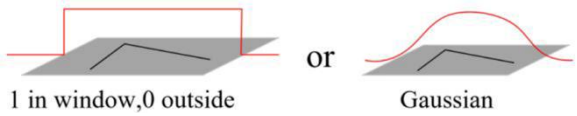


Fig 3. The window function is represented by binary Gaussian function

Then expand the formula and finally update the expression to:

$$E(u,v) \cong [u,v]M\begin{bmatrix} u \\ v \end{bmatrix} (3)$$

The *M* matrix needs to be filtered to each pixel of the image R_{roi} to obtain the horizontal gradient, the sum of the first derivative of the intensity information of the point in the *x* and *y* directions, and finally the *M* matrix is established:

$$M = \sum_{x,y} w(x,y)\begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} (4)$$

After the expression of *M* is obtained, Formula 5 is used to determine whether an area contains corner features:

$$R = \det(M) - k(trace(M))^2 (5)$$

Among them, det and trace are determinant and trace operators, which are constants with values ranging from 0.04 to 0.06 and satisfy[1]:

1. det(M)=λ 1λ 2;
2. trace(M)=λ 1+λ 2;
3. λ 1 and λ 2 are the eigenvalues of M. These eigenvalues determine whether a region is a corner, edge or flat surface.

When λ 1 and λ 2 are large and approximately equal, that is, when *R* is large, the area is a corner.

After the Harris corner is determined, each Harris corner is compared with all its adjacent points and compared with its eight adjacent points of the same scale and the eighteen points corresponding to the upper and lower adjacent scales, a total of 26 points are compared to ensure that feature points are detected in both scale space and image space. If a Harris corner is the largest or smallest among the 26 points, the corner is a feature point. The set of feature points $j_1, j_2, \cdots j_n$ and $g_1, g_2, \cdots g_n$ in R_{roi} and Z_{roi} is obtained, and then proceed to the next step 6;

Step6: Feature points matching. After feature points extraction in step 5, the matching process can be roughly divided into the following parts:

- Set the feature points of R_{roi} and Z_{roi} in step 5 as feature point descriptors: feature 1, feature 2;
- Calculate the distance between all feature vectors between feature 1 and feature 2 for matching. In order to avoid the problem of poor matching effect, the next part is added for matching correction;
- The accuracy of matching is improved by estimating the geometric transformation of matching points. The predictive affine change is used to remove outliers that do not conform to the change, then retain the correct matching points;
- The matching feature points *n* which are positive integers, and the height of feature points is determined by step 7.

Step7: Feature point height judgment. *n* pairs of feature points in the images collected at two times are extracted and matched, where *n* is a positive integer, and the distance index of the *i*-th pair of feature points is defined as $l_i, i = 1, 2, \cdots n$. Establish a rectangular coordinate system in pixels with the upper left corner of the image collected at time *t* as the origin, which is used as the image coordinate system of the image collected at that time and set the midpoint coordinates of the image collected at time *t* as (u_0, v_0). The coordinate of the feature point of the image collected at time *t* in the *i*-th pair is (u_i, v_i), d_i can be got by calculating the distance from the thermal imager at the current moment to the object point corresponding to the *i*-th pair of feature points. The schematic diagram of VIDAR calculation for judging the height of feature points is shown in Fig. 4.

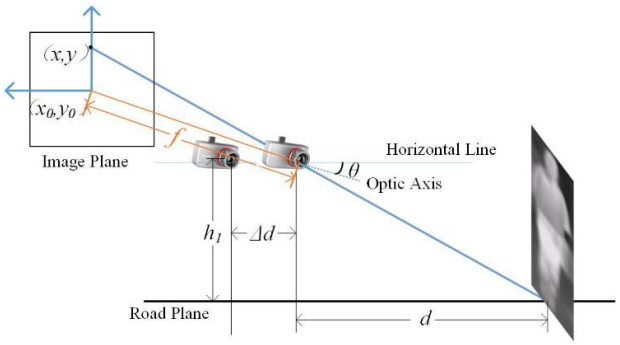


Fig.4. Schematic diagram of VIDAR calculation for judging the height of feature points

$$d = \frac{h_1}{\tan\left[\theta + \arctan\frac{(y_0 - y)\times\mu}{f}\right]} \tag{6}$$

In Formula 6, h_1 is the installation height of the thermal imager, f is the focal length of the thermal imager, θ is the pitch angle of the thermal imager, μ is the pixel size of the thermal imager, and the thermal imager is iray T3s.

The distance from the last moment to the *i*-th pair of feature points is T_i, and the calculation formula is as follows:

$$T_i = \frac{h_1}{\tan\left[\theta + \arctan\frac{(v_0 - v_i)\times\mu}{f}\right]} \tag{7}$$

In Formula 7, the thermal imager model and parameters are the same. v_0, v_i corresponds to y_0, y.

According to $l_1 = |T_i - d_i - \Delta d|$, where Δd is the distance travelled by the car, calculate the distance index of the *i*-th pair of feature points. If l_1>0, then it is judged that the *i*-th pair of feature points has a height, and it is on a three-dimensional obstacle; if $l_1 \leq 0$, then it is judged that the feature points corresponding to the *i*-th pair of feature points is on a horizontal ground and the target does not belong to three-dimensional obstacles, output the image Q after the height judgment and eliminate the pseudo-obstacles, proceed to the step 8;

Step8: Feature point clustering and pseudo-obstacle elimination. The improved DBSCAN clustering method OPTICS[24](Ordering points to identify the clustering structure) is used to cluster the feature points with height in the output image P in step 8, input the pixels of image Q as data samples, and initialize the reachable distance and core distance of all points as Max, radius ε, and minimum number of points Minpts. The algorithm is as follows:

1. Two queues are set up, which are ordered queue (core point and its direct density reachable point), and result queue (storing sample output and processing order);

2. If all the data in *D* is processed, the algorithm ends. Otherwise, select an unprocessed point as the core object from *D* and put the core point into the result queue. The direct density reachable point of the core point is put into the ordered queue, and the direct density reachable point is listed in ascending order according to the reachable distance;

3.If the ordered sequence is empty, go back to part 2, otherwise the first point is taken from the ordered queue;

- Judge whether the point is a core point. If not, return to part 3. If yes, the point will be stored in the result queue;
- If the point is the core point, find all the direct density

reachable points, and put these points into the ordered queue, and reorder the points in the ordered queue according to the reachable distance. If the point is already in the ordered queue and the new reachable distance is small, the reachable distance of the point will be updated;

- Repeat part 3 until the ordered queue is empty.

4. The algorithm ends.

The sample points with reachable distance information are output, and the cluster after clustering is the final obstacle, and the pseudo-obstacle target without height is eliminated. Output image *O* for next step 9 tracking;

Step9: Obstacle tracking. Image feature points are used as description features of moving objects. First, feature points are extracted and matched from the video images collected by the thermal imager. In the tracking module, the thermal images are processed to get *n* pairs of matched feature points. In this image, the feature points are extracted and matched with the frame, the previous frame, so as to realize the tracking of the target. The current frame of the new target is the first frame of the region, the rest is the same as above. The specific processes are as follows:

- Obtain a frame of image l_1;
- If it is the first frame image, i.e., $i=1$, the l_1 feature point is extracted and returned to 1;
- l_1 feature points are extracted and matched with l_1-1 frame feature points;
- The relevant data is calculated according to the formula to track the target;
- Update the set of target feature points.

After a new feature point appears in the target area, the feature points matching and feature points height judgment are carried out, and steps 6 to 9 are repeated.

4. Experiments and Analysis

In this section, the real vehicle experiment of thermal imaging obstacle detection based on VIDAR is carried out, and the results are analyzed. The experimental results are analysed, and compared with the traditional thermal imaging method, the detection speed and accuracy of the method are presented.

4.1 Test

4.1.1 Experimental setup: The thermal imager we used in the experiment is iray T3s, which supports USB (Universal Serial Bus) type-C connection to smart phone. Fig. 5 shows the main structure of the thermal imager.

Fig.5. iray T3sthermal imager

In the experimental section, we collect infrared data information. The device is fixed on the front bar of the smart car, which is on the same plane as the steering wheel and connects the USB cable of the thermal imager to transmit the thermal video to the computer for subsequent processing of the thermal image. The installation height of the thermal imager is 0.76m, and the inertial measurement unit (IMU) is installed at the bottom of the inspection vehicle, which can locate and obtain its own motion state information in real time. The electric power steering is used as the actuator to control the inspection vehicle to complete the actions of obstacle avoidance and lane change. The experimental vehicle is shown in Fig.6.

Fig.6. Experimental vehicle

4.1.2 Image Processing: In Fig.7, left-shown figures are the thermal images collected at a certain moment. After the second part of the steps, we obtain the processed right-shown figures.

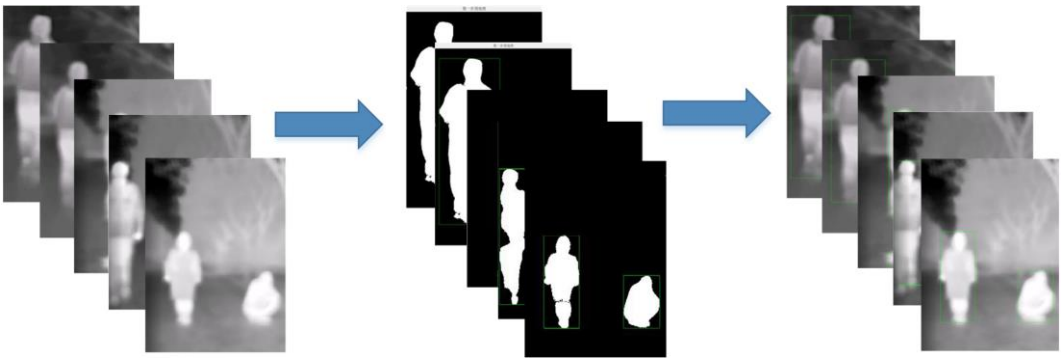


Fig.7. Thermal image processing of pedestrians with different postures

It can be found that the method of extracting target obstacles based on color features is easy to identify plane objects that do not affect driving. In order to solve this problem, we will carry on the next operation. The process successfully eliminates the pseudo-obstacles without image enhancement. Using the characteristics of VIDAR technology innovatively, the process of thermal imager detecting thermal obstacles is simplified and the accuracy of detection is guaranteed.

4.1.3 Feature point extraction and matching: In Fig.8, the feature points of the thermal image are extracted and matched, and the feature points of the two frames are drawn as scatter diagram to get the left-shown figures. The heating well cover without height is also detected as an obstacle, but it will not affect driving safety. In order to eliminate non-3d pseudo-obstacles, the next step is to determine the height of feature points. The specific feature point extraction process and feature point judgment process are shown in the second section. As shown in the middle-shown figure. After judging the height of the feature point, we eliminate the feature point without height, judge it as a pseudo-obstacle, and get the right-shown figures. After elimination, the feature points are tracked until new suspected obstacles appear, and then continue to judge.

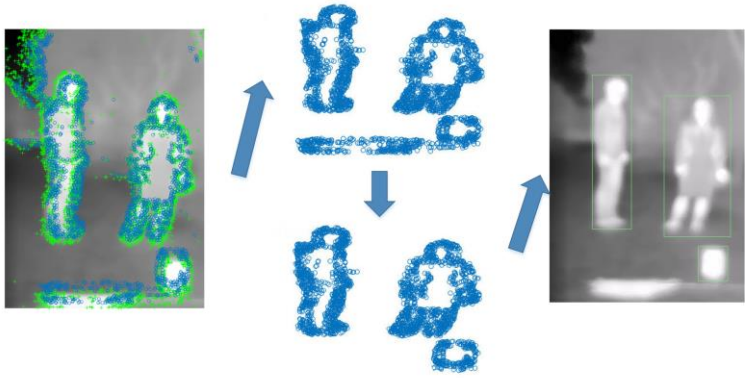


Fig.8. Removal of pseudo-obstacle

4.2. Evaluation method

Compared with the traditional thermal imaging camera to distinguish the thermal obstacles and detect the obstacles, the method in this paper is more effective in the traditional thermal imaging detection technology. Through the comparison of the number of feature points detected, it can be intuitively concluded that after elimination, the planar thermal obstacles such as heating manhole covers will not be detected and followed up.

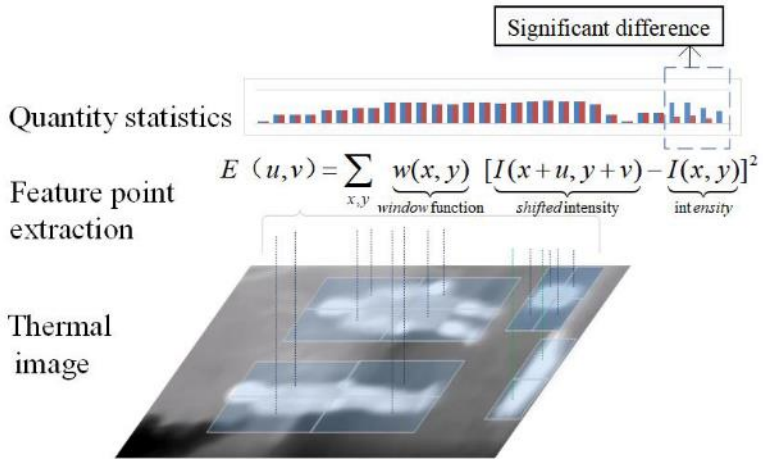


Fig.9. Visual contrast chart of feature point number.

In the statistical part of the number of feature points (see Fig. 9), the blue bar graph counts all detected feature points detection results, and red bar graph statistics feature points which have no height counts all detected feature points detection results after eliminating. This test demonstrates the accuracy and practicability of this method by showing the change in the number of feature points before and after using VIDAR to eliminate pseudo-obstacles. Specifically, in order to verify the advantages of this method in detection accuracy, the feature points of the captured video frames are detected and counted, and the matched feature points are counted to characterize the difference of detection accuracy before and after removing the pseudo-obstacles. The feature points without height are eliminated by the proposed method and the feature points removed will not be counted, according to the number of feature points on the intuitive comparison, can show that this method has advantages in the detection accuracy, can achieve the function of rapid screening of pseudo-obstacles. Compared with the traditional thermal imaging method, the detection accuracy has been improved significantly.

5. Discussion

It can be seen that a simple thermal sensor can no longer meet the needs of contemporary cars driving safety. Compared with visual sensors, thermal imaging is more accurate and efficient in detecting target obstacles at night and in low

illumination environment. For the deep learning category, the method in this paper saves the training step of the data set and the image enhancement technology and can accurately frame the target obstacle. The following uses thermal imaging combined with YOLO v5 to detect obstacles and compare with the method in this paper.

Considering that IMU is used to obtain the moving data of each camera frame, it cannot use a wide range of databases. In order to strengthen the credibility of this study, we have built our own data set: VIDAR-iray database.

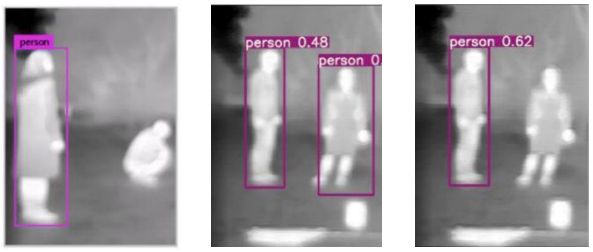
The acquisition frequency is 77 frames/ min, while there are 2544 images in total, including 1625 from short video 1.mp4, 649 from 2.mp4, 270 from 3.mp4. The VIDAR-iray database (see Fig.10) is based on real-time data, acquired by iray T3s thermal imager. Pedestrians and pets were labelled as 5332 pedestrian targets and 962 other heat source targets. The data set was configured, and the configuration file of the model was accordingly edited. The data in yolov5s.yaml were used to modify the parameters, whereas model training followed, after creation. In the last step, the self-defined dataset VIDAR-iray was finalized.

Fig.10 The VIDAR-iray database (part of data).

In this study, a total of 450 experimental pictures are used for testing, including 783 pedestrian targets and 717 effective pedestrian detections, while the detection accuracy is 91.6 %. Target misdetection is 42% and false detection rate is at 5.4 %. There were 18 missed targets, deriving a missed detection rate of 3.2 %, while the target detection speed was 0.164s. In the processing of multi-objective thermal images, the time for the proposed method to detect obstacles after a series of processing is 0.271s. The total time required, plus clustering and tracking time, is 0.302s. Although the proposed method takes more time than

YOLO v5, it eliminates the steps of training sample library and meets the real-time requirements.

However, if you want to make good use of YOLOv5, you need to input a large number of models, such as all kinds of pedestrians, vehicles, common pets, road obstacles and other types of targets. If the data set fails to meet the requirements, it will make false detection for pedestrians with special walking posture or squatting posture, which is a great threat to pedestrian safety. The experimental results show that the detection speed of YOLOv5 is slow in the detection of low pixel thermal image, and even miss detection occurs.



(a) ***(b)*** ***(c)***

Fig.12. YOLO v5 test results

(a) YOLOv5 can't detect squatting pedestrians, ***(b)*** YOLO v5 can't detect three-dimensional obstacles, ***(c)*** Detection results under different image enhancement processing

After the corresponding image enhancement of the collected thermal images, we use YOLO v5 for target recognition. In Fig. 12a, we can see that YOLO v5 cannot detect squatting pedestrians. In Fig. 12b and Fig. 12c, after different image enhancement processing, the detection effect is different. In Fig. 12b, we cannot detect obstacles with height, which have a huge impact on driving safety.

The thermal imaging model used in this paper is iray T3s, which has low cost. However, there are some problems, such as low pixels, unclear picture information and so on. The final thermal image is fuzzy, and the edge contour is not very clear. For the target detection of the YOLO series, complex image enhancement and other pre-processing must be carried out, and even the results are not satisfactory.

YOLO mechanisms need to be further using high resolution, large lens thermal imager to improve the quality of the original image, using the updated image processing algorithms improve image detail, by comparison advantage one of this article lies in, under the premise of not to image enhancement, the thermal target can be quickly detected by simple pre-treatment box, the second advantage is that no complex model training, can take with the goal of thermal target goal box, for pedestrians can quickly detect lies down posture as the goal, contrast with YOLO can greatly reduce the error checking. The third advantage is that in the case of pedestrian overlap, YOLO has a high false detection rate, and the frame

outline is not very clear. However, the proposed method has a clear definition of the appearance outline of thermal obstacles.

In a specific environment, the obstacle detection method based on VIDAR and thermal imaging in this paper has higher accuracy than YOLO v5. In thermal target detection, statistics on the number of false detection frames can provide a more intuitive comparison of the proposed method to the YOLO v5 method, in terms of accuracy.

The following figure shows the pixel distribution of the image number 3 in the experimental frame. According to the pixel level, we detect 450 frames of images, if the number of pixels of the detected target is less than 80%, the target is missing. This stipulates that, target miss detection or target false detection in the detected frames are recorded as false detection frames.

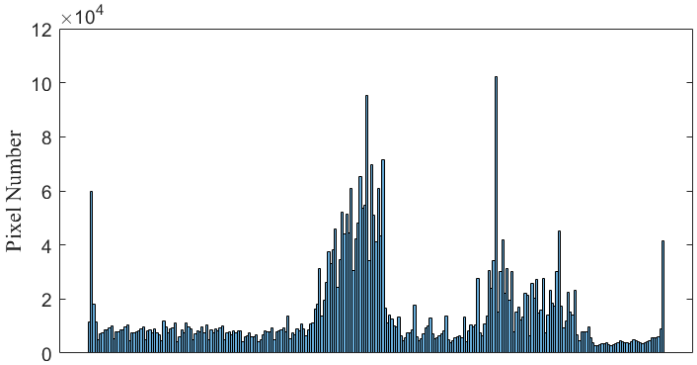


Fig. 13. Statistics of false detections in a total of 450 detected images (The abscissa is the number of detected images, while the ordinate is the number of false detected images)

The final results are as follows.

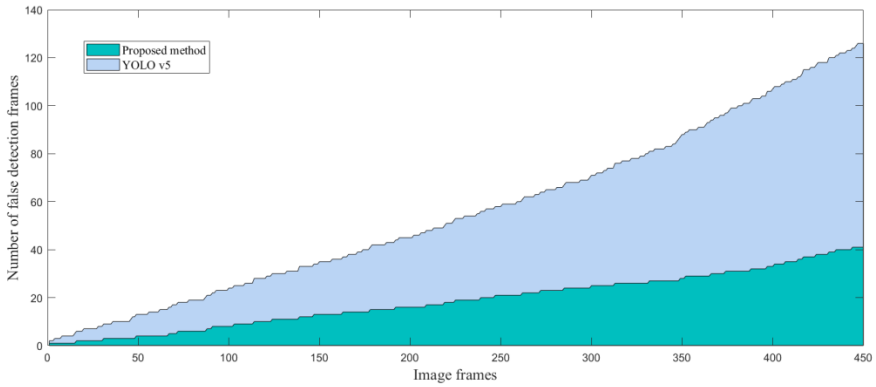


Fig. 14. Statistics of false detections in a total of 450 detected images (The abscissa is the number of detected images, while the ordinate is the number of false detected images)

Fig. 14 shows that the method of this paper has obvious advantages in the detection accuracy under specific conditions, and with the increase of the number of detection frames, the false detection rate of YOLOv5 has an obvious upward

trend, while the false detection number of this method tends to be flat and has obvious advantages compared with the YOLO method.

6. Conclusions

To sum up, the thermal imaging technology is applied to obstacle detection in low illumination environment, and the common pedestrian segmentation and matching steps are omitted. Combined with VIDAR, the obstacles without height are eliminated. In this paper, a real-time and safe detection method is provided for the detection process of thermal obstacles in low illumination environment. Combined with the research method of machine vision, the thermal obstacles are detected without using or establishing a sample database. Firstly, the obstacle information in front of the road is visually collected. According to the characteristics of thermal imaging, the position of the target thermal obstacle is obtained by image processing technology, and the pseudo obstacle in the detection process is removed by VIDAR, and finally the detection is completed to ensure the accuracy of detection in a variety of thermal obstacle environments. After experimental comparison, compared with traditional thermal imaging detection methods and YOLO v5, this method has higher recognition accuracy and stable recognition speed. It can be seen from the results that under the premise of ensuring the accuracy, the former has a significant improvement in accuracy, which also makes up for the deficiency of the future visible light camera in the low illumination environment. In this study, we compared the time and accuracy of the proposed method with YOLOv5, which saves the training steps, and has obvious contrast to the problems of pedestrian overlap, pedestrian detection in squatting posture and so on.

REFERENCES

[1] *Tyrrell, R.A., Wood, J.M., Owens, D.A., Borzendowski, S.W., and Sewall, A.S.,* "The Conspicuity of Pedestrians at Night: A Review", Clinical and Experimental Optometry, 2016, 99.

[2] *Chang, L., Nash, J., and Prior, S.*, "A Low-Cost Vision-Based Unmanned Aerial System for Extremely Low-Light Gps-Denied Navigation and Thermal Imaging", in, ICIUS 2015: International Conference on Intelligent Unmanned Systems, (2015)

[3] *Wong, H.*, "Bitmap Generation from Computer-Aided Design for Potential Layer-Quality Evaluation in Electron Beam Additive Manufacturing", Rapid Prototyping Journal, 2020, ahead-of-print, (ahead-of-print).

[4] *Yeom, S.*, "Multi-Level Segmentation of Infrared Images with Region of Interest Extraction", International Journal of Fuzzy Logic & Intelligent Systems, 2016, 16, (4), pp. 246-253.

[5] *Wang, G. and Liu, Q.*, "Far-Infrared Based Pedestrian Detection for Driver-Assistance Systems Based on Candidate Filters, Gradient-Based Feature and Multi-Frame Approval Matching", Sensors, 2015, 15, (12), pp. 32188-32212.

[6] *Bai, X., Wang, Y., Liu, H., and Guo, S.*, "Symmetry Information Based Fuzzy Clustering for Infrared Pedestrian Segmentation", IEEE Transactions on Fuzzy Systems, 2017.

[7] *Wang, P. and Bai, X.*, “Thermal Infrared Pedestrian Segmentation Based on Conditional Gan”, IEEE Transactions on Image Processing, 2019, PP, (99).
[8] *Tattersall, G.*, “Thermimage: Functions for Handling Thermal Images”, R Package Version 1.0. 1', (2015)
[9] *Jan, J.*, “Image Enhancement”, (Medical Image Processing, Reconstruction and Analysis, 2019)
[10] *Liu, F., Cheng, Z., Jia, P., Zhang, B., and Hu, R.*, “Impact of Thermal Control Measures on the Imaging Quality of an Aerial Optoelectronic Sensor”, Sensors, 2019, 19, (12), p. 2753.
[11] *Bhatnagar, G. and Wu, Q.*, “A Fractal Dimension Based Framework for Night Vision Fusion”, IEEE/CAA Journal of Automatica Sinica, 2019, 6, (01), pp. 223-230.
[12] *Wang, L., Gui, J., Lu, Z.M., and Liu, C.*, “Fast Pedestrian Detection and Tracking Based on Vibe Combined Hog-Svm Scheme”, International Journal of Innovative Computing Information and Control, 2019, 15, (6), pp. 2305-2320.
[13] *Vipparthi, S.K., Murala, S., Gonde, A.B., and Wu, Q.*, “Local Directional Mask Maximum Edge Patterns for Image Retrieval and Face Recognition”, Iet Computer Vision, 2016, 10, (3), pp. 182-192.
[14] *Yin, H.F. and Wu, X.J.*, “A New Feature Fusion Approach Based on Lbp and Sparse Representation and Its Application to Face Recognition”, International Workshop on Multiple Classifier Systems, 2013.
[15] *Qingxin, H.U. and Peng, L.*, “Pedestrian Detection in Infrared Images Based on Multi-Feature Fusion”, Journal of Computer Applications, 2016.
[16] *Huckridge, D.A., Ebert, R., Lee, S.T., Rodger, I., Connor, B., and Robertson, N.M.*, “Classifying Objects in Lwir Imagery Via Cnns”, in, Electro-optical & Infrared Systems: Technology & Applications XIII, (2016)
[17] *Tewary, S., Akula, A., Ghosh, R., Kumar, S., and Sardana, H.K.*, “Hybrid Multi-Resolution Detection of Moving Targets in Infrared Imagery”, Infrared Physics & Technology, 2014, 67, pp. 173-183.
[18] *Hui, L., Wu, X.J., and Kittler, J.*, “Infrared and Visible Image Fusion Using a Deep Learning Framework”, International Conference on Pattern Recognition 2018, 2018.
[19] *Xu, L., Liu, G., Cao, B., Zhang, P., and Liu, S.*, “Infrared Target Recognition Based on Improved Convolution Neural Network”, in, ICNCC 2019: 2019 The 8th International Conference on Networks, Communication and Computing, (2019)
[20] *Brehar, R.D., Muresan, M.P., Mariţa, T., Vancea, C.-C., Negru, M., and Nedevschi, S.*, 'Pedestrian Street-Cross Action Recognition in Monocular Far Infrared Sequences', IEEE Access, 2021, 9, pp. 74302-74324.
[21] *Altay, F. and Velipasalar, S.*, 'Pedestrian Detection from Thermal Images Incorporating Saliency Features', in, 2020 54th Asilomar Conference on Signals, Systems, and Computers, (IEEE, 2020)
[22] *Wang, Y., Wei, X., Tang, X., Shen, H., and Zhang, H.*, 'Adaptive Fusion Cnn Features for Rgbt Object Tracking', IEEE Transactions on Intelligent Transportation Systems, 2021.
[23] *Gonzalez, R.C. and Woods, R.E.*, “Digital Image Processing”, Prentice Hall International, 2008, 28, (4), pp. 484 - 486.
[24] *Kanagala, H.K. and Krishnaiah, V.*, “A Comparative Study of K-Means, Dbscan and Optics”, in, 2016 International Conference on Computer Communication and Informatics, (2016)