

Research Article

Automatic Detection and Classification of Road, Car, and Pedestrian Using Binocular Cameras in Traffic Scenes with a Common Framework

Yongchao Song ¹, Jieru Yao,^{1,2} Yongfeng Ju ¹, Yahong Jiang,³ and Kai Du¹

¹School of Electronic and Control Engineering, Chang'an University, Xi'an, China

²School of Automation, Northwestern Polytechnical University, Xi'an, China

³School of Economics and Management, Chang'an University, Xi'an, China

Correspondence should be addressed to Yongfeng Ju; yfju@chd.edu.cn

Received 15 December 2019; Revised 20 April 2020; Accepted 24 April 2020; Published 14 May 2020

Academic Editor: Guang Li

Copyright © 2020 Yongchao Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to solve the problems of traffic object detection, fuzzification, and simplification in real traffic environment, an automatic detection and classification algorithm for roads, vehicles, and pedestrians with multiple traffic objects under the same framework is proposed. We construct the final V view through a considerate U-V view method, which determines the location of the horizon and the initial contour of the road. Road detection results are obtained through error label reclassification, omitting point reassignment, and so on. We propose a peripheral envelope algorithm to determine sources of vehicles and pedestrians on the road. The initial segmentation results are determined by the regional growth of the source point through the minimum neighbor similarity algorithm. Vehicle detection results on the road are confirmed by combining disparity and color energy minimum algorithms with the object window aspect ratio threshold method. A method of multifeature fusion is presented to obtain the pedestrian target area, and the pedestrian detection results on the road are accurately segmented by combining the disparity neighbor similarity and the minimum energy algorithm. The algorithm is tested in three datasets of Enpeda, KITTI, and Daimler; then, the corresponding results prove the efficiency and accuracy of the proposed approach. Meanwhile, the real-time analysis of the algorithm is performed, and the average time efficiency is 13 pfs, which can realize the real-time performance of the detection process.

1. Introduction

With the rapid development of driverless and assisted driving technologies, autonomous vehicles should have safety functions such as obstacle collision warning, road departure warning, and speed maintenance [1]; then, they can analyze and understand the environment around them and discriminate roads, cars, pedestrians, buildings, and so on in a traffic scene [2]. Road detection is a basic task for many computer vision applications, such as road network extraction, robot autonomous navigation, global navigation satellite system (GNSS), and unmanned aerial vehicle images [3]. Vehicle and pedestrian detection and classification on the road are one of the challenges of advanced driver

assistance systems (ADAS) and are essential for traffic safety applications. Based on this, road detection [4], vehicle detection [5, 6], and pedestrian detection [7] are the key steps to realize vehicle autonomous driving technology.

As we all know, compared to monocular vision, stereo vision can provide richer information such as depth. Labayrade et al. [8] introduced stereo vision information into the traffic scene and used V view to realize obstacle detection in road environment. In recent years, stereo vision technology has been widely used in road detection, obstacle detection, and other fields [9, 10].

Stereo vision-based road detection can obtain accurate road contour estimates and provide clear path driving information, which has been studied for many years. Oh et al.

[11] proposed a road detection method combining illumination invariant algorithm and stereo vision. The disparity image and road direction with illumination invariant were estimated; the road probability map was calculated, whether each pixel belongs to the road was estimated, and then the joint bilateral filter was used to optimize the road detection result. Vitor et al. [12] proposed a road detection method that combines 2D image segmentation and 3D image processing. The features obtained by the two-dimensional segmentation technique based on the watershed transform were combined with v-disparity classification elements to form the feature descriptor of the artificial neural network, and the road area, obstacles, and nonclassified area information were estimated. Wu et al. [13] proposed a non-parametric technique road surface detection algorithm that only uses depth cues (abbreviated as the NT-RD method). Relying on the four inherent properties of the disparity image of the road scene, combined with the U view and V view, the road pavement extraction was realized. Wang et al. [14] proposed a method of combining the initial road surface obtained from road surface axis calibration in logarithmic color space and the road surface extracted from stereo vision V view to filter out false detection pixels and obtain accurate road detection results. Guo et al. [15] proposed an automatic estimation method of stereo vision homography matrix based on feature correspondence and region optimization and constructed a road boundary detection algorithm based on stereo vision homography and HMM. An observation probability function based on the state sequence was proposed to obtain the optimal boundary between road and nonroad areas. Li et al. [16] proposed a stereo vision road boundary detection method based on multicues fusion and established three Bayesian models based on the boundary region normal vector, height, and color cues. The points with the highest confidence level were fitted to the road boundary curve using the support vector regression (SVR) method. Xie et al. [17] proposed a binocular vision drivable region detection algorithm based on the cascading framework. The corrected binocular image pair disparity image was given, and the U-V view was calculated in a probabilistic manner. The road region was obtained by the RANSAC plane fitting method. Cheng et al. [18] proposed a stereo vision road edge detection method that integrates 16-dimensional descriptors such as appearance, geometry, and parallax information. The Dijkstra road boundary model with vanishing point constraints was used to search the two minimum cost paths according to the constructed cost map to obtain the curb detection results. Su et al. [19] proposed a texture voting strategy based on stereo vision V view and vision ranging technology to achieve fast vanishing point detection. Formulated lane detection problem was a graph search process. A Dijkstra minimum path lane model with vanishing point limitation was constructed to realize lane detection. Zhang et al. [20] proposed a Dijkstra road model based on vanishing point constraints to implement stereo vision road detection. A weighted sampling RANSAC sample line fitting strategy was used to detect the horizontal line. A vanishing point estimation method for horizontal line and pavement area constraints was proposed. A Dijkstra minimum cost

map with vanishing point limitation was constructed to implement road boundary segmentation.

Obstacle detection in stereo vision is an effective method for safe driving of vehicles and avoiding collisions. Reliable obstacle detection methods have been the focus of research in the industry. Wang et al. [21] proposed using the convex hull method to extract the region of interest of the obstacle, performing U-view operation on the region of interest and using the connected region extraction method to detect multiple obstacles. Fakhfakh et al. [22] proposed an improved V view that fuses confidence values, used U and V views to estimate potential obstacle boundaries, and used a weighted V-view method to detect obstacles. Kang et al. [23] proposed a probabilistic polar coordinate grid map to analyze the structural characteristics and ergodicity of the volume of the grid map, generated the nearest obstacle and a larger range of potential obstacles in each search direction, and obtained obstacle detection results. Yoo et al. [24] extracted the three features of disparity, superpixel, and pixel gradient, calculated the disparity reliability from superpixel segment and pixel gradient, and proposed the use of reliability voting, CIELAB color similarity, and distance similarity between superpixels to achieve obstacle detection. Burlacu et al. [25] proposed a multicharacterized stereo image obstacle detection framework and used multiple representations of the disparity image: V view, U view, and θ view to achieve obstacle detection.

However, the abovementioned method only performs single traffic object detection or general obstacle detection, and the obstacle detection research is oriented to the entire scene, not directly to traffic participants, and does not accurately classify multiple traffic objects. In recent years, with the rapid development of deep learning technology, many methods of applying deep learning have appeared to realize the detection of traffic objects. Han et al. [26] proposed road detection methods based on GANs semisupervised and CGANs weak supervised learning. The method uses a large number of labeled and unlabeled images for training. Dairi et al. [27] proposed the problem of urban scene monitoring and obstacle tracking based on unsupervised deep learning methods. A hybrid encoder integrating a deep Boltzmann machine (DBM) and an automatic encoder (AE) is designed. It combines the greedy learning characteristics of DBM and the dimensionality reduction capability of AE and can accurately and reliably detect the existence of obstacles. Unfortunately, most deep learning-based approaches have high execution time and hardware requirements and are expensive and costly, and only a few of them are suitable for real-time applications [28, 29].

In contrast, our approach can deal with the above problems. Based on the above, we propose to implement the classification detection of roads, vehicles, and pedestrians under the same framework. Bicubic interpolation is used to obtain the corrected stereo disparity image. The considerate U-V view method is used to determine the initial contour of the road and obtain the road detection results. We propose a peripheral envelope algorithm to obtain the area of interest on the road and determine the source points of vehicles and pedestrians. Based on the vehicle's good color similarity and

disparity similarity, we construct a minimum energy algorithm to complete the vehicle information and extract the vehicle using the aspect ratio threshold method. A method of multifeature fusion such as aspect ratio, perspective ratio, and area ratio is proposed to obtain the pedestrian target area, and the neighborhood target similarity and energy minimization algorithm is used to accurately extract and segment the pedestrian.

1.1. Contributions. Briefly, this paper makes the following some main contributions:

- (i) For the complex problem of traffic multiobject classification detection in traffic environment perception, we propose to realize the automatic detection and classification of roads, vehicles, and pedestrians under a common framework, which avoids the detection of a single traffic object or the detection of general obstacles and improves the pertinence of detection objects.
- (ii) A new disparity image correction method is proposed to provide conditions for accurate classification and detection of subsequent traffic objects such as roads, vehicles, and pedestrians.
- (iii) A considerate U-V view method is proposed to obtain the final V view and the initial contour of the road, which avoids the incorrect estimation of the road contour caused by the traditional straight-line fitting method for the V view of uphill and downhill sections.
- (iv) A new method for detecting and classifying vehicles and pedestrians on the road is proposed. A peripheral envelope algorithm is used to obtain the source points of vehicles and pedestrians on the road. Multifeature fusion and threshold segmentation methods are used in combination with the minimum energy and similarity algorithm to achieve classification detection.

1.2. Organization. The rest of the paper is organized as follows. Section 2 describes the stereo vision road scene model applied by the method in this paper. Section 3 introduces the process of obtaining the initial contour of the road by the proposed considerate U-V view method. Section 4 describes the detailed steps of multitraffic object road, vehicle, and pedestrian classification detection proposed in this paper. Section 5 mainly presents the experiments, including datasets, standards, compared methods, results, and necessary discussions. Finally, Section 6 concludes this paper and describes some future research directions.

2. Stereo Vision Road Scene Model

The stereo cameras are installed on the same plane, and the horizontal axes of the two cameras are on the same line and have the same parameters [30]. Assume that the tilt angle of the camera relative to the vertical plane is θ , the distance between the optical centers of the stereo camera is b , the focal

length of the left and right cameras is f , the height of the camera from the horizontal plane is h , and a point in the road scene is $p(x, y, z)$. The points projected to the left and right image planes are $p_l(x_l, y_l)$ and $p_r(x_r, y_r)$, respectively, as shown in Figure 1.

The center of the two cameras is the origin of the world coordinate system. Because the left and right cameras have relative displacement only in the horizontal direction, then

$$b = |O_l - O_r|, \quad (1)$$

$$\frac{x_l}{f} = \frac{X + b/2}{Z}, \quad \frac{x_r}{f} = \frac{X - b/2}{Z}. \quad (2)$$

Disparity in 3D coordinates:

$$d = x_l - x_r = f \frac{b}{Z}. \quad (3)$$

(u_0, v_0) is the projection coordinate of the optical center of the camera. According to the imaging principle, the formula for projecting a point in the world coordinate system onto the left and right camera imaging planes is

$$\begin{cases} u_{l,r} = u_0 + f \frac{X \pm b/2}{Y \cdot \sin \theta + Z \cdot \cos \theta} \\ v = v_0 + f \frac{Y \cdot \cos \theta - Z \cdot \sin \theta}{Y \cdot \sin \theta + Z \cdot \cos \theta} \end{cases} \quad (4)$$

The disparity value corresponding to each pixel constitutes a disparity image. Given the value $\Delta = u_l - u_r$, we can obtain disparity:

$$\Delta = f \frac{b}{Y \cdot \sin \theta + Z \cdot \cos \theta} \quad (5)$$

3. Road Initial Contour Detection by Considerate U-V View

Stereo vision-based traffic object detection has significant advantages. It does not require prior knowledge and model building of traffic objects and is not sensitive to background changes caused by weather changes such as shadows, lighting, and reflections. At the same time, U and V views are the basic method of stereo vision processing. How to process the initial U and V views to meet specific traffic objects: road, vehicle, and pedestrian detection is one of the research focuses in this paper.

3.1. Disparity Image Acquisition and Correction. We use the ‘‘SemiGlobal’’ [31] disparity estimation algorithm to obtain disparity images, as shown in Figure 2(a). In stereo vision technology, due to uncontrollable factors such as occlusion, complex textures, and reflective lights, the pixels of the reference image cannot match the corresponding points in the target image, which causes invalid matching points in the disparity image that cannot obtain depth information and affect subsequent detection algorithm. An improved bicubic interpolation algorithm is proposed to reassign invalid matching points. Combined with the actual traffic scene, the

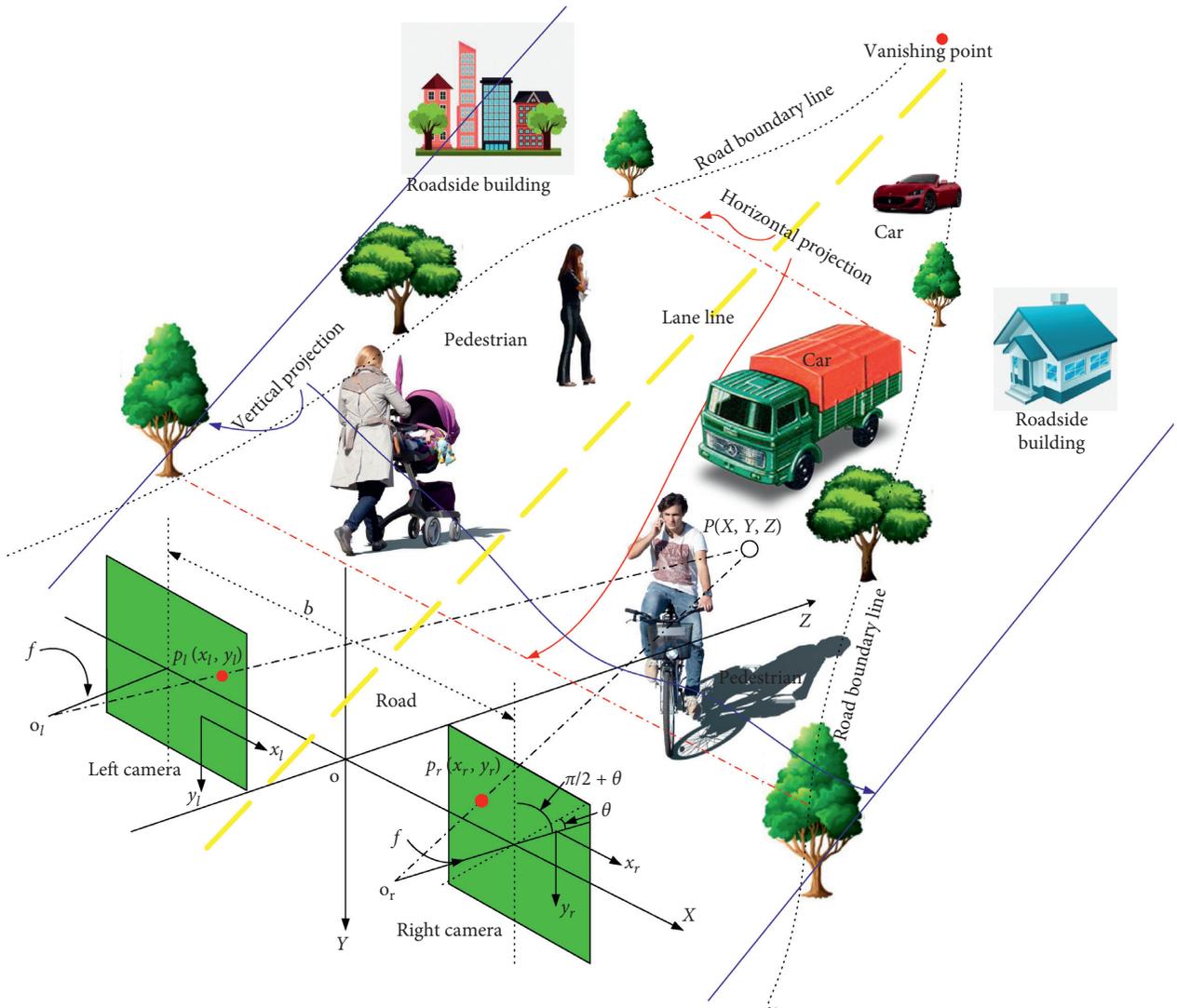
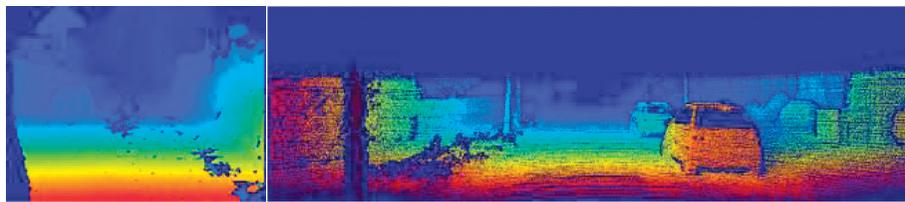
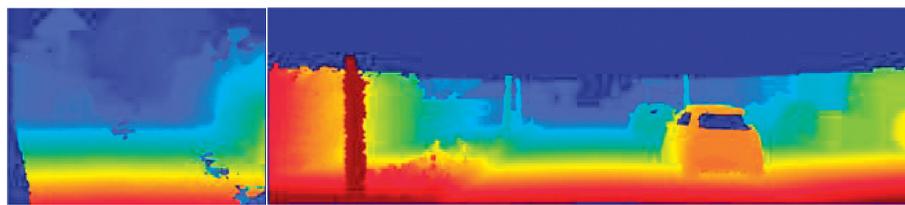


FIGURE 1: Stereo vision road scene model.



(a)



(b)

FIGURE 2: (a) The original disparity image and (b) the disparity image of interpolation revised.

bicubic interpolation kernel function is redefined, the nuclear response range is expanded, and a 5×7 convolution kernel is used. At the same time, the disparity values of all invalid matching points are set to 0, and the influence of the invalid matching points on the correction result of the disparity image is removed. The interpolation basis function is expressed as follows:

$$S(x) = \begin{cases} 1 - (\lambda + 3) \cdot |x|^2 + (\lambda + 2) \cdot |x|^3, & (|x| < 2), \\ -(4 \cdot \lambda) + (8 \cdot \lambda) \cdot |x| - (5 \cdot \lambda) \cdot x^2 \\ \quad + \lambda \cdot |x|^3, & (2 \leq |x| < 4), \\ 0, & (|x| \geq 4). \end{cases} \quad (6)$$

The target pixel value $f(i + u, j + v)$ is obtained by the following bicubic interpolation formula:

$$f(i + u, j + v) = [A] * [B] * [C]. \quad (7)$$

Among them,

$$[A] = [S(u + 2)S(u + 1)S(u)S(u - 1)S(u - 2)], \quad (8)$$

$$[B] = \begin{bmatrix} f(i - 2, j - 3) & \cdots & f(i - 2, j + 3) \\ \vdots & \ddots & \vdots \\ f(i + 2, j - 3) & \cdots & f(i + 2, j + 3) \end{bmatrix}, \quad (9)$$

$$[C] = [S(v + 3)S(v + 2) \cdots S(v - 2)S(v - 3)]^T. \quad (10)$$

The disparity image p after interpolation correction obtained by equation (7) is shown in Figure 2(b).

3.2. Considerate U-V View. We propose a considerate U-V view method, including obstacle removal on the initial U and V views, inverse view transformation, approximate classification, minimum distance calculation, and noise removal, which constructs the final V view, obtains the initial contour of the road and the horizon position, and provides a basis for the detection of specific traffic objects (Figure 3).

3.2.1. Stereo Vision Road Scene Characteristics. In a stereoscopic road scene, the road surface is on a horizontal plane. Pedestrians and vehicles on the road may be in the vertical and oblique vertical planes directly in front of them or on the vertical planes on the sides of the road. The buildings and trees are on the vertical sides of the sides of the road. The specific scene is shown in Figure 1.

The stereo vision road scene features are as follows:

- (a) The road is from near to far; the farther away from the camera, the smaller the corresponding road scene disparity value and the approximately uniform change.
- (b) Obstacle has approximately equal disparity value on the vertical plane. U view mainly reflects the characteristics of obstacle, including the size and range of obstacles. The road has approximately equal

disparity value on the horizontal plane. The V view mainly reflects the characteristics of the road, including road extension and road shape.

- (c) After the larger obstacles (buildings and trees) are removed by U view, the road is the main part of the current scene. After the road is determined, the corresponding area of interest is obtained, and the traffic object is on it.
- (d) In the captured image, when the obstacle and the road point are projected on the same line, the distance from the obstacle to the camera is smaller than the distance from the road point to the camera.

3.2.2. Final V View. The road surface can be divided into straight, uphill, and downhill sections according to the shape of the terrain profile. Therefore, the road surface can be projected as a straight line or a curve in the V view. Obviously, how to obtain the final V view is a necessary step to obtain the initial contour of the road and implement road detection.

A more detailed description of the final V disparity image acquisition algorithm is presented in Algorithm 1:

3.2.3. Horizon. The horizon is the dividing line between the road and background. It points to the end of the road, and the accuracy of its position will directly affect the segmentation effect of the road. Pixels on the horizon are infinitely far from the camera, so disparity value is zero. The horizon position can be obtained from the terminal of the point in the V view and the intersection of the obstacle lines in the U view. This article uses the final V view to stop extending to the left; the final number of rows, that is, the number of rows with disparity value of 0 is determined as the horizon.

The specific position of the horizon can be formulated as follows:

$$HL = x(v)_{\Delta_v=0}, \quad (11)$$

where HL represents the horizon of the original image and x represents the number of lines in the V view.

4. Multiple Traffic Object Classification Detection

Roads, vehicles, and pedestrians constitute the three most basic elements of road traffic scenarios. Research on them can provide basic guarantees for the management and control of urban traffic. It is of great significance for alleviating traffic congestion, reducing traffic accidents, ensuring travel safety and efficiency, and improving the intelligent construction of urban traffic. We propose to realize the classification detection of roads, vehicles, and pedestrians under the same framework in the stereoscopic traffic scene. The overall flow chart of the algorithm is shown in Figure 4.

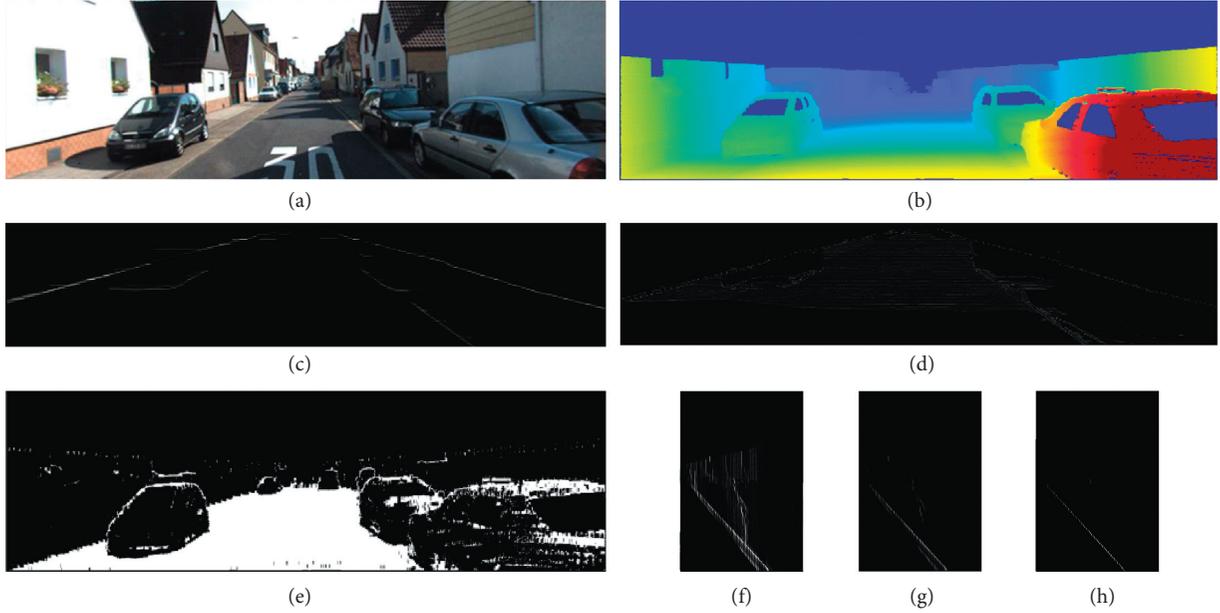


FIGURE 3: Considerate operation details of U and V views. (a) Original image. (b) Modified disparity image. (c) Initial U view. (d) U view after removing obstacles (enlarged). (e) Disparity image after removing obstacles. (f) Initial V view. (g) V view after removing obstacles. (h) Final V view.

Require: DI : Disparity image, Δ_{\max} , Δ_{\min} : Disparity values corresponding to the maximum and minimum intensity points for each line, Δ_{v_i} : Disparity value of the v th row in the V view, t_d : Disparity threshold, λ_d : Euclidean distance, S : Represents the cumulative sum of the Euclidean distances of the disparity values of adjacent rows in the V view;

Ensure: Final V view;

- (1) **for** each frame
- (2) Original DI performs horizontal and vertical projection to obtain initial U and V views;
- (3) Remove large obstacles in the initial U view;
- (4) The new U view is inversely transformed into a new disparity image and a new V view is obtained. The disparity value range of each row of the new V view is $(\Delta_m - t_d, \Delta_m + t_d)$;
- (5) **While** $(\Delta_v \neq 0)$ **do**
- (6) The new V view is scanned progressively from bottom to top, removing disparity value points that do not meet the $\min(\Delta_{v+1}) \leq \max(\Delta_v) \leq \max(\Delta_{v+1})$ condition;
- (7) **if** the new V view still has noise
- (8) Find the line where the noise is located, make the point where the disparity point of this line meets $S = \sum_{v(\Delta_v, \Delta_{v-1})} \lambda_d(\Delta_v, \Delta_{v-1})$ as the reference point of the road, and update the range of road deviation where the noise is located to $\Delta_{\min} \pm t_d$;
- (9) **end if**
- (10) Obtain the Final V view using the considerate U-V view method.
- (11) **end while**
- (12) **end for**
- (13) return "Final V view".

ALGORITHM 1: Road peripheral enveloping algorithm.

4.1. Road Detection. The three road shapes of straight line section, uphill section, and downhill section present simple straight shapes in the V view. Therefore, the traditional Hough transform [32] and RANSAC [33] straight line fitting can no longer accurately reflect the road information. We combine the final V view and horizon position obtained by the considerate U-V view method, as shown in Figure 5(b). Rough detection of the road surface is achieved

through inverse transformation to obtain the initial contour of the road, as shown in Figure 5(c), but there are still spot spots that are misclassified or missing. Therefore, we propose to reclassify road spots that may be misjudged or missed based on the initial contour of the road by reclassifying the wrong class and reassigning missing points to obtain road detection results. The specific method is as follows:

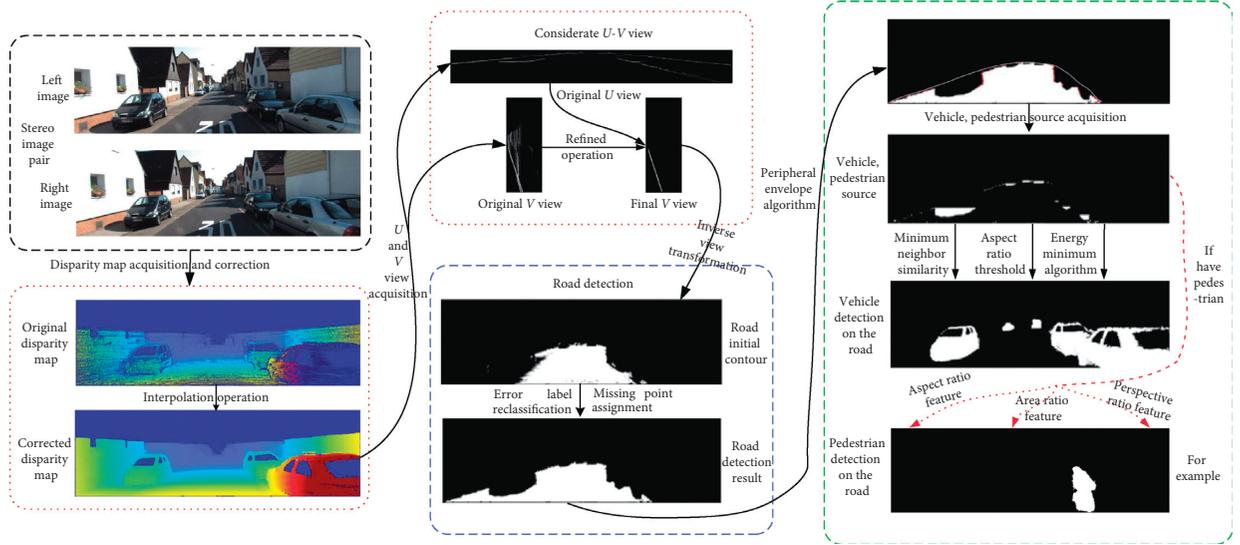


FIGURE 4: The overall flow chart of the algorithm.

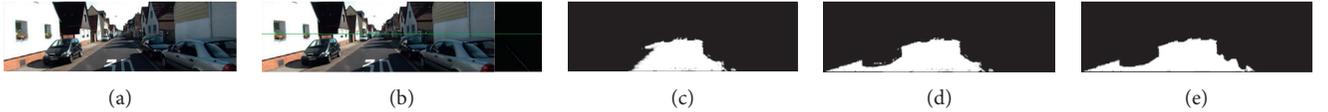


FIGURE 5: Road detection algorithm. (a) Original image. (b) Horizon position. (c) Road initial contour. (d) Road point completion. (e) Detection result.

Step 1: the initial contour of the road should be flat and continuous. However, due to the slight fluctuations of the road surface, there is a small area of nonroad category on the road surface. Use the following equation (12) to reclassify the nonroad category of small area:

$$c(x, y, l) = \frac{1}{25} \sum_{i=-2}^2 \sum_{j=-2}^2 |l(x, y) - l(x+i, y+j)|. \quad (12)$$

Among them, l is the class label at coordinates (x, y) in the binary image obtained in (1) (as shown in Figure 5(d)). When the point is off-road, $l = 0$; when the point is the road, $l = 1$. The class label value on each (x, y) should ensure that the value of c is minimal.

Step 2: after the reclassification of the error class, the small area is completed, and there may also be the error classification point of the larger area. We calculate the initial position and length of the continuous class of each row to achieve reassignment of larger misclassification points:

$$\begin{aligned} &\text{if } l_{n-1} = l_{n+1} \neq l_n \text{ and } D_n < t_e, \\ &\text{then } l_{n-1} = l_{n+1} = l_n. \end{aligned} \quad (13)$$

Among them, l_n represents the n th region with continuous class labels in a row and t_e represents the error classification point threshold. When the n th consecutive label area in a row is inconsistent with the left and right areas, and the length D_n of the n th consecutive label area is less than t_e , the label of this area is reassigned to the same label as the left and right area labels. The road detection result obtained after reclassifying the misjudgment points by the abovementioned method is shown in Figure 5(e).

4.2. Vehicle Detection on the Road. The vehicle travels on the road and occupies the inherent range of the road. The acquisition of road contour is the basis of vehicle detection on the road. The algorithm process of vehicle detection on the road proposed in this paper is

Step 1: determine the road detection result as the region of interest, which contains the source points I_p required for vehicle detection on the road. In order to obtain the peripheral contour of the road area, a road peripheral contour estimation algorithm is proposed and named as the peripheral envelope algorithm. The detailed implementation process is described in Algorithm 2. The estimation result is shown in Figure 6(b), and the vehicle source point I_p is shown in Figure 6(c).

Step 2: according to the disparity range of I_p and the similarity algorithm between neighbors, I_p performs region growth. First, select the points in I_p as the initial

Require: ROI: Road area detection results, t : number of valid lines for road detection;

Ensure: Road peripheral contour;

(1) **While** ($t \notin \varphi$) **do**

(2) Obtain the coordinate point sets $pl_i = (x_{l_i}, y_{l_i})$ and $pr_i = (x_{r_i}, y_{r_i})$ of the left and right edges of the ROI, and count the number of left and right coordinate points t_1 and t_2 ;

(3) **While** $i < t_1, t_2$ **do**

(4) Find the maximum values $\max(x_{l_i})$ and $\max(x_{r_i})$ in pl_i and pr_i , and use their corresponding coordinate points as the initial coordinate points pl_0 and pr_0 ;

(5) According to the vector outer product formula: $|c| = |a| * |b| * \sin \theta$, with pl_0 and pr_0 as the starting points, scan other coordinate points pl_i and pr_i , respectively, and sort them in ascending order from the longitudinal distance from their respective starting points;

(6) The obtained vertices are sequentially connected to obtain the road outline.

(7) **end while**

(8) **end while**

(9) return "Road peripheral contour".

ALGORITHM 2: Road peripheral enveloping algorithm.



FIGURE 6: The detection algorithm of vehicle on the road. (a) Original image. (b) Road periphery envelope. (c) Vehicle source point. (d) Detection result.

seed points and search its 8 neighborhoods. For the pixels in the range of $[\Delta_s - t_v, \Delta_s + t_v]$, consider it as new seed points (where t_v is the vehicle threshold and Δ_s is the disparity value of the seed point), iteratively search on this principle until no disparity value is found that meets the rules.

Step 3: after the initial segmentation of the vehicle according to Step 2, it is then accurately segmented by the minimum energy algorithm:

$$E = E_d + E_c. \quad (14)$$

Among them,

$$E_d(x, y) = \sum_{\forall (\Delta_x, \Delta_y) \in O} \lambda_d(\Delta_x, \Delta_y). \quad (15)$$

$E_d(x, y)$ represents the disparity similarity between two adjacent pixels, the disparity on the same object should not be much different, O represents an obstacle, (Δ_x, Δ_y) represents the set of all adjacent pixels, and $\lambda_d(\Delta_x, \Delta_y)$ represents the disparity distance between adjacent pixels.

$$E_c(x, y) = \sum_{\forall (C_x, C_y) \in O} 1 - \frac{C_x}{\|C_x\|} \cdot \frac{C_y}{\|C_y\|}. \quad (16)$$

$E_c(x, y)$ represents the color similarity of two adjacent pixels, the colors of the same object should have similarity, and C_x and C_y are color vectors of adjacent pixels.

Step 4: for the object aggregation region obtained in Step 3, determine the final vehicle detection result on the road by setting the object aspect ratio threshold, as shown in Figure 6(d).

4.3. Pedestrian Detection on the Road. Similar to the initial steps of the vehicle detection method on the road, that is, the human source point on the road is determined by the contour of the road area, as shown in Figure 7(b). For pedestrian detection on the road, we propose to use a multifeature fusion method to obtain the pedestrian target area and combine the disparity neighbor similarity algorithm and the energy minimum algorithm to accurately extract and segment the pedestrian target area. Multifeature fusion target extraction methods mainly use features such as aspect ratio, perspective ratio, and area ratio.

4.3.1. Aspect Ratio R_{lw} . Different objects on the road have their corresponding fixed size aspect ratios. Because pedestrians are flexible and show different postures, there will be a small range of fluctuations in the aspect ratio of pedestrians on the road. The aspect ratio can be defined as follows:

$$R_{lw} = \frac{P_l}{P_w}, \quad (17)$$

where P_l represents the height of pedestrian on the road circumscribed rectangle and P_w represents the width of pedestrian circumscribed rectangle.

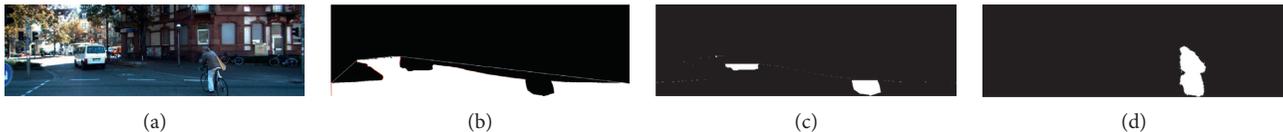


FIGURE 7: The detection algorithm of pedestrian on the road. (a) Original image. (b) Road periphery envelope. (c) Pedestrian source point. (d) Detection result.

4.3.2. *Perspective Ratio R_p* . The width of the road surface gradually narrows from near to far and intersects the horizon. Traffic objects on the road are distributed in different parts of the image, and the width of pedestrians gradually decreases with the depth of field. The pedestrian perspective R_p can be defined as follows:

$$R_p = \frac{P_l}{R_w}, \quad (18)$$

where R_w represents the width of the road surface corresponding to the row where the center coordinates of the pedestrian circumscribed rectangle are.

4.3.3. *Area Ratio R_a* . Pedestrians swing their arms or wave their arms when they walk, and they cannot be accurately detected as pedestrians on the road using only aspect ratio and perspective ratio features. We fuse the area ratio feature and define the area ratio as the ratio of the actual number of pedestrian pixels detected to the total number of pixels wrapped around the pedestrian rectangle. The specific formula is

$$R_a = \frac{\int A_c(x, y) ds}{P_l \cdot P_w}, \quad (19)$$

where $A_c(x, y)$ represents the contour curve of the pedestrian on the road.

After using the multifeature fusion method to obtain the pedestrian target area, due to the existence of defects and incomplete areas, it is proposed to use the neighborhood similarity and energy minimum algorithm to accurately segment the pedestrian target to determine the final pedestrian detection result, as shown in Figure 7(d).

5. Experiment

In this section, we use multiple experiments to evaluate the performance of the traffic object classification detection system proposed in this paper. We selected and tested multiple image sequences including different road scenes, different weather conditions, different city streets, and suburban roads:

- (1) Enpeda dataset [34] is a synthetic stereo vision dataset, which contains 496 pairs of traffic scene images with a resolution of 640×480 . The traffic scene in the dataset is relatively simple. Roads are divided into two types: planar and nonplanar, and there are more vehicles.
- (2) KITTI dataset [35] is a real-life stereo image dataset that contains 194 pairs of discontinuous traffic scene

images with different resolutions. The dataset has complex traffic scenes and contains occluded areas.

- (3) Daimler dataset [36] is a real-life stereo image dataset with a large number of images, over 21,790 pairs, collecting 27-minute continuous urban road traffic images with a resolution of 640×480 . The dataset contains a wealth of traffic objects such as roads, vehicles, and pedestrians, including continuous and complex road conditions.

The original images, disparity images, and interpolation-corrected disparity images of each standard dataset are shown in Figure 8, where Figure 8(a) is the original image, Figure 8(b) is the original disparity image, and Figure 8(c) is the corrected disparity image.

5.1. *Road Detection Results*. In this paper, the road surface is obtained by the proposed algorithm. It is experimentally determined that the threshold t_d of Enpeda dataset is 3 and KITTI and Daimler datasets is 5. For the threshold t_e for removing misclassification points, the experimental setting is 3. According to our proposed algorithm and NT-RD algorithm, experiments are performed on three standard datasets, and the road detection results of some typical sections are shown in Figure 9.

Figures 9(a) and 9(b) are the road detection results of the selected Enpeda dataset, Figure 9(a) is the uphill section, Figure 9(b) is the downhill section, and Figure 9(b) the road section contains vehicles on the road. Figures 9(c) and 9(d) are the road detection results of the selected KITTI dataset, and they are all horizontal sections. The road conditions in Figure 9(c) are complicated, including multiple vehicles, and the road is relatively congested. Figures 9(e) and 9(f) are road detection results of the selected Daimler dataset, Figure 9(e) is a downhill section with pedestrians on the road, and Figure 9(f) is an uphill section with vehicles on the road and a large pavement area. The roads in Figures 9(b)–9(d) contain a lot of shadows, and the roads in Figures 9(e) and 9(f) have traffic markings.

It can be seen from Figure 9 that, in the Enpeda dataset, the detection results of our method and the NT-RD method are basically the same. In the KITTI and Daimler datasets, the detection results of our method are significantly better than the NT-RD method. For different road environments, our algorithms are robust and can successfully detect road surfaces.

Considering the large number of images in the standard dataset, for quantitative analysis, we use a random sampling method to randomly extract 100 pairs of images from each of the three standard datasets to mark, use the proposed

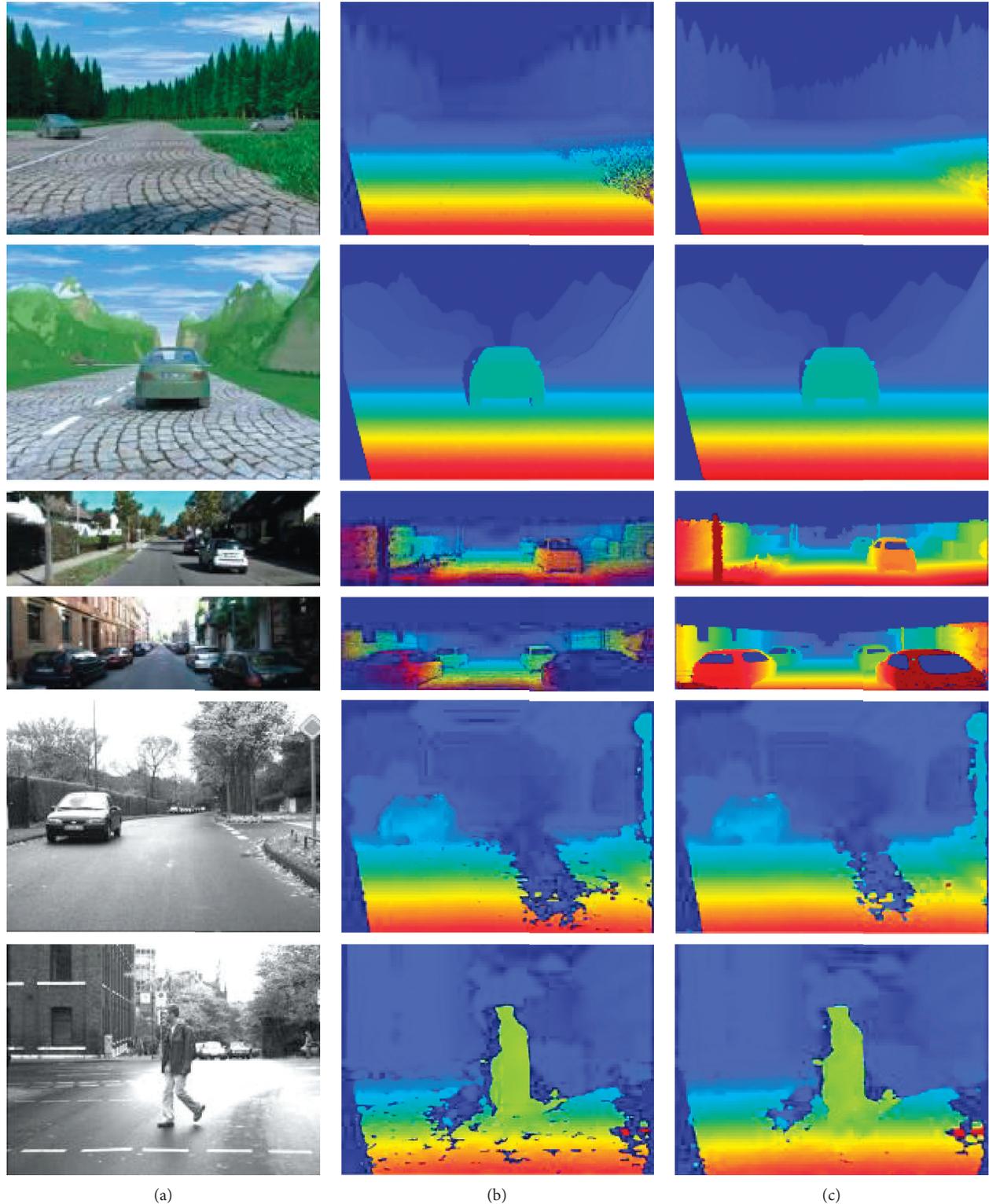


FIGURE 8: Comparison of disparity figure before and after correction, from top to bottom is Enpeda, KITTI, and Daimler dataset, respectively. (a) Original image. (b) Original disparity image. (c) Corrected disparity image.

method for road surface detection and analysis, and randomly repeat sampling five times. The following performance indexes are defined for quantitative analysis of the road detection results: (1) accuracy rate P , (2) recall rate R ,

and (3) comprehensive evaluation index F . Let ζ be the real area of the road and ξ be the actual detection result of the road; then, $P = \sum (\zeta \cdot \xi) / \sum \xi$, $R = \sum (\zeta \cdot \xi) / \sum \zeta$, and $F = 2PR / (P + R)$, $PRF \in [0, 1]$. P and R are

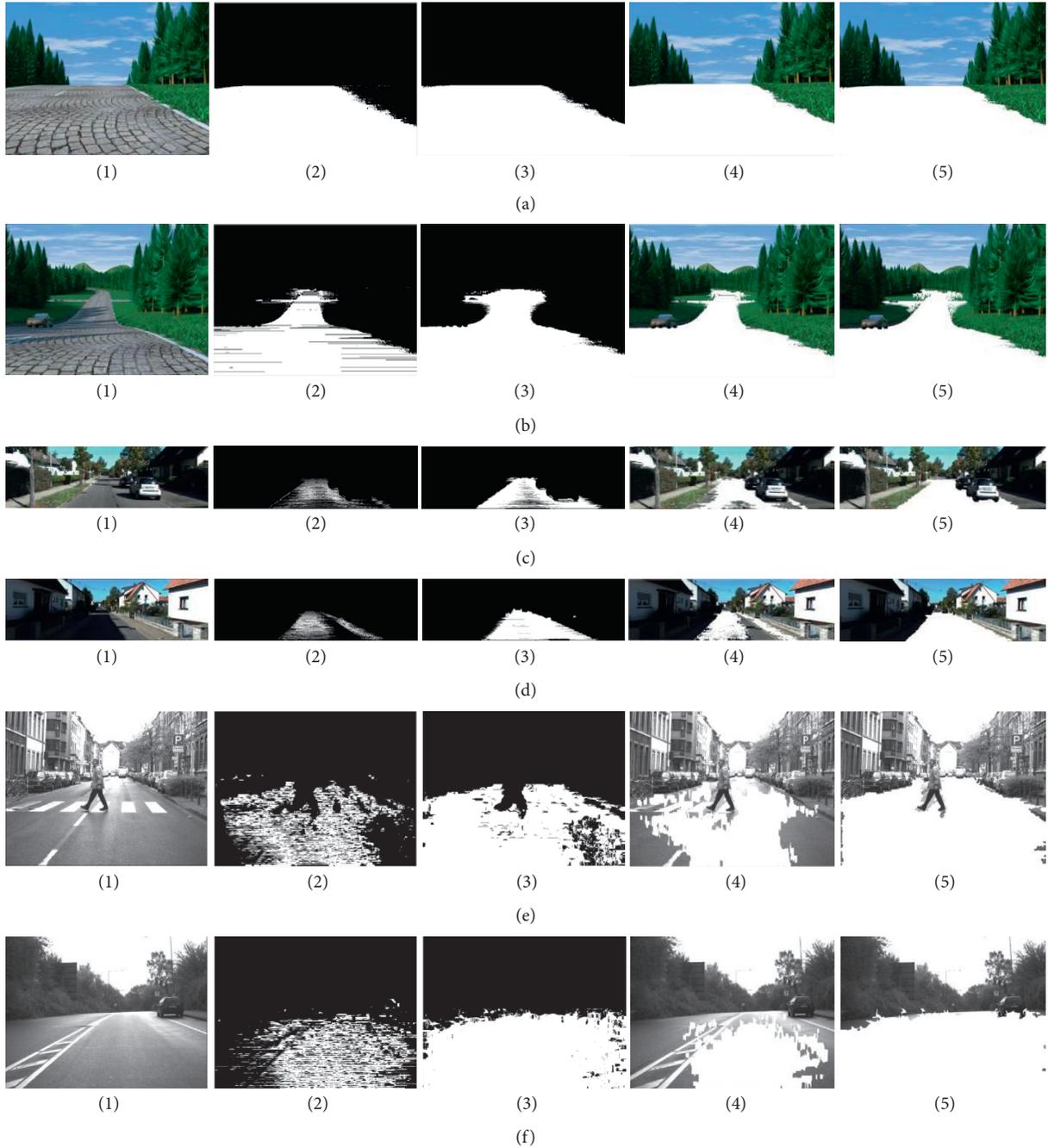


FIGURE 9: Typical road pavement detection results. (1) Original image (left). (2) Initial road contour for NT-RD method. (3) Initial road contour for ours. (4) Road detection result for NT-RD method. (5) Road detection result for ours.

complementary. F is the weighted harmonic average of P and R and can reconcile and comprehensively reflect the accuracy rate P and the recall rate R . The closer F is to 1, the better the road detection is. Due to the random sampling method in this article, there are slight fluctuations in the P , R , and F values, and the standard deviation of the fluctuations is σ . The comparison of the road surface performance indexes between ours and NT-RD method is shown in Table 1. The comparison of the comprehensive evaluation index F of the detection results is shown in Figure 10.

It can be seen from Table 1 that, in the Enpeda, KITTI, and Daimler datasets, the performance indexes P , R , and F of our method are significantly higher than the NT-RD method. It can be seen from Figure 10 that the average value of the comprehensive evaluation index F of the proposed method is large and the fluctuation is small. The experimental results show that the results of the proposed road detection algorithm can achieve the expected results and are significantly better than the NT-RD method in terms of P , R , and F .

TABLE 1: Road detection algorithm performance comparison of our and NT-RD method.

Dataset	Method	$P \pm \sigma$	$R \pm \sigma$	$F \pm \sigma$
Enpeda	NT-RD [†]	0.9603 ± 0.0741	0.9132 ± 0.0390	0.9362 ± 0.0511
	Ours [‡]	0.9847 ± 0.0217	0.9662 ± 0.0173	0.9754 ± 0.0193
KITTI	NT-RD	0.8816 ± 0.1264	0.9323 ± 0.0968	0.9062 ± 0.1096
	Ours	0.9372 ± 0.0389	0.9637 ± 0.0255	0.9503 ± 0.0308
Daimler	NT-RD	0.9106 ± 0.0832	0.9331 ± 0.1046	0.9217 ± 0.0927
	Ours	0.9335 ± 0.0568	0.9418 ± 0.0874	0.9376 ± 0.0689

[†] is algorithm used in Reference [10]. [‡] is algorithm proposed in this paper.

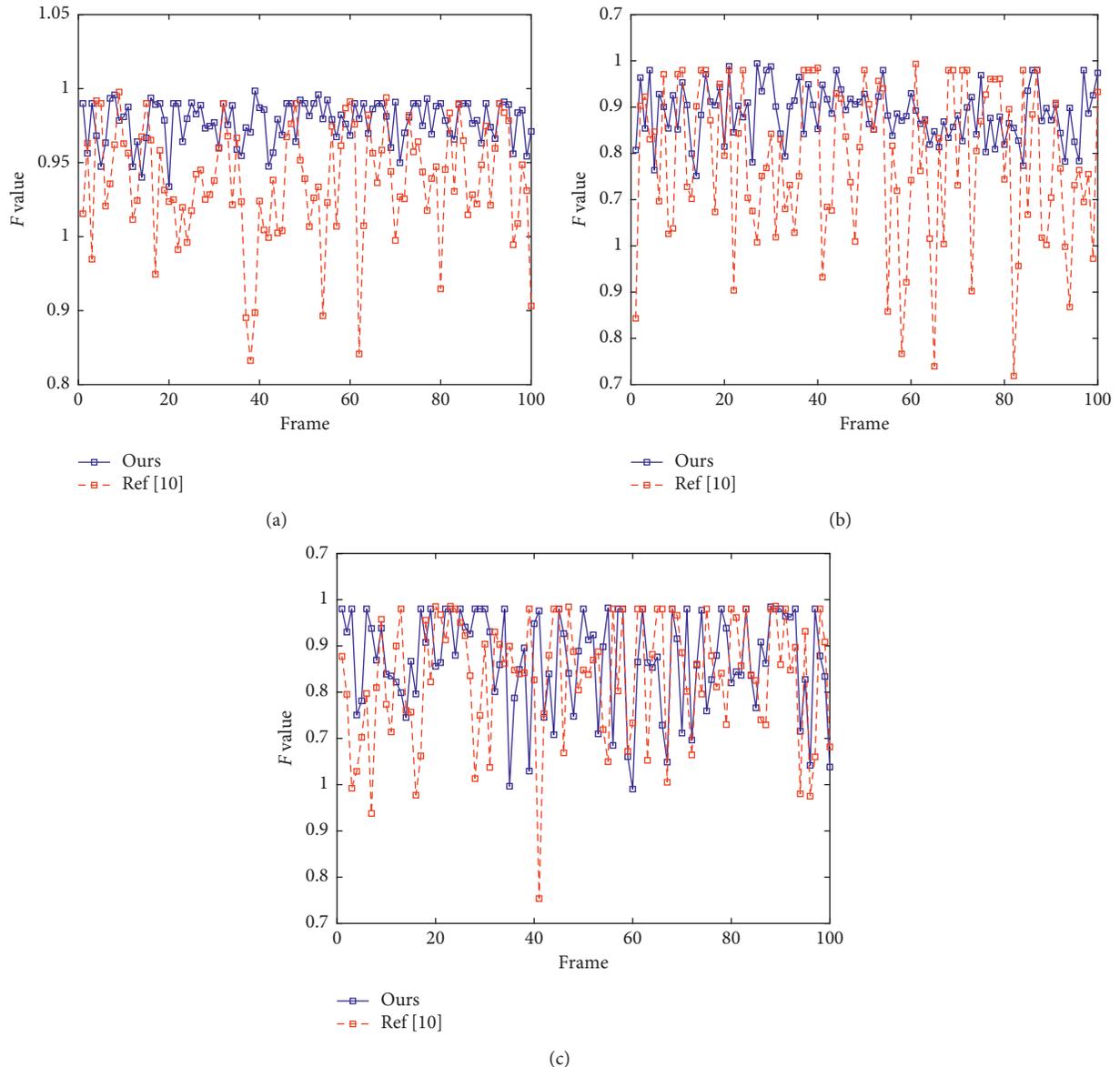


FIGURE 10: Comprehensive evaluation index F curve comparison of three datasets. (a) F value comparison of Enpeda dataset. (b) F value comparison of KITTI dataset. (c) F value comparison of Daimler dataset.

5.2. *Vehicle Detection Results on Road.* It is proposed to perform vehicle detection on the road based on the results of road detection. In the experiment, the vehicle threshold t_v is set to 5, and the vehicle aspect ratio threshold range is 0.5 to 2.

Considering that binocular vision has not been used for accurate vehicle detection in related studies, this paper will separately explain the accuracy of the proposed vehicle detection algorithm from qualitative and quantitative

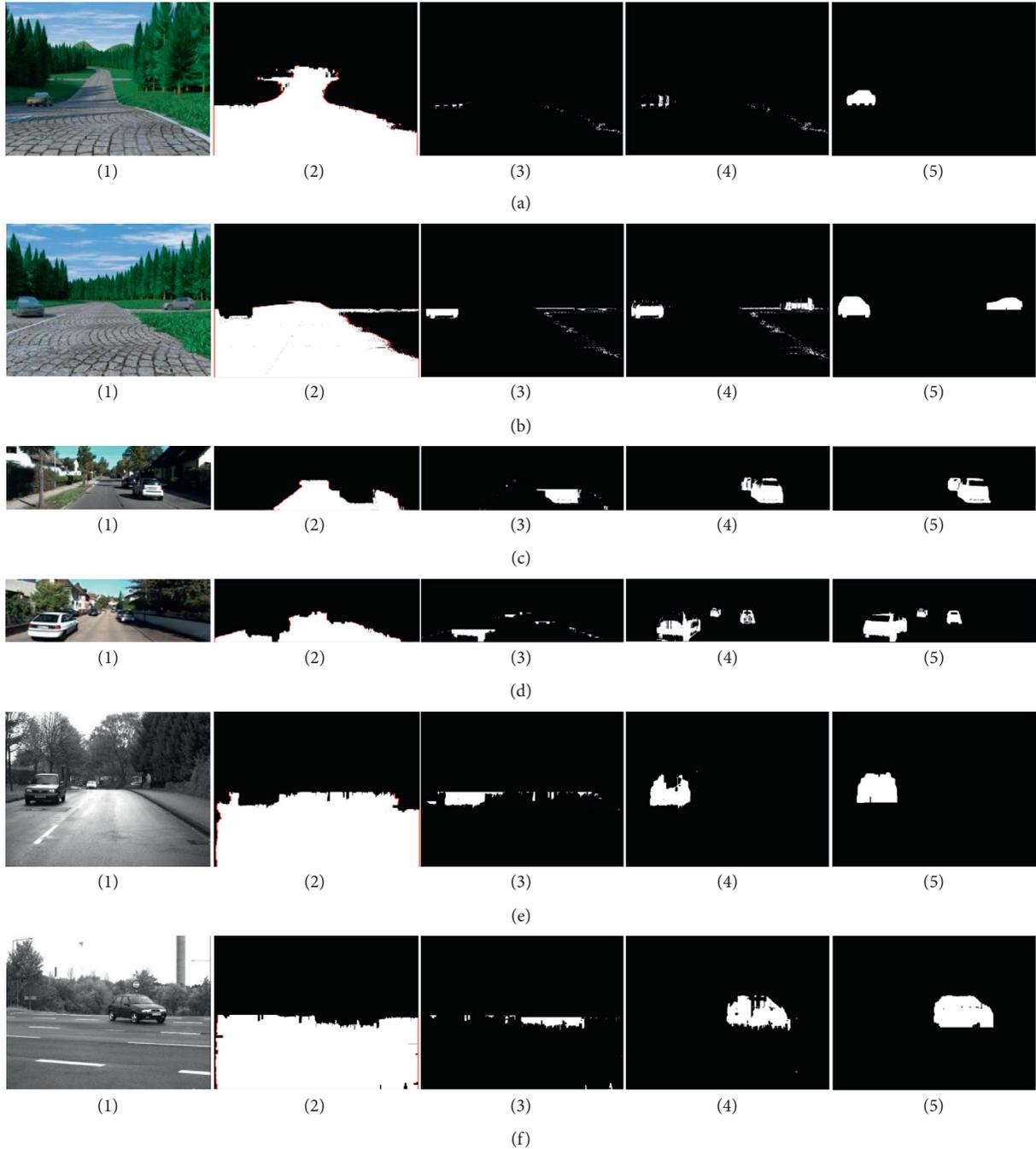


FIGURE 11: Typical vehicle detection result on the road. (1) Original image (left). (2) Region of interest. (3) Vehicle source point. (4) Initial detection result. (5) Detection result.

aspects. The experiments are performed on the image pairs containing valid vehicles in three standard datasets. The Enpeda dataset contains 477 valid vehicles, the KITTI dataset contains 173 valid vehicles, and the Daimler dataset contains 462 valid vehicles in a time with more vehicles. Some typical vehicle detection results on the road are shown in Figure 11.

Among them, Figures 11(a) and 11(b) are vehicle detection results of the Enpeda dataset, Figures 11(c) and 11(d) are vehicle detection results of the KITTI dataset, and

Figures 11(e) and 11(f) are vehicle detection results of the Daimler dataset. From the qualitative detection result in the above figures, it can be seen that the proposed algorithm can accurately detect whether it is a single-car or multi-car scene. Vehicles failed to be detected in Figures 11(d) and 11(e) because the vehicles in Figure 11(e) are not included in the road detection results and the vehicles are far away in Figure 11(d); the body and background are integrated.

Similarly, the results of the proposed vehicle detection algorithm are evaluated using Precision, Recall, and

TABLE 2: The vehicle detection result mixing matrix of three standard datasets.

			Algorithm Assessment			
			Enpeda	KITTI	Daimler	Recall
Ground Truth	Enpeda	477	418	0	0	87.63
	KITTI	173	0	163	0	94.21
	Daimler	462	0	0	402	87.01
	Precision		89.13	89.56	89.14	89.44

comprehensive evaluation index F . The accuracy rate is for the prediction result, which indicates how many of the results detected as vehicles are real vehicles calibrated in the dataset; the recall rate is for the original sample, which indicates how many real vehicles calibrated in the dataset are accurately detected. The dark gray area indicates the type of dataset, and the light gray area indicates the number of real and effective vehicles in the three datasets and the actual number of vehicles detected by the algorithm. From these data, the accuracy rate, recall rate, and comprehensive evaluation index can be calculated. The calculation formulas are shown in (20)–(22) [37]:

$$P = \frac{TP}{TP + FP}, \quad (20)$$

$$R = \frac{TP}{TP + FN}, \quad (21)$$

$$F = 2 \frac{P \cdot R}{P + R}. \quad (22)$$

Among them, TP is the true rate, that is, it is detected as a vehicle and it is indeed a vehicle; FP is a false positive rate, that is, it is detected as a vehicle, which is actually a non-vehicle object; TN is a true negative rate, that is, the vehicle is not detected but is actually a vehicle; and FN is a false negative rate, that is, the vehicle is not detected and is indeed a nonvehicle object. The detection results are shown in Table 2.

From the above experimental results, it can be known that the vehicle detection on the road algorithm proposed in this paper can accurately measure the area where the vehicle is located and has better robustness under complex background. The algorithm only detects traffic objects in the ROI region, reducing the range of possible vehicles, thereby reducing the calculation time, increasing the calculation speed, and meeting the real-time requirements. However, the algorithm still has missing detection and false detection. The main reasons are as follows: (1) there is a lot of noise and invalid areas in the disparity image. If such areas are large, incomplete correction will affect the detection results. (2) There are errors in road detection results, resulting in inaccurate ROI regions and missed detection of road vehicles. (3) There is not much difference in disparity and color similarity between the road near the vehicle and the vehicle, which affects the accurate division of the vehicle.

5.3. Pedestrian Detection Results on the Road. We select road scenes from typical datasets to verify the proposed

pedestrian detection algorithm on the road. Experiments have determined that the range of the pedestrian's aspect ratio threshold R_{lw} is set to 1.5–5, the perspective ratio R_p threshold is set to 0.12–0.45, and the area ratio R_a is less than 0.75. Some typical pedestrian detection results on the road are shown in Figures 12; Figures 12(a) and 12(c) are single pedestrian detection results, and Figures 12(b) and 12(d) are multiple pedestrian detection results. At the same time, the road scenes in Figures 12(b)–12(d) are more complicated, and there are more types of nonpedestrian obstacles.

From the above qualitative experimental results, it can be known that the pedestrian detection algorithm proposed in this paper can accurately measure the area where pedestrians are located and has better robustness under complex backgrounds.

5.4. Real-Time Analysis. The experiment in this paper uses a Windows 10 64 bit operating system, a CPU with Intel Core i5, 3.2 GHz, and a 32G memory PC as the experimental platform. The experimental environment is MATLAB 2017b.

For real-time analysis, this section uses a random sampling method to randomly extract 100 pairs of images from the three standard datasets for labeling. The method for automatic detection and classification of traffic objects proposed in this paper can be summarized into five stages: (1) acquisition and correction of disparity images, (2) considerate U-V view, (3) road surface detection, (4) vehicle detection on the road, and (5) pedestrian detection on the road. Figure 13 shows the average processing time for each stage of each frame in each dataset. The processing time for road detection, road vehicle detection, and pedestrian detection is calculated separately. It can be seen from Figure 13 that the total processing time of the KITTI dataset is the longest, which is about 99.8 ms (10pfs). It is mainly a road in the city and there are many objects in the scene. It takes a relatively long time in road detection, vehicle detection, and pedestrian detection on the road, and its image resolution is larger than that of the Enpeda and Daimler datasets. The least time consuming is the Enpeda dataset, with a time of about 59.4 ms (17pfs). This dataset is artificially synthesized, the target scene is relatively simple, and the pedestrian scene is lacking. The total processing time of the Daimler dataset is about 79.2 ms (13pfs). The scenes of roads, vehicles, and pedestrians on the road are rich, and the time is between the two datasets.

The real-time analysis results show that, in the actual application process, a certain number of interval frames can be selected for processing, and the real-time requirements

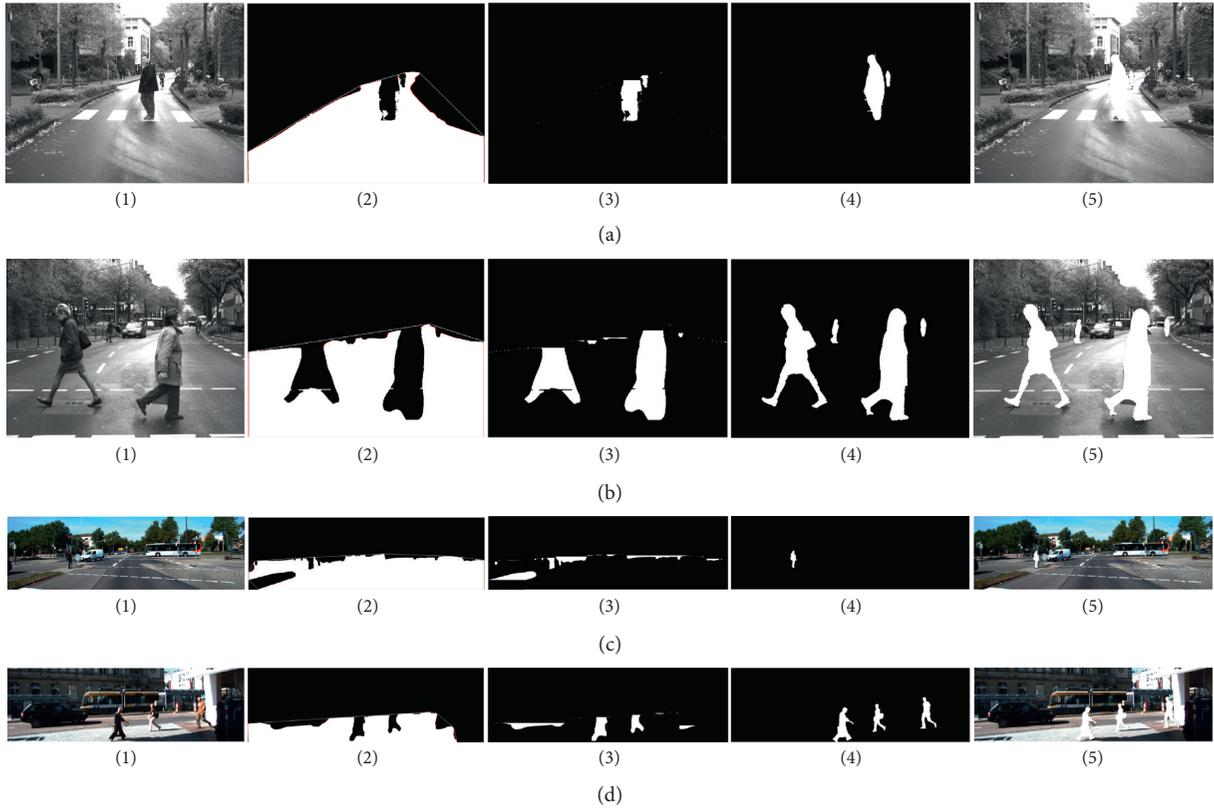


FIGURE 12: Typical pedestrian detection result on the road. (1) Original image (left). (2) Region of interest. (3) Pedestrian source point. (4) Pedestrian detection result. (5) Comparison.

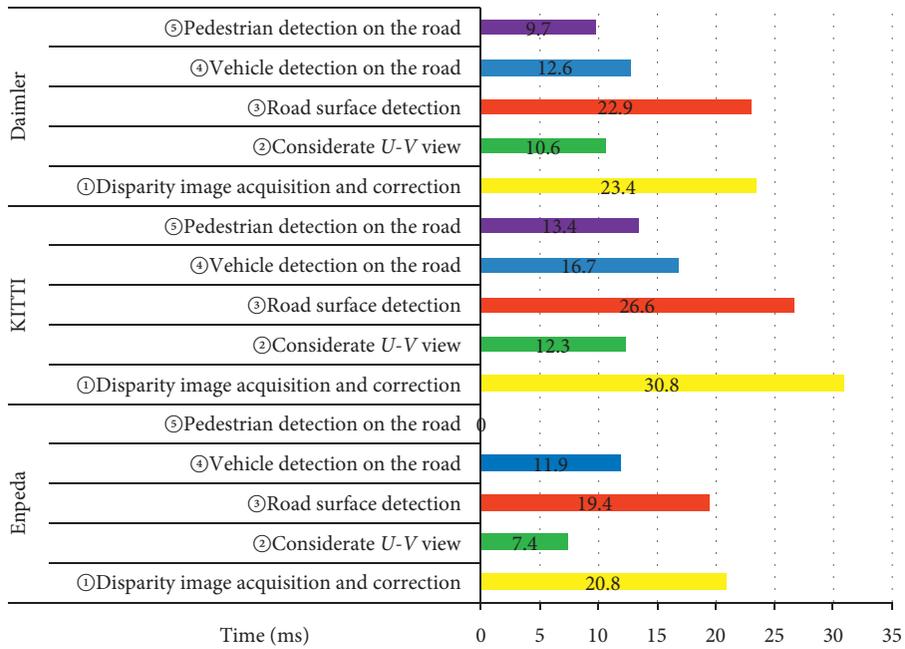


FIGURE 13: The real-time analysis result of three standard datasets.

can be met without affecting the final results of the experiment.

6. Conclusions

- (1) The classification detection of roads, vehicles, and pedestrians with multiple traffic objects under the same detection framework is proposed, which makes up for the ambiguity and unity of traditional detection methods. At the same time, it is the first time to classify detection of vehicles and pedestrians on the road.
- (2) The proposed road detection link is only based on a considerate U-V view method, using geometric characteristics without the need to build a complex road model, and the final V view can obtain accurate uphill and downhill road detection results without straight and curve fitting.
- (3) The proposed method for detecting vehicles and pedestrians on the road carries the results of road detection, uses the proposed peripheral envelope algorithm to determine the contour of the road periphery, defines the source points of vehicles and pedestrians, combines the characteristics of vehicles and pedestrians, proposes multifeature and multi-threshold fusion, and constructs disparity similarity and energy minimization algorithm to obtain the detection results of vehicles and pedestrians on the road.
- (4) The effectiveness of the proposed method is tested on three standard datasets of Enpeda, KITTI, and Daimler, including different traffic environments, different road alignments, different vehicles, pedestrian distribution, and different road scenes. The comprehensive evaluation index F values for road detection are 97.54%, 95.03%, and 93.76%; the overall comprehensive evaluation index F value for vehicle detection on the road is 89.44%; the pedestrian detection on the road also obtains better detection results.
- (5) The random sampling method is adopted to reflect the time efficiency of the entire dataset by the detection time of some frames of the dataset. The overall detection process is broken down into five stages to gradually determine the corresponding time, and the average time efficiency is 13 pfs, which can achieve real-time detection.

Data Availability

Previously reported datasets with Enpeda dataset, KITTI dataset, and Daimler dataset were used to support this study and are available at DOI: 10.1109/IVS.2011.5940480, DOI: 10.1109/CVPR.2012.6248074, and DOI: 10.1109/TPAMI.2008.260, respectively. These prior studies (and datasets) are cited at relevant places within the text as references [[34]–[36]]. The datasets also are available from the corresponding author and the first author upon request.

Conflicts of Interest

The authors declare no potential conflicts of interest.

Authors' Contributions

Yongchao Song and Jieru Yao contributed equally to this work. Yongchao Song and Jieru Yao conceived and designed the experiments; Yahong Jiang and Kai Du presented tools and carried out the data analysis; Yongchao Song and Jieru Yao wrote the paper; Yongfeng Ju and Kai Du guided and revised the paper; Yongchao Song and Yahong Jiang rewrote and improved the theoretical part; Jieru Yao and Yahong Jiang collected the materials and did a lot of format editing work.

Acknowledgments

The authors thank the Fundamental Research Funds for the Central Universities, CHD (Grant no. 300102320721) for their support. This research was funded by the General Program of National Natural Science Foundation of China (Grant no. 61603057). This research was also supported by the Fundamental Research Funds for the Central Universities (Grant no. 300102328403) and Natural Science Basic Research Plan in Shaanxi Province of China (Grant no. 2019JQ-073 and 2020JM-255).

References

- [1] Mh Xu, Kx Zhang, and Zl Jiang, "Algorithm design and implementation for a real-time lane departure pre-warning system," *Journal of Traffic and Transportation Engineering*, vol. 16, no. 3, pp. 149–158, 2013.
- [2] H. Fan, X. Mei, D. Prokhorov, and H. Ling, "Multi-level contextual RNNs with attention model for scene labeling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 11, pp. 3475–3485, 2018.
- [3] F. Yang, H. Fan, P. Chu, E. Blasch, and H. Ling, "Clustered object detection in aerial images," *IEEE*, vol. 34, pp. 8311–8320, 2019.
- [4] Y. Song, Y. Ju, K. Du, W. Liu, and J. Song, "Online road detection under a shadowy traffic image using a learning-based illumination-independent image," *Symmetry*, vol. 10, no. 12, p. 707, 2018.
- [5] T. Chen, R. Wang, B. Dai, D. Liu, and J. Song, "Likelihood-field-model-based dynamic vehicle detection and tracking for self-driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11, pp. 3142–3158, 2016.
- [6] F. Yang, M. Yi, Y. Cai et al., "Multitask assessment of roads and vehicles network (MARVN)," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 10, 2018.
- [7] Y. Lin, C. Yue-Lin, and L. Jing-Liang, "Pedestrian fast detection based on histograms of oriented gradient," *Computer Engineering*, vol. 36, no. 22, pp. 206–210, 2010.
- [8] R. Labayrade, D. Aubert, and J. P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through v-disparity representation," *IEEE*, vol. 56, pp. 646–651, 2002.
- [9] N. Soquet, D. Aubert, and N. Hautiere, "Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation," *IEEE*, vol. 45, pp. 160–165, 2007.

- [10] F. Oniga and S. Nedevschi, "Processing dense stereo data using elevation maps: road surface, traffic isle, and obstacle detection," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1172–1182, 2009.
- [11] C. Oh, B. Kim, and K. Sohn, "Automatic illumination invariant road detection with stereo vision," *IEEE*, vol. 67, pp. 889–893, 2012.
- [12] G. B. Vitor, D. A. Lima, A. C. Victorino, and J. V. Ferreira, "A 2d/3d vision based approach applied to road detection in urban environments," *IEEE*, vol. 76, pp. 952–957, 2013.
- [13] M. Wu, S. K. Lam, and T. Srikanthan, "Nonparametric technique based high-speed road surface detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 874–884, 2015.
- [14] B. Wang and V. Frémont, "Fast road detection from color images," *IEEE*, vol. 34, pp. 1209–1214, 2013.
- [15] G. Chun-Zhao, T. YAMABE, and M. Seiichi, "Drivable road boundary detection for intelligent vehicles based on stereo-vision with plane-induced homography," *Acta Automatica Sinica*, vol. 39, no. 4, pp. 371–380, 2013.
- [16] L. Wang, T. Wu, Z. Xiao, L. Xiao, D. Zhao, and J. Han, "Multi-cue road boundary detection using stereo vision," *IEEE*, vol. 1–6, 2016.
- [17] Y. Xie, S. Zeng, Y. Zhang, and L. Chen, "A cascaded framework for robust traversable region estimation using stereo vision," *IEEE*, vol. 43, pp. 3075–3080, 2017.
- [18] M. Cheng, Y. Zhang, Y. Su, J. M. Alvarez, and H. Kong, "Curb detection for road and sidewalk detection," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10330–10342, 2018.
- [19] Y. Su, Y. Zhang, T. Lu, J. Yang, and H. Kong, "Vanishing point constrained lane detection with a stereo camera," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2739–2744, 2017.
- [20] Y. Zhang, Y. Su, J. Yang, J. Ponce, and H. Kong, "When Dijkstra meets vanishing point: a stereo vision approach for road detection," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2176–2188, 2018.
- [21] B. Wang, S. A. R. Florez, and V. Frémont, "Multiple obstacle detection and tracking using stereo vision: application and analysis," *IEEE*, vol. 56, pp. 1074–1079, 2014.
- [22] N. Fakhfakh, D. Gruyer, and D. Aubert, "Weighted v-disparity approach for obstacles localization in highway environments," *IEEE*, vol. 78, pp. 1271–1278, 2013.
- [23] J. Kang and M. J. Chung, "Stereo-vision based free space and obstacle detection with structural and traversability analysis using probabilistic volume polar grid map," *IEEE*, vol. 98, pp. 245–251, 2011.
- [24] H. Yoo, J. Son, B. Ham, and K. Sohn, "Real-time rear obstacle detection using reliable disparity for driver assistance," *Expert Systems with Applications*, vol. 56, pp. 186–196, 2016.
- [25] A. Burlacu, S. Bostaca, I. Hector et al., "Obstacle detection in stereo sequences using multiple representations of the disparity map," *IEEE*, vol. 45, pp. 854–859, 2016.
- [26] X. Han, J. Lu, C. Zhao, S. You, and H. Li, "Semisupervised and weakly supervised road detection based on generative adversarial networks," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 551–555, 2018.
- [27] A. Dairi, F. Harrou, M. Senouci, and Y. Sun, "Unsupervised obstacle detection in driving environments using deep-learning-based stereo-vision," *Robotics and Autonomous Systems*, vol. 100, pp. 287–301, 2018.
- [28] A. D. Costea and S. Nedevschi, "Traffic scene segmentation based on boosting over multimodal low, intermediate and high order multi-range channel features," *IEEE*, vol. 12, pp. 74–81, 2017.
- [29] A. Daniel Costea, R. Varga, and S. Nedevschi, "Fast boosting based detection using scale invariant multimodal multi-resolution filtered features," *IEEE*, vol. 56, pp. 6674–6683, 2017.
- [30] Z. Hu and K. Uchimura, "UV-disparity: an efficient algorithm for stereo-vision based scene analysis," *IEEE*, vol. 67, pp. 48–54, 2005.
- [31] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," *IEEE*, vol. 2, 2005.
- [32] J. K. Suhr and H. G. Jung, "Dense stereo-based robust vertical road profile estimation using Hough transform and dynamic programming," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1528–1536, 2014.
- [33] H.-C. Choi, J.-M. Park, W.-S. Choi, and S.-Y. Oh, "Vision-based fusion of robust lane tracking and forward vehicle detection in a real driving environment," *International Journal of Automotive Technology*, vol. 13, no. 4, pp. 653–669, 2012.
- [34] T. Vaudrey, C. Rabe, R. Klette, and J. Milburn, "Differences between stereo and motion behaviour on synthetic and real-world stereo sequences," *IEEE*, vol. 56, pp. 1–6, 2008.
- [35] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," *IEEE*, vol. 34, pp. 3354–3361, 2012.
- [36] C. G. Keller, M. Enzweiler, and D. M. Gavrila, "A new benchmark for stereo-based pedestrian detection," in *Proceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV)*, pp. 691–696, Baden-Baden, Germany, June 2011.
- [37] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," *IEEE*, vol. 25, pp. 1693–1700, 2013.