IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 October 2023, accepted 28 October 2023, date of publication 1 November 2023, date of current version 6 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329056

RESEARCH ARTICLE

Robust Stereo Road Image Segmentation Using Threshold Selection Optimization Method Based on Persistent Homology

WENBIN ZHU[1], HONG GU[1], ZHENHONG FAN[1], AND XIAOCHUN ZHU[2]
[1]School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[2]School of Automation, Nanjing Institute of Technology, Nanjing 211167, China

Corresponding authors: Wenbin Zhu (wenbinzhu@njust.edu.cn) and Xiaochun Zhu (zhuxc@njit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62231016, and in part by the Jiangsu Funding Program for Excellent Postdoctoral Talent.

ABSTRACT This paper introduces a novel method for road target segmentation in the context of autonomous driving based on stereo disparity maps. The proposed method utilizes topological persistence threshold analysis to address the challenges of selecting appropriate thresholds. The approach involves converting stereo road images into uv-disparity maps, extracting road planes using v-disparity maps, and calculating occupancy grid maps using u-disparity maps. Persistence diagrams are then constructed by generating segmentation results under various threshold parameters. By establishing persistence boundaries in these diagrams, the most significant regions are identified, enabling the determination of robust segmentation thresholds. Experimental validation using KITTI stereo image datasets demonstrates the effectiveness of the proposed method, with low error rates and superior performance compared to other segmentation methods. The research holds potential for application in autonomous driving systems.

INDEX TERMS Disparity map, persistent homology, image segmentation, threshold selection optimization.

I. INTRODUCTION

Significant progress has been made in the field of autonomous driving in recent years, with stability-assisting functions such as lane line extraction [1], path finding [2], and multi-object recognition [3]. Stereo images obtain richer information of various traffic elements, such as disparity and depth in traffic scenes, having the advantages of a simple system structure and flexible operational capabilities. Therefore, the construction of stereo disparity maps using stereo cameras has gradually become a prominent vision-based method with great development potential [4], [5]. Autonomous driving requires automatic detection and recognition of targets in front of the vehicle, involving first extracting the road surface, followed by segmenting road obstacles, and finally localizing and identifying these obstacles. One of the key technologies is obstacle segmentation based on stereo camera vision. The objects to be segmented are mainly road surface, road traffic signs, vehicles, pedestrians, and other road target information in front of the vehicle [6], [7], [8], [9], [10]. These methods all require setting thresholds, but accurate and reasonable thresholds are often difficult to obtain due to various factors such as noise. Moreover, the target segmentation results are sensitive to threshold selection, resulting in poor robustness of the segmentation results in practical applications due to the limitations of threshold selection [11], [12], [13].

Numerous studies have reported road target segmentation based on stereo disparity maps. For instance, Chen et al. [14] utilized a depth slicing technique to segment the stereo disparity map and then employed a region growing method to accurately label the object boundaries, thereby enhancing the segmentation of obstacles on the road. Similarly, Kormann et al. [15] proposed a region growing technique for vehicle segmentation, using a planar segmented mean shift clustering method and modeling vehicles as rectangles.

Wang et al. [16] presented a robust obstacle segmentation method based on g-disparity and effective disparity map computation, employing sample strips to construct road models

The associate editor coordinating the review of this manuscript and approving it for publication was Long Xu.

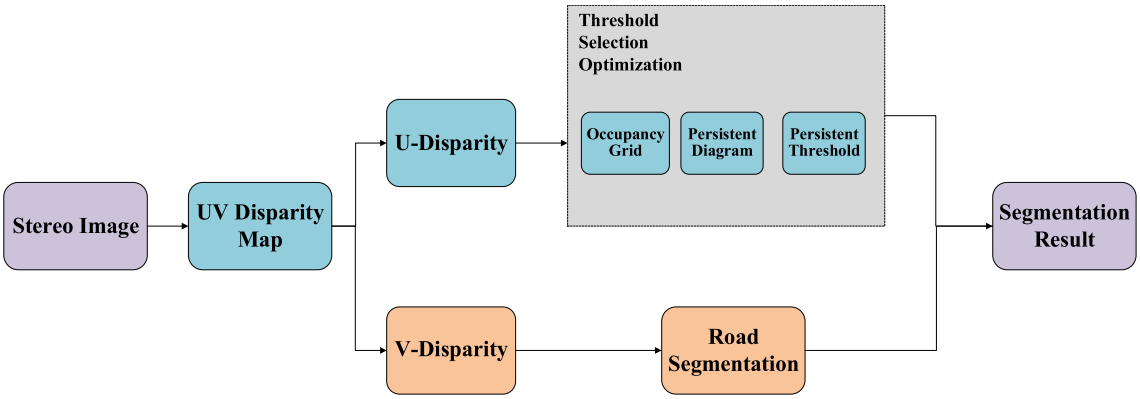

FIGURE 1. Overview of the proposed method.

and uv-disparity mapping to segment obstacles. Additionally, Lefebvre et al. [17] estimated 3D point clouds directly from dense disparity maps computed from stereo pairs and applied mean shift segmentation to achieve vehicle segmentation. Erbs et al. [18] computed dynamic stixels from stereo disparity maps and modeled real-world 3D road scenes using dynamic Stixels, achieving optimal segmentation through iterative dynamic planning. Furthermore, the group proposed an approach [19] for traffic scene understanding and driver assistance systems that combines Bayesian segmentation methods, enhancing the algorithm's robustness through the Stixel representation of images, and ensuring stable operation even under adverse weather conditions. Lee et al. [20] employed road features and disparity histograms for vehicle segmentation, extracted road features from v-disparity maps, utilized disparity histograms to localize obstacles, segmented them into multiple obstacles, and remerged them using four criterion parameters: obstacle size, distance, angle between obstacles, and disparity value difference, ultimately optimizing obstacle segmentation. Fritsch et al. [21] proposed an improved v-disparity map with fused confidence values. They estimated potential obstacle two-dimensional bounding boxes using the u-v disparity map, obtained obstacle object state vectors, and employed a weighted v-disparity map method. Cao et al. [22] proposed the V-intercept method for analyzing obstacles in the disparity space, which is fast but requires a robust threshold value. Robust segmentation is required for safe driving, and the target segmentation results are generally sensitive to the selection of thresholds. However, these road target segmentation methods often fail to provide robust threshold selection, as accurate and reasonable thresholds are often difficult to obtain due to various factors, such as lighting conditions and object texture.

In this paper, we propose a novel method based on persistent homology thresholding analysis to achieve uv-disparity road target segmentation (as shown in Fig. 1). By establishing persistence boundaries in these diagrams, the most significant regions are identified, enabling the determination of robust segmentation thresholds. Compared with other road target segmentation methods, this approach does not rely on a predefined threshold during segmentation. Instead, it obtains persistence boundary thresholds by analyzing all clusters in the persistence graph, allowing the method to achieve better performance. Additionally, the method utilizes the persistence graph to visualize the segmentation results corresponding to different thresholds. Through the analysis of these visualizations, more robust persistence thresholds can be obtained.

II. THRESHOLD SELECTION OPTIMIZATION BASED ROAD TARGET SEGMENTATION METHOD

The commonly used uv-disparity map [23] target segmentation method is to set a threshold on the occupied grid by a priori knowledge, but accurate and reasonable thresholds are often difficult to obtain due to various factors, such as lighting conditions and object textures, which lead to poor quality disparity pictures and bring errors. And the target segmentation results are generally sensitive to the threshold selection, i.e., a small threshold change can lead to a large difference in the segmentation region. In this regard, we propose a topological segmentation method based on persistent homology to solve the above problem in order to obtain more robust segmentation results, and the algorithm process is shown in Fig. 1. As shown in Fig. 1, our proposed method first transforms stereo images into UV disparity maps by SGBM method, then extract road from V-disparity, and threshold selection optimization based on persistent homology is performed on U-disparity.

A. CONSTRUCTION OF U-V DISPARITY MAP

Firstly, we utilize the Semi-Global Block Matching (SGBM) algorithm [24] to derive the disparity map from the stereo image. This map represents a 3D point cloud where each point's coordinates are denoted as (u, v, d). Here, (u,v) corresponds to the (x,y) coordinates in the 2D image, while d indicates depth information, allowing us to associate each point in the depth image with a point in the real 3D world. Assuming the camera's declination, pitch, and rotation angles

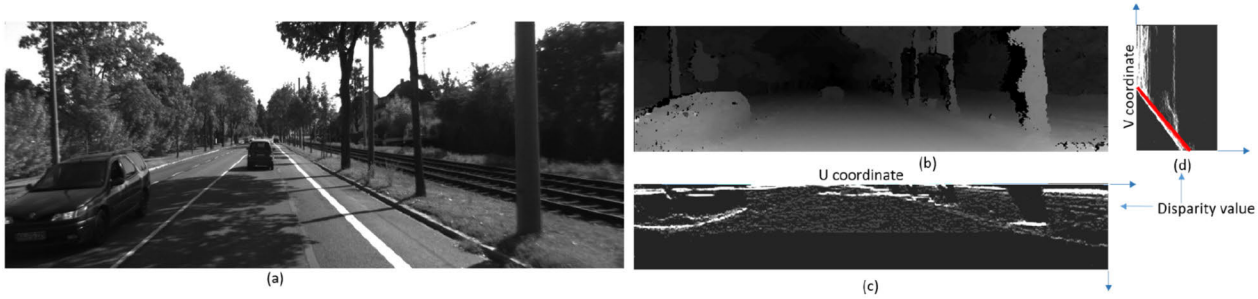

FIGURE 2. Disparity maps of the sample images of KITTI dataset (a) left grayscale image; (b) uv-disparity map of Fig. (a); (c) u-disparity map; (d) v-disparity map.

are calibrated to 0, we establish transformation equations between world coordinate points and their corresponding points in the depth image.

By projecting all points onto the ud-plane and vd-plane, two sub-images can be generated: the u-disparity and v-disparity maps. Fig. 2 exhibits an example disparity map obtained using the aforementioned method with the KITTI stereo dataset [25], [26]. Fig. 2(a) displays the left grayscale image of the stereo image, Fig. 2(b) represents the corresponding disparity map, and Fig. 2(c) and Fig. 2(d) show the u-disparity map and v-disparity map, respectively.

$$\begin{aligned} u &= u_0 + a_u \cdot \frac{z_W - x_0 - b_s/2}{y_W - y_0} \\ v &= v_0 + a_v \cdot \frac{z_W - z_0}{y_W - y_0} \\ d &= a_u \cdot \frac{b_s}{y_W - y_0} \end{aligned} \tag{1}$$

where (x_0, y_0, z_0) is the center coordinate of the stereo camera, a_u and a_v is the intrinsic focal length parameter, and b_s is the baseline distance between stereo cameras.

B. ROAD SEGMENTATION FROM V-DISPARITY

Based on the disparity map obtained in Section II-A, this study extracts the road using the v-disparity map. In real road scenarios, the road surface area can be approximated as a continuous horizontal surface, causing the disparity in the v-disparity map to decrease from bottom to top along the V-direction. Consequently, under ideal circumstances, the disparity statistics for the pavement area in the v-disparity diagram form a continuous diagonal line segment, encompassing the longitudinal extent of the pavement area in the original image. While the shape of obstacles may vary, along with the distance of each point on them from the acquisition device, the disparity value associated with a particular determined obstacle tends to remain within a specific range and exhibit continuity relative to other obstacles. Therefore, in an ideal state, the disparity statistics point corresponding to the obstacle in the v-disparity map approximately forms a continuous vertical line segment. The intersection of this vertical line segment and the diagonal line segment represents the coordinate of the vehicle's contact point with the road surface in the original vehicle image.

Accordingly, the pixel value corresponding to each row of the roads in the v-disparity map corresponds to the minimum pixel value within that specific row. As a result, the projection of the road on the v-disparity map appears as a line, which can be approximated by fitting a straight line within the v-disparity map.

To obtain a robust estimate of the road plane, a straight line is fitted to the v-disparity map by selecting the global minimum of the matching cost function, which is defined as $g_{ground}(d)$. The noise and the error in generating the disparity map may cause some error in fitting the road plane, so we set a threshold value h_l and h_h with the assumption that obstacles are within the thresholds to reduce the fitting error:

$$g_{ground}(u,d)+\left(\frac{\alpha_u h_l}{\alpha_v b}\right) d \le g_0(u,d) \le g_{ground}(u,d)+\left(\frac{\alpha_u h_h}{\alpha_v b}\right) d \tag{2}$$

where α_u and α_v are the camera intrinsic focal length parameters and b is the baseline distance of the stereo camera system. This value can be set according to the training dataset. For example, in the KITTI dataset, we set $h_l = 200$ and $h_h = 1700$, thus the ground plane 0.2m above the fitting line is preserved to avoid fitting errors and plane 1.7m above the fitting line is cut to avoid the influence of obstacles above the car.

C. OCCUPANCY GRID FROM U-DISPARITY

Given a point s in a u-disparity map with coordinates of (s_u, s_d), we define two binary random variables V_s and C_s denote the visibility of the point ($V_s = 1$ indicates the point is visible and $V_s = 0$ indicates the point is invisible) and the obstacle confidence ($C_s = 1$ indicates the presence of an obstacle at s), respectively. Then we can obtain the probability of occupation O_s at point s.

$$P(O_s) = \sum P(V_s = v, C_s = c) \bullet P(O_s | V_s = v, C_s = c) \tag{3}$$

We assume that V_s and C_s are independent of each other, which means $P(V_s = v, C_s = c) = P(V_s = v) P(C_s = c)$.

In order to obtain expressions for $P(V_s = v)$ and $P(C_s = c)$, we first need to calculate.

$$\begin{aligned} N_P(s) &= |A_P(s)| \\ &= \left|\left\{(u, v)\,|u = s_u, v \in \left[g_{lower}(s_d), g_{upper}(s_d)\right]\right\}\right| \quad (4) \\ N_O(s) &= |A_O(s)| = |\{(u, v)\,|I_D(u, v) = s_d\} \cap A_P(s)| \quad (5) \\ N_V(s) &= |A_V(s)| = |\{(u, v)\,|I_D(u, v) \leq s_d\} \cap A_P(s)| \quad (6) \end{aligned}$$

where I_D is the disparity map, (u, v) is the coordinate in the disparity image. $N_P(s)$ is the total number of measured points on point s in the image, $N_O(s)$ is the total number of obstacle points, and $N_V(s)$ is the total number of visible points of the points.

Then the visibility probability of point s in the u-disparity map is defined as

$$P(V_s) = \frac{N_V(s)}{N_P(s)} \quad (7)$$

The confidence probability of the observation is:

$$P(C_s) = 1 - e^{-\lambda \frac{N_O(s)}{N_V(s)}} \quad (8)$$

where λ is a constant. Then define $P\,(O_s|V_s = v, C_s = c)$ as:

$$\begin{aligned} &P\,(O_s|V_s = 0, C_s = c) = 0 \\ &P\,(O_s|V_s = 1, C_s = 1) = 1 - P_{FP} \\ &P\,(O_s|V_s = 1, C_s = 0) = P_{FN} \end{aligned} \quad (9)$$

where P_{FP} and P_{FN} represent the false positive and false negative of occupancy respectively.

D. THRESHOLD SELECTION OPTIMIZATION BASED ON PERSISTENT HOMOLOGY

The u-disparity map contains the probabilities of the target regions or edges in the disparity map, from which we want to obtain regions with high probability values or regions surrounded by edges with high probability values.

When studying the topological properties of a space, the homogeneous type of this space is generally constructed because the topological properties of two spaces with the same embryo are the same. However, it is difficult to construct the homology of an image space, so we use the idea of continuous homology to approximate this space by constructing a simplicial complex with a series of continuous parameters [27]. In this paper, we use the Vietoris-Rips complex to construct the simplicial complex.

We compute a set of simplicial complexes at different scales in a continuous homogeneous manner to find features that are closer to the essential properties of the data by computing features that are consistently stable over a range of scales. This process is called filtration S_ε. When $\varepsilon_1 < \varepsilon_2$, S_ε satisfies $S_{\varepsilon 1} \subseteq S_{\varepsilon 2}$.

The values corresponding to the appearance (birth) and disappearance (death) of topological features in the filtering process are called the appearance time and disappearance time of the kth n-dimensional feature, respectively. The set of points consisting of emergence and disappearance times (b_n^k, d_n^k), is called a persistence graph. The duration graph has stability, i.e., if there is a change in the point cloud set, there will be a corresponding change in the duration graph. A diagonal line can be drawn along the bottom left to the top right of the continuum graph. Since the vanishing time of the connected components in the screening process is always greater than the appearance time, the point on the continuum graph must be at the top left of this diagonal line. The value of each point on the continuum graph is defined as the continuum interval $\lambda = d_n^k - b_n^k$.

When distinguishing a topological space, we want to find its topologically invariant features, i.e., topological invariants. The homology group is an important topological invariant characteristic of a topological space, and the homology of a topological space can be measured by the Betti number [28]. Betti number is the rank of the homology group, and the nth dimensional Betti number is the number of n-dimensional holes on the topological space [29], [30]. In this paper, we focus on the 0-dimensional Betti number, i.e., the number of connected components.

As shown in the filtering process in Fig. 3(a), the emergence (birth) and disappearance (death) of the connected components within the filtering process for the upper level set occupying the probability fs of the grid can be visualized. When $\tau = 0.02$, the cyan area of the right vehicle appears. When $\tau = 0.03$, another red area appears, but it merges with the cyan area at the time $\tau = 0.10$, i.e., it disappears, and the red area lasts for an interval of $\lambda_{per} = 0.07$. When $\tau = 0.53$, the cyan area continues to exist and expands in size. When $\tau = 0.61$, the cyan area merges with other areas. If we choose to keep only the region with the duration interval $\lambda_{per} > 0.2$, then the cyan region will be kept and the red region will be removed, and we can see from the image that the red region is not the vehicle target we want and can be considered as noise in the probability map.

In the duration map in Fig. 3(b), the appearance and disappearance information of all regions is shown. Each point in the persistence diagram represents the period from appearance to disappearance of the connected components occupying the grid in the filtering process as the threshold value changes, and its horizontal coordinate is the threshold value corresponding to the moment of appearance and the vertical coordinate is the threshold value corresponding to the moment of disappearance, so the persistence diagram can be understood as a function relationship about the threshold value. Since the disappearance is generally caused by the merging of small connected components with each other, the graph contains information about the merging of these regions. In order to obtain a robust segmentation of the target, a persistence boundary λ_{per} is set to select the most salient regions in the clustering process of the data. This selection process is represented in the persistence diagram as the feature selected above a specific straight line. As the line increases, the targets close to each other in the image

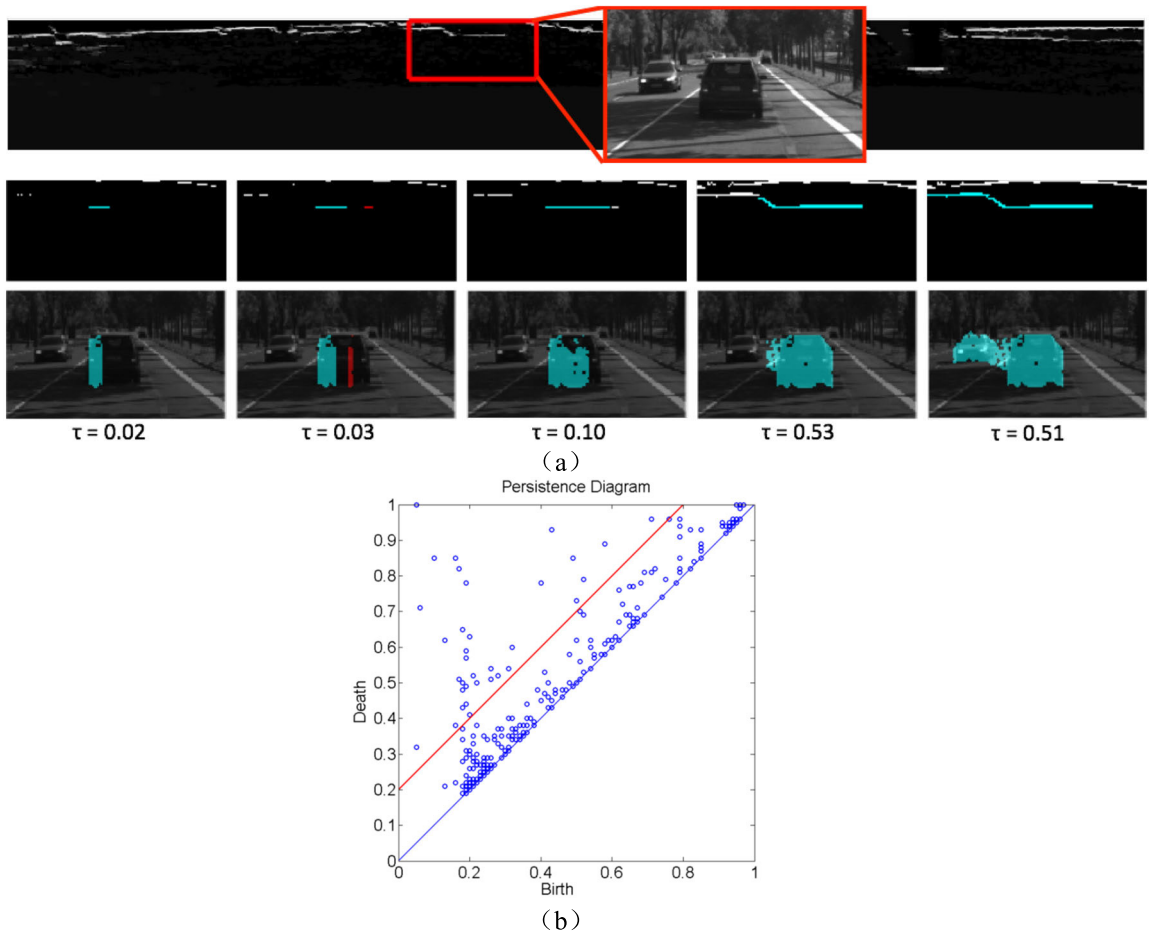

FIGURE 3. Road target segmentation screening and duration map (a) u-disparity map corresponding to the screening process with different thresholds; (b) continuity map.

space are first merged, so that a smaller number of regions can be segmented but occupy a larger area. In Fig. 3(b), only the region larger than the persistence boundary $\lambda_{per} = 0.2$, i.e., the red line on the way, is retained. In order to mark the corresponding segmentation results in the u-disparity map, we need to determine the support domain corresponding to the persistence diagram. Here we choose the region and the threshold corresponding to the region when its point set is maximum before it disappears. Since in this case the different regions selected may have overlapping parts, the overlapping region in the u-disparity map is labeled to the region with the earliest vanishing time.

III. EXPERIMENTAL RESULTS AND ANALYSIS

To illustrate and evaluate the performance of the road segmentation method proposed in this paper, experiments were conducted using images from two stereo image datasets, KITTI 2012 and KITTI 2015, by MATLAB R2022a on Intel(R) Core(TM) i7-11800H 2.3 GHz CPU and NVIDIA Geforce RTX 3060 GPU Windows system. Section III. A describes the experimental setup and dataset; Section III-B and III-C conduct comparison and validation experiments on several stereo image datasets; the algorithm proposed in this paper is compared and discussed with traditional methods in Section III-D, and the role of persistence bounds is discussed.

A. EXPERIMENT SETUP

In this paper, three publicly available online stereo image datasets are used for experimental evaluation and comparison.

(1) KITTI 2012: a real-world dataset containing different traffic scenes was collected, and images were obtained from a moving platform recording on a Volkswagen station wagon driving around Karlsruhe, Germany. The stereo image dataset consists of 194 training image pairs and 195 test image pairs, saved in lossless png format.

(2) KITTI 2015: KITTI 2015 was collected in the same way as KITTI 2012 data, consisting of 200 training scenes and 200 test scenes (4 color images per scene, saved in lossless png format).

(3) Daimler dataset: a real-world dataset containing 21, 790 image pairs (640*480 pixels) with 56492 manual labels, and images of various objects were captured.

(4) Enpeda dataset: Enpeda is a synthetic stereo image dataset containing 496 image pairs with 640*480 pixels.

PSMNet [31] and Displets v2 [32] are two of the leading methods in the KITTI stereo leaderboard compared in this paper.

The Displets v2 method performs specification at larger distances and determines the target parallax position (displets) by constructing an ensemble topological model of the object surface. Displets considers that the shape of a particular class of objects is not arbitrary, it has a typical regular structure. While most binocular vision stereo matching algorithms concentrate on textual features and smoothing assumptions, ignoring the importance of semantic information. The Displets v2 method takes into account the weak textuality, reflectivity and translucency of the target class, and improves the matching effect by increasing the distance between possible targets using the knowledge of target recognition. It obtains planar parameters by matching local planes and parallax maps, and establishes an energy cost function to estimate pixel point parallax values. Combining the above ideas, experiments are conducted for vehicle targets, and the method is ranked top in the KITTI dataset.

PSMNet argues that the current architecture relies on patch-based Siamese networks, but it is still difficult to find accurate counterparts in inherently ill-defined regions, such as occluded regions, repetitive patterns, texture-free regions, and reflective surfaces. Simply applying the intensity consistency constraint between different viewpoints is not sufficient for accurate correspondence estimation in ill-defined regions, and is useless for texture-free regions. Therefore, PSMNet incorporates region support from global contextual information into stereo matching. PSMNet also extends pixel-level features to accept region-level features at different scales using spatial pyramid pooling (SPP) and null convolution to enlarge the perceptual field. In addition, a stacked hourglass 3D CNN and intermediate supervision are designed to regulate the cost volume. The stacked hourglass 3D CNN reprocesses the cost volume in a top-down/bottom-up manner to further improve the utilization of global contextual information.

B. COMPARISON EXPERIMENT RESULTS

KITTI 2012 ranks the methods according to the number of non-obscured error pixels at the specified disparity/endpoint error threshold. The evaluation metrics are defined as follows:

1) Out-Noc: the percentage of erroneous pixels in the non-occluded region.

2) Out-All: percentage of the total number of erroneous pixels.

3) Avg-Noc: the average disparity/endpoint error in the non-occluded region.

4) Avg-All: average difference/total endpoint error.

The error thresholds were set to 3, 4 and 5 px (pixel), and the segmentation of KITTI was performed using the method in this paper and both Displets v2 and PSMNet, and the results are shown in Tables 1 to 3.

As seen from Tables 1 to 3, the algorithm in this paper performs well in both Out-Noc and Out-All regions, with 1.17% and 1.54% at 3px, 0.92 and 1.21 at 4px, and 0.77 and 1.01 at 5px, respectively, which are significantly better than the Displets v2 and PSMNet methods. The indicators are close to PSMNet, but they are still the best performers among the three methods.

TABLE 1. Test results of KITTI dataset (error threshold of 3px).

	Out-Noc/%	Out-All/%	Avg-Noc(px)	Avg-All(px)
Displets v2	2.37	3.09	0.7	0.8
PSMNet	1.49	1.89	0.5	0.6
Proposed	1.17	1.54	0.5	0.5

TABLE 2. Test results of KITTI dataset (error threshold of 4px).

	Out-Noc/%	Out-All/%	Avg-Noc(px)	Avg-All(px)
Displets v2	1.97	2.52	0.7	0.8
PSMNet	1.12	1.42	0.5	0.6
Proposed	0.92	1.21	0.5	0.5

TABLE 3. Test results of KITTI dataset (error threshold of 5px).

	Out-Noc/%	Out-All/%	Avg-Noc(px)	Avg-All(px)
Displets v2	1.72	2.17	0.7	0.8
PSMNet	0.90	1.15	0.5	0.6
Proposed	0.77	1.01	0.5	0.5

The KITTI 2015 dataset adds the use of semi-automated methods to obtain the true value of dynamic images compared to KITTI 2012, so the evaluation algorithm is implemented by calculating the ratio of false detected pixels to true pixels in the test images.

The evaluation metrics for this set of experiments are as follows:

1) D1: the percentage of stereo disparity anomalies in the first frame.

2) bg: the percentage of anomalous values averaged over the background region only.

3) fg: percentage of outliers averaged over the foreground region only.

4) all: the percentage of outliers averaged over all ground truth pixels.

The results of the segmentation of KITTI using the algorithm of this paper and recent methods such as Displets v2 and PSMNet are shown in Table 4.

The performance of the proposed model is quantitatively evaluated using the percentage of erroneous pixels in the

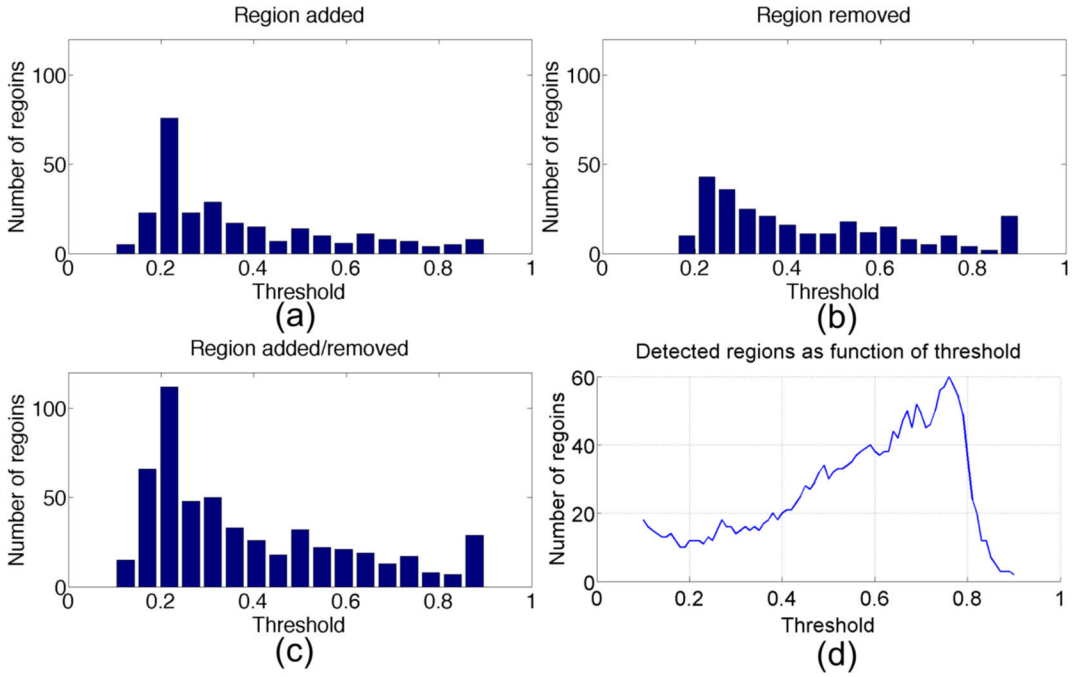

FIGURE 4. Variation of segmented regions with the conventional method (a) the number of regions added; (b) the number of regions removed; (c) the total number of regions changed; (d) the variation of the total number of regions with [the threshold width in (a)-(c) is 0.05].

TABLE 4. Test results of KITTI dataset.

	All pixels			Non-occluded pixels		
	D1-bg(%)	D1-fg(%)	D1-all(%)	D1-bg(%)	D1-fg(%)	D1-all(%)
Displets v2	3.00	5.56	3.43	2.73	4.95	3.09
PSMNet	1.86	4.62	2.32	1.71	4.31	2.14
Proposed	1.51	2.88	1.74	1.32	2.87	1.58

TABLE 5. Quantitave analysis result of proposed method on various datasets.

	Precision	Recall	FA
KITTI	0.91	0.96	0.25
Daimeler	0.93	0.96	0.19
Enpeda	0.95	0.97	0.12

background (D1-bg), foreground (D1-fg), and all pixels (D1-all), and as can be seen in Table 4, the proposed model performs significantly better than other methods for the texture-free and foreground regions on the KITTI dataset.

C. VALIDATION EXPERIMENT RESULTS

In this section, experiments are carried out on Dameler and Enpeda stereo image dataset to validate proposed method. The following metrics from literature are used to quantify the segmentation performance:

1) Precision: correct matches / total groundtruth objects.
2) Recall: correct matches / total objects.
3) FA: false alarms.

The results are shown in table 5. Our evaluations on KITTI, Daimeler and Enpeda datasets demonstrate that the proposed method give pretty accurate and precise segmentation result. The robustness of the proposed method is discussed in the next section.

D. EXPERIMENT DISCUSSION AND ANALYSIS

Although the threshold method is a simple method to occupy obstacle segmentation in probability maps, it is very sensitive to the choice of parameter values. In contrast, the method based on persistence analysis can obtain robust thresholds and keep track of all generated segmentation results in a topological manner. The segmentation results from the validation experiments, KITTI dataset, show that the method proposed in this paper can obtain more stable results.

To further analyze the robustness of the algorithm in this paper, the relationship between the traditional method and the method and segmentation region in this paper are investigated in Fig. 4 and Fig. 5, respectively. The histogram of the increase or removal of the connected components when the threshold value of the two methods changes is visualized and quantified.

Fig. 4(a) and (b) show the histograms of the corresponding changes in the emergence and disappearance values in the persistence analysis when the threshold value τ is changed using the traditional method. Fig. 4(c) shows the change in the total number of regions for different threshold parameters. Fig. 4(d) presents the relationship between the segmented regions as a function of threshold value τ. As seen in Fig. 4(c), when τ changes from 0.45 to 0.5, about 30 regions are added or removed, and some of them are removed by the post-processing process. When τ changes from 0.5 to 0.55, about

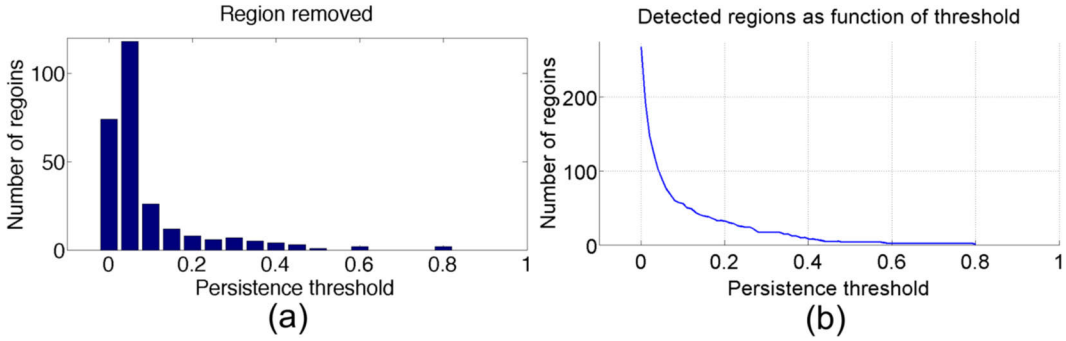

FIGURE 5. Variation of segmented regions in this paper with (a) the number of regions removed; (b) the variation of the total number of regions with (where the threshold width is 0.05).

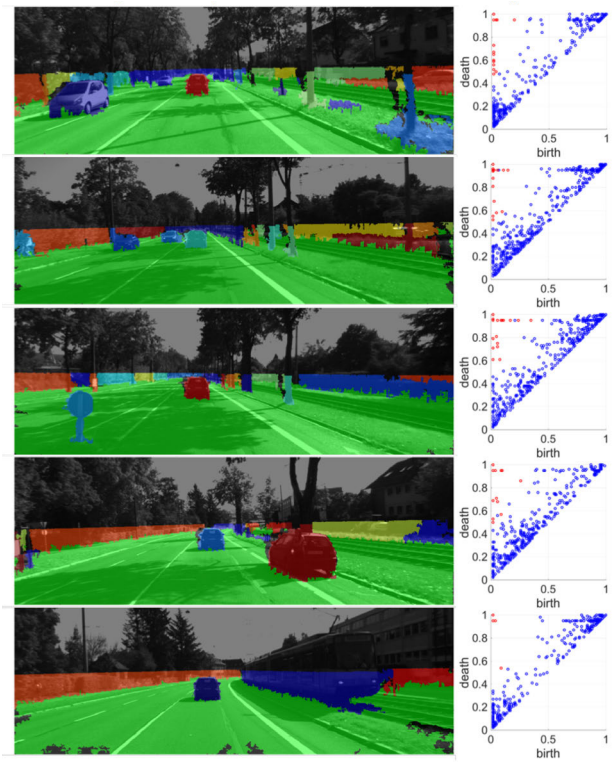

FIGURE 6. Road segmentation results of the proposed method and the corresponding persistence diagrams.

20 regions are added or removed. We note that the curve of the total number of regions segmented versus the threshold value τ is not smooth and thus the robustness of the threshold segmentation is not good.

In this paper, the histogram of region removal λ_{per} is obtained by merging the regions, i.e., with the horizontal coordinate λ_{per} and the number of regions as the vertical coordinate, and the results are shown in Fig. 5(a).

As seen in Fig. 5(a), the λ_{per} changes from 0.15 to 0.2 or from 0.2 to 0.25 results in less than 10 regions. The total number of regions as a function of λ_{per} is shown in Fig. 5(b), which is directly related to the histogram in Fig. 5(a), and the curve is smooth, indicating good robustness of the threshold segmentation.

Persistence boundaries λ_{per} are used to select the most prominent regions in the clustering process of the data. This selection process can be shown in the persistence plot as the features selected above a specific straight line. As this straight line moves upward with the increase of λ_{per}, obstacles close to each other in the image space will be merged, so that a smaller number of regions but occupying a larger area can be selected.

Figure 6 shows several results of our image segmentation using our method. It can be seen that our method is able to segment the ground and obstacles correctly. For cars that are not too far from the cameras, they are consistently detected as a single region. In the second row, the method is also able to detect a person who is riding a bicycle. Also, most of the trees and bushes on both sides of the road were properly detected and segmented. While the bushes on the left side were always detected, they were sometimes segmented into single regions or merged into a single region. Overall, the method successfully recognizes and segments the ground and obstacles, and is particularly good at detecting vehicles in close proximity, recognizing them as a single area. In addition, the method was able to detect bicyclists, as well as trees and bushes on both sides of the road.

IV. CONCLUSION

In conclusion, this paper proposes a robust method for extracting road planes and segmenting obstacles from stereo disparity maps. The approach utilizes the uv-disparity method to transform the stereo traffic scene into a uv-disparity map, extracts road planes using the v-disparity map, and calculates the occupancy grid map based on the u-disparity map. Different from other segmentation methods, this method obtains persistence boundary thresholds by tracking all clusters in the persistence diagram. The persistence diagram can visualize segmentation results at various thresholds, enabling the identification and selection of more precise persistence parameters. Overall, the segmentation results of the method are satisfactory in most cases. Experiment results on KITTI, Daimeler and Enpeda datasets validate the proposed method

on different stereo image datasets. Experiment results also demonstrate that the proposed method achieves a lower error rate compared to methods such as PSMNet and Displets v2 according to the KITTI System Benchmark.

REFERENCES

[1] W. Wang, F. Berholm, K. Hu, L. Zhao, S. Feng, A. Tu, and E. Fan, "Lane line extraction in raining weather images by ridge edge detection with improved MSR and Hessian matrix," *Inf. Technol. Control*, vol. 50, no. 4, pp. 722–735, Dec. 2021, doi: 10.5755/j01.itc.50.4.29094.

[2] U. Malūkas, R. Maskeliūnas, R. Damaševičius, and M. Woźniak, "Real time path finding for assisted living using deep learning," *J. Universal Comput. Sci.*, vol. 20, no. 4, pp. 475–487, Apr. 2018, doi: 10.3217/jucs-024-04-0475.

[3] B. Shi, X. Li, T. Nie, K. Zhang, and W. Wang, "Multi-object recognition method based on improved YOLOv2 model," *Inf. Technol. Control*, vol. 50, no. 1, pp. 13–27, Mar. 2021, doi: 10.5755/j01.itc.50.1.25094.

[4] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1341–1360, Mar. 2021, doi: 10.1109/TITS.2020.2972974.

[5] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah, "A survey of deep learning applications to autonomous vehicle control," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 712–733, Feb. 2021.

[6] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, Y. Li, and D. Cao, "Deep learning for image and point cloud fusion in autonomous driving: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 722–739, Feb. 2022, doi: 10.1109/TITS.2020.3023541.

[7] G. Li, Y. Chen, D. Cao, X. Qu, B. Cheng, and K. Li, "Extraction of descriptive driving patterns from driving data using unsupervised algorithms," *Mech. Syst. Signal Process.*, vol. 156, Jul. 2021, Art. no. 107589, doi: 10.1016/j.ymssp.2020.107589.

[8] Z. Qiu and Y. Li, "Real-time binocular foreground depth estimation algorithm based on sparse convolution," *J. Comput. Appl.*, vol. 41, no. 12, pp. 3680–3685, Mar. 2021.

[9] M. Hallek, H. Boukamcha, A. Mtibaa, and M. Atri, "Dynamic programming with adaptive and self-adjusting penalty for real-time accurate stereo matching," *J. Real-Time Image Process.*, vol. 19, no. 2, pp. 233–245, Apr. 2022, doi: 10.1007/s11554-021-01180-1.

[10] W. Ma and S. Zhu, "A multifeature-assisted road and vehicle detection method based on monocular depth estimation and refined U-V disparity mapping," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16763–16772, Sep. 2022, doi: 10.1109/TITS.2022.3195297.

[11] S. Chattopadhyay, Q. Ge, W. Wei, and E. Lobaton, "Robust multi-target tracking in outdoor traffic scenarios via persistence topology based robust motion segmentation," in *Proc. IEEE GlobalSIP*, Orlando, FL, USA, Dec. 2015, pp. 805–809.

[12] C. Wei, Q. Ge, S. Chattopadhyay, and E. Lobaton, "Robust obstacle segmentation based on topological persistence in outdoor traffic scenes," in *Proc. IEEE Symp. Comput. Intell. Vehicles Transp. Syst. (CIVTS)*, Oralando, FL, USA, Dec. 2014, pp. 92–99.

[13] Q. Ge, "Robust image segmentation: Applications to autonomous car and foraminifera morphology identification," Ph.D. thesis, Dept. Elect. Eng., NC State Univ., Raleigh, NC, USA, 2019.

[14] L.-C. Chen, X.-L. Nguyen, and C.-W. Liang, "Object segmentation method using depth slicing and region growing algorithms," in *Proc. Int. Conf. 3D Syst. Appl.*, Tokyo, Japan, 2010, pp. 87–90.

[15] B. Kormann, A. Neve, G. Klinker, and W. Stechele, "Stereo vision based vehicle detection?" in *Proc. VISAPP*, Angers, France, 2010, pp. 431–438.

[16] Y. Wang, Y. Gao, A. Achim, and N. Dahnoun, "Robust obstacle detection based on a novel disparity calculation method and G-disparity," *Comput. Vis. Image Understand.*, vol. 123, pp. 23–40, Jun. 2014, doi: 10.1016/j.cviu.2014.02.014.

[17] S. Lefebvre and S. Ambellouis, "Vehicle detection and tracking using mean shift segmentation on semi-dense disparity maps," in *Proc. IEEE Intell. Vehicles Symp.*, Madrid, Spain, Jun. 2012, pp. 855–860, doi: 10.1109/IVS.2012.6232280.

[18] F. Erbs, A. Barth, and U. Franke, "Moving vehicle detection by optimal segmentation of the dynamic stixel world," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 951–956, doi: 10.1109/IVS.2011.5940532.

[19] F. Erbs, B. Schwarz, and U. Franke, "From stixels to objects—A conditional random field based approach," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Gold Coast, QLD, Australia, Jun. 2013, pp. 586–591, doi: 10.1109/IVS.2013.6629530.

[20] M. Lee, "Integrated position and motion tracking method for online multi-vehicle tracking-by-detection," *Opt. Eng.*, vol. 50, no. 7, Jul. 2011, Art. no. 077203, doi: 10.1117/1.3595429.

[21] J. Fritsch, T. Kühnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, The Hague, The Netherlands, Oct. 2013, pp. 1693–1700, doi: 10.1109/ITSC.2013.6728473.

[22] T. Cao, Z.-Y. Xiang, and J.-L. Liu, "Perception in disparity: An efficient navigation framework for autonomous vehicles with stereo cameras," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2935–2948, Oct. 2015, doi: 10.1109/TITS.2015.2430896.

[23] Y. Gao, X. Ai, J. Rarity, and N. Dahnoun, "Obstacle detection with 3D camera using U-V-disparity," in *Proc. Int. Workshop Syst., Signal Process. Their Appl. (WOSSPA)*, Tipaza, Algeria, May 2011, pp. 239–242, doi: 10.1109/wosspa.2011.5931462.

[24] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008, doi: 10.1109/tpami.2007.1166.

[25] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. CVPR*, Providence, RI, USA, 2012, pp. 3354–3361, doi: 10.1109/CVPR.2012.6248074.

[26] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, Sep. 2013, doi: 10.1177/0278364913491297.

[27] J. R. Clough, N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schnabel, and A. P. King, "A topological loss function for deep-learning based image segmentation using persistent homology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8766–8778, Dec. 2022, doi: 10.1109/TPAMI.2020.3013679.

[28] R. Paul and S. Chalup, "Estimating Betti numbers using deep learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Budapest, Hungary, Jul. 2019, pp. 1–7.

[29] S. K. Giri and G. Mellema, "Measuring the topology of reionization with Betti numbers," *Monthly Notices Roy. Astronomical Soc.*, vol. 505, no. 2, pp. 1863–1877, Jun. 2021, doi: 10.1093/mnras/stab1320.

[30] M. Valvekens, "L^2-Betti numbers of C-tensor categories associated with totally disconnected group," *Int. Math. Res. Notices*, vol. 2022, no. 14, pp. 10704–10766, Jul. 2022, doi: 10.1093/imrn/rnab066.

[31] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 5410–5418, doi: 10.1109/CVPR.2018.00567.

[32] F. Güney and A. Geiger, "Displets: Resolving stereo ambiguities using object knowledge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 4165–4175, doi: 10.1109/CVPR.2015.7299044.

WENBIN ZHU was born in Jiangsu, China, in 1990. He received the B.Sc. degree from Nanjing University, in 2012, the M.S. degree in electrical engineering from North Carolina State University, in 2015, and the Ph.D. degree from the Nanjing University of Science and Technology, in 2022. He is currently a Postdoctoral Researcher with the Nanjing University of Science and Technology. His research interests include radar signal processing, image processing, and target recognition.

HONG GU received the B.Sc. degree from the East China Institute of Technology, in 1988, the M.S. degree from the Nanjing University of Science and Technology (NJUST), in 1991, and the Ph.D. degree from Xidian University, in 1995, all in electronic engineering. He is currently a Professor with the School of Electronic and Optical Engineering, NJUST. His research interests include signal processing and target recognition.

ZHENHONG FAN was born in Jiangsu, China, in 1978. He received the M.Sc. and Ph.D. degrees in electromagnetic field and microwave technique from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 2003 and 2007, respectively. In 2006, he was with the Center of Wireless Communication, City University of Hong Kong, Hong Kong, as a Research Assistant. He is currently a Professor with the School of Electronic and Optical Engineering, NJUST. He is the author or coauthor of more than 50 technical papers. His research interests include computational electromagnetics, electromagnetic scattering, and radiation.

XIAOCHUN ZHU was born in Jiangsu, China, in 1963. He received the B.S. degree from the Nanjing Institute of Technology, and the M.S. degree from the Nanjing University of Science and Technology. He is currently a Professor with the School of Automation, Nanjing Institute of Technology. His research interests include intelligent sensing, numerical control, signal processing, and automation.