



Image-Based Maritime Obstacle Detection Using Global Sparsity Potentials

Xiaozheng Mou* and Han Wang, *Member, KIICE*

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798

Abstract

In this paper, we present a novel algorithm for image-based maritime obstacle detection using global sparsity potentials (GSPs), in which “global” refers to the entire sea area. The horizon line is detected first to segment the sea area as the region of interest (ROI). Considering the geometric relationship between the camera and the sea surface, variable-size image windows are adopted to sample patches in the ROI. Then, each patch is represented by its texture feature, and its average distance to all the other patches is taken as the value of its GSP. Thereafter, patches with a smaller GSP are clustered as the sea surface, and patches with a higher GSP are taken as the obstacle candidates. Finally, the candidates far from the mean feature of the sea surface are selected and aggregated as the obstacles. Experimental results verify that the proposed approach is highly accurate as compared to other methods, such as the traditional feature space reclustering method and a state-of-the-art saliency detection method.

Index Terms: Global sparsity potentials, Horizon detection, Maritime images, Obstacle detection

I. INTRODUCTION

Obstacles in a maritime scenario can be big cargo ships, vessels, yachts, and/or small buoys. Successful detection of these obstacles provides important information for either navigation or object detection and the tracking of unmanned surface vehicles (USVs) [1-3]. The methods for obstacle detection in maritime images can be roughly categorized into two-dimensional (2D) image-based and three-dimensional (3D) stereo vision-based methods. The 3D methods [4, 5] use two stereo images to obtain a 3D point cloud of the scene, which is then applied to fit the sea surface plane and cluster the points above this plane as obstacles. The 2D methods can either use graphic models to separate the sky, horizon, and sea areas and then segment the obstacles from the sea area [6], or apply saliency detection to estimate the

obstacles [7]. In this work, we use a 2D image-based method because of the low camera budget and the relatively small computational burden.

Sparsity potential (SP), originally proposed in [8] for object detection, is a measure that captures the sparseness or similarity of an image patch with respect to its neighborhood. In [8], image patches with a high value of SP are considered to be more discriminative and chosen for training and testing with Hough Forests. However, there are some limitations of applying such a local SP to maritime images because of the unique properties of a sea surface. For example, the neighboring patches of an image patch of the top of a sea wave may contain the bottom of the sea wave; thus, the image patch in the center would be highly distinct from its neighbors and have a low self-similarity or SP value, which would lead to a high probability of being

Received 13 April 2016, Revised 15 April 2016, Accepted 30 April 2016

*Corresponding Author Xiaozheng Mou (E-mail: xmou1@ntu.edu.sg)

School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798.

Open Access <http://dx.doi.org/10.6109/jicce.2016.14.2.129>

print ISSN: 2234-8255 online ISSN: 2234-8883

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

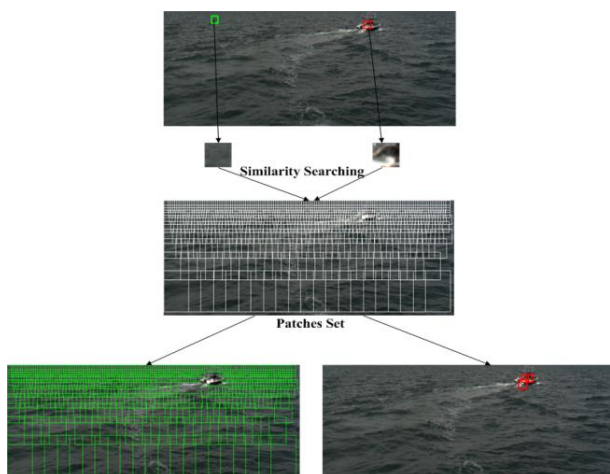


Fig. 1. Each image patch exhibits a different global sparsity potential. The top image shows two image patches from the sea surface (green) and the obstacle (red). After they go through the entire patch set (middle image) for similarity searching, patches similar to them can be retrieved as shown in the two bottom images.

classified as a patch of obstacles. In contrast, if the image patch is sampled from a big cargo ship, it may have a very similar appearance to its neighboring patches; this may lead to a high SP value, and the image patch may be wrongly classified as the background. Therefore, to reduce the abovementioned detection error in maritime images, we propose “global sparsity potential (GSP)”, which computes the self-similarity of an image patch within the entire sea surface area. By taking into account the entire sea surface area, we find that the image patch on the sea has more similar patches in the entire set of patches, while the image patch on the obstacles has the opposite. Thereafter, the discriminative power of image patches increases, and it becomes easier to separate the foreground (obstacles) patches from the background (sea) ones. As illustrated in Fig. 1, one can see that compared to the patch of obstacles, which only has three similar patches in the patch set, the patch of the sea is considerably less sparse and more similar to most of the patches.

In [9], the researchers proposed the use of variable-size image windows and feature space reclustering for detecting obstacles in maritime images. In their work, an iterative reclustering method is proposed to determine the centroid of the main cluster (sea), whose outliers are considered obstacles. Nevertheless, the clustering process in this method is sensitive to the outliers in the computation of the mean or the median of the feature set, thus leading to poor performance when there are a larger number of obstacles or more white wake outliers in the image. To solve this problem, in this study, we omitted the reclustering process and used GSP for estimating the mean feature of the sea. Then, as in [9], the outliers were considered obstacles.

In summary, two contributions are made in this paper.

One is that we introduce a new measure, GSP, to represent the sparseness of an image patch with respect to the entire sea surface area. The other is a novel image-based obstacle detection algorithm using GSP for USVs; this algorithm has been experimentally proven to be more accurate and robust than the traditional method [9] and the state-of-the-art saliency detection method [10].

The rest of this paper is organized as follows: Section II introduces the proposed GSP measure and the proposed algorithm for obstacle detection in maritime images. Section III presents the experimental results with our own dataset and a comparison with other related work. Finally, Section IV concludes this work and discusses the future work in this area of study.

II. PROPOSED ALGORITHM

To ensure that only the sea surface area is processed and the sea surface is the dominating cluster, the proposed algorithm for maritime obstacle detection is based on two assumptions:

- The horizon can be detected in the images.
- Obstacles form a small part (<50%) of the sea area in the images.

Therefore, only the obstacles below the horizon line in the images are considered the detection targets. In general, the proposed algorithm can be divided into three procedures: horizon detection, sampling and representation of image patches, and obstacle detection using GSP.

A. Horizon Detection

In [11], four different horizon detection methods were compared and analyzed, and it was concluded that the Random Sample Consensus (RANSAC) method provides the best results with high accuracy. Therefore, in this work, we apply the RANSAC method for horizon detection. To reduce the computational expense and the noise effect, it is usually better to set a region of interest (ROI) that contains the horizon. However, rather than predefine a fixed ROI for every frame as in [11], we propose a more general method to adaptively estimate the ROI.

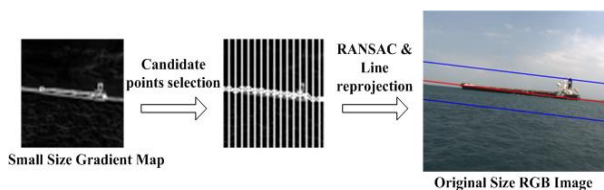


Fig. 2. ROI (area between the two blue lines) estimation for horizon detection. For visual purposes, the size ratio between the small size gradient map and the original image is enlarged.

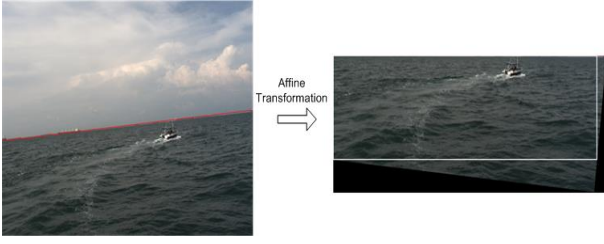


Fig. 3. ROI for obstacle detection. Affine transformation is applied on the left image to horizontalize the horizon line (red), and then, a rectangle area without artificial pixels is cropped as the ROI (white rectangular area in the right image).

We first resize the original image to a smaller size (e.g., 64×64 in this work), because downsampling eliminates a considerable amount of noise, and the horizon can still be roughly estimated without significantly biasing the ground truth. In the downsized image, shown in Fig. 2, first, the gradient map is computed using a Sobel operator, and then, locations with the maximum gradient values along each sampled column are selected as the candidate points; RANSAC is used by randomly selecting two candidate points at each iteration to fit the horizon line. Finally, after reprojecting the estimated horizon in the small image onto the original image, we can define the ROI in the original image by moving the horizon vertically up and down for the same distance to form the upper and lower boundaries, respectively. Thereafter, RANSAC is used again as in [11] in the ROI for estimating a more accurate horizon in the original image.

As shown in Fig. 3, after the detection of the horizon line, the ROI for obstacle detection can be obtained via affine transformation and cropping. Then, further processing is performed on the ROI.

B. Patch Sampling and Representation

In 2D images, considering the geometric relation between the camera and the sea surface, the resolution of the observation that is close to the horizon is smaller than that of the observation close to the image bottom. Similar to [9], here, square image patches are sampled from the ROI by using variable-size image windows with an overlap rate of α and an expansion rate of β ; the minimum window size is $\omega \times \omega$ pixels. An example of this image patch sampling method can be seen in the middle image of Fig. 1, in which the white rectangles denote the sample patches from the whole image.

To represent the abovementioned sample patches, we adopt a gray-level co-occurrence matrix-based texture analysis [12] as in [9]. In this method, all patches are first resized to the same size ($\omega \times \omega$) and then, each image patch f is represented by a four-dimensional vector: $f = [Energy, Entropy, Contrast, Homogeneity]$. Here,

$$Energy = \sum_i \sum_j I^2(i, j), \quad (1)$$

$$Entropy = \sum_i \sum_j I(i, j) \log(I(i, j) + 1), \quad (2)$$

$$Contrast = \sum_i \sum_j (i - j)^2 I(i, j), \quad (3)$$

$$Homogeneity = \sum_i \sum_j I(i, j) / (1 + |i - j|), \quad (4)$$

where i and j denote the row and column indices of the image patch, respectively, and $I(i, j)$ represents the intensity value at pixel location (i, j) .

C. Obstacle Detection Using GSP

Here, we denote an image patch with a feature vector expression as f_k , and the entire image patch set as $\mathbf{F} = \{f_k, k = 1, 2, \dots, N\}$, where N represents the total number of patches sampled in the ROI of an image.

1) Measure of GSP

In this work, the GSP of an image patch is measured by its global self-similarity, which computes the similarity of a query patch to the entire patch set. Different from [13], which extracts the global self-similarity descriptors by performing a cross correlation of the patches in the entire image for object classification and detection, we measure the texture similarity of a query patch to the entire patch set by respectively computing their Mahalanobis distances. The smaller the distance between two patches, the higher is the similarity between them. Then, all the computed distances to the patch set are summed and their average is taken as the global self-similarity measure of this query patch. Eq. (5) formulates the global self-similarity G_k of patch f_k in an image.

$$G_k = \frac{1}{N-1} \sum_{h=1, h \neq k}^N \sqrt{(f_k - f_h)^T C^{-1} (f_k - f_h)}, \quad (5)$$

where C denotes the covariance matrix of the feature set. Then, G_k is normalized to be in $[0, 1]$ with 0 representing the most similar and 1 the most dissimilar.

2) Clustering of Features

In [9], the centroid of the main cluster (sea) is estimated using an iterative procedure. However, this method may be sensitive to the outliers, because at each iteration, it treats all image patches equally in order to compute the mean or median feature. Therefore, including these outliers in the estimation of the centroid of the sea may decrease the accuracy. In this work, to overcome the abovementioned drawback, we propose to select image patches with a high probability to be a sea surface, i.e., having a relatively low GSP value, to estimate the centroid (mean feature) of the sea.

Thereafter, the procedure for feature clustering can be summarized as follows:

1. Image patches with GSP $G_k \leq \tau_1$ (τ_1 denotes the threshold) are selected to compute the mean feature μ of the sea.
2. The Mahalanobis distance d_r between image patches with GSP $G_k > \tau_1$ and μ are calculated respectively, and patches with $d_r > \tau_2$ (τ_2 denotes the threshold) are considered to the patches of obstacles.
3. All the detected obstacle patches are grouped and merged to form the obstacle boundaries, and then, an inverse affine transform with respect to the horizon line is applied to map the location of the detected obstacles in the original image.

III. EXPERIMENTAL RESULTS

Since there are few available public datasets for maritime obstacle detection, we built our own dataset, the details of which are described in Section III-A. Using this new dataset, we evaluated the accuracy of the proposed algorithm and compared its performance with that of the traditional method [9] and that of a state-of-the-art saliency detection approach [10] in Section III-B.

A. Dataset

Our maritime obstacle detection dataset consists of four sequences (S_#1, S_#2, S_#3, and S_#4), which were captured by a Point Grey grasshopper CCD camera mounted on a moving USV on the sea. Each sequence contains 600 RGB frames (size: 684×548 pixels). The obstacle in this dataset is a moving target boat, which varies its distance (approximately from 50 m to 500 m) to the USV (camera boat). The different sequences present different challenges:

S_#1 characterizes the detection ability for a short distance, in which the target boat moves close to the USV (within 100 m);

S_#2 contains many white wake outliers generated by the fast moving target boat, and the distance is around 100 m to 200 m.

S_#3 has a majority of the frames without the obstacle shown, and the target boat quickly moves 200 m away from the USV, and from the left to the right of the image in a few frames at the middle of the sequence; this also provides some white wake outliers.

S_#4 renders the challenge of distant obstacle detection, and the target boat moves a distance of 200 m to 500 m away.

B. Performance Evaluation

Since the proposed algorithm is based on the prior

knowledge of the horizon line, only images whose horizons are detected can be processed for the obstacle detection. Thus, images without a detected horizon are discarded, and not considered in the accuracy or the false rate calculation of the obstacle detection.

The parameters for variable-size windows in the image patch sampling part are set as follows: overlap rate $\alpha = 33\%$ expansion rate $\beta = 6\%$, and the minimum window size is 16×16 . In Section II-C-2), the thresholds τ_1 and τ_2 are set to 0.1 and 0.9, respectively.

Similar to [14], the accuracy evaluation is performed visually as follows:

1. For frames with an obstacle presented, the detection result for each frame is classified as the detected bounding box b being either *correct* (most part of b contains the obstacle), *halfcorrect* (about half part of b contains the obstacle), or *false* (very small or no part of b contains the obstacle). Accordingly, the assigned numeric value $\rho(b)$ is 1, 0.5, or -1, respectively.
2. For frames without the obstacle shown, the detected bounding box b is classified as either *correct* (b is not detected in the result, i.e., void detection) or *false* (b is detected in the result, i.e., false detection). Accordingly, the assigned numeric value $\rho(b)$ is 1 or -1, respectively.

Integrating the two above-discussed cases, we can express the score assigned to b as follows:

$$\rho(b) = \begin{cases} 1, & b = \text{correct}, \\ 0.5, & b = \text{half} - \text{correct}, \\ -1, & b = \text{false}. \end{cases} \quad (6)$$

Finally, as formulated in (7), the detection accuracy ξ and the false detection rate η can be calculated using all assigned values $\rho(b)$ in each sequence.

$$\begin{cases} \xi = \frac{1}{n_+} \sum_{l=1}^{n_+} \rho(b_l), & \rho(b_l) > 0, \\ \eta = \frac{1}{n_-} \sum_{l=1}^{n_-} |\rho(b_l)|, & \rho(b_l) < 0, \end{cases} \quad (7)$$

where n_+ denotes the sum of the total number of ground-truth obstacles and the total number of frames without an obstacle in each sequence, and n_- represents the total number of frames in each sequence.

Table 1 summarizes the performance of obstacle detection using the proposed algorithm and two comparative algorithms.

C. Comparisons and Analysis

The main difference between the proposed algorithm and the feature space reclustering method [9] is the computation

of the centroid of the sea features. In [9], the authors proposed the use of all the features of the sampled image patches to estimate the centroid iteratively, while our method involves the selection of image patches with small values of GSP and the calculation of their mean to estimate the centroid. Theoretically, the proposed algorithm is more insensitive to the outliers of the sea features, because rather than taking all the features, which may contain many outliers, to compute the mean or median feature, we just use features with high probabilities to be the sea to compute the mean feature. We reimplemented the method of [9] with the same parameter settings for variable-size window sampling and feature extraction on our maritime obstacle detection dataset. Experimental results show that the proposed algorithm is more accurate than that proposed in [9].

As shown in Table 1, the accuracy of the proposed algorithm is more than 10% compared to that proposed in [9] in the first three sequences, which contain many outliers caused by the white wake. Nevertheless, in sequence S_#4, our method performs only slightly better than that proposed in [9]. This could be attributed to the fact that the obstacles in most frames of S_#4 are far away from the camera, so there are very few outliers, such as white wake, in the images. Fewer outliers lead to more accurate estimation of the centroid of the sea; thus, only a small accuracy gap exists between our method and that proposed in [9]. In

Table 1. Comparison of accuracy (Acc) and false rate (FR) for maritime obstacle detection using different methods

Seq.	[10]		[9]		Proposed algo.	
	Acc	FR	Acc	FR	Acc	FR
S_#1	0.662	0.214	0.782	0.216	0.893	0.132
S_#2	0.056	0.933	0.652	0.381	0.842	0.230
S_#3	0.827	0.149	0.814	0.195	0.917	0.124
S_#4	0.053	0.925	0.927	0.151	0.965	0.058
Avg.	0.400	0.555	0.794	0.236	0.904	0.136

addition, our method exhibits a smaller false detecting rate than that proposed in [9].

Some advantages of the proposed algorithm over the method proposed in [9] can be seen in Fig. 4. In Fig. 4(a) and (c), false detection caused by white wake happens in the case of the method proposed in [9]. In Fig. 4(f), the method proposed in [9] has two detections, in which one is correct and the other is false and is caused by a sea wave; the same false detection is also observed in the scenario of Fig. 4(d). Only a small portion of the boat is detected by the method proposed in [9] in Fig. 4(e), and this situation is classified as *false* in Eq. (6). Fig. 4(b) shows the missed detection for the method proposed in [9], which wrongly detects nothing for this frame.

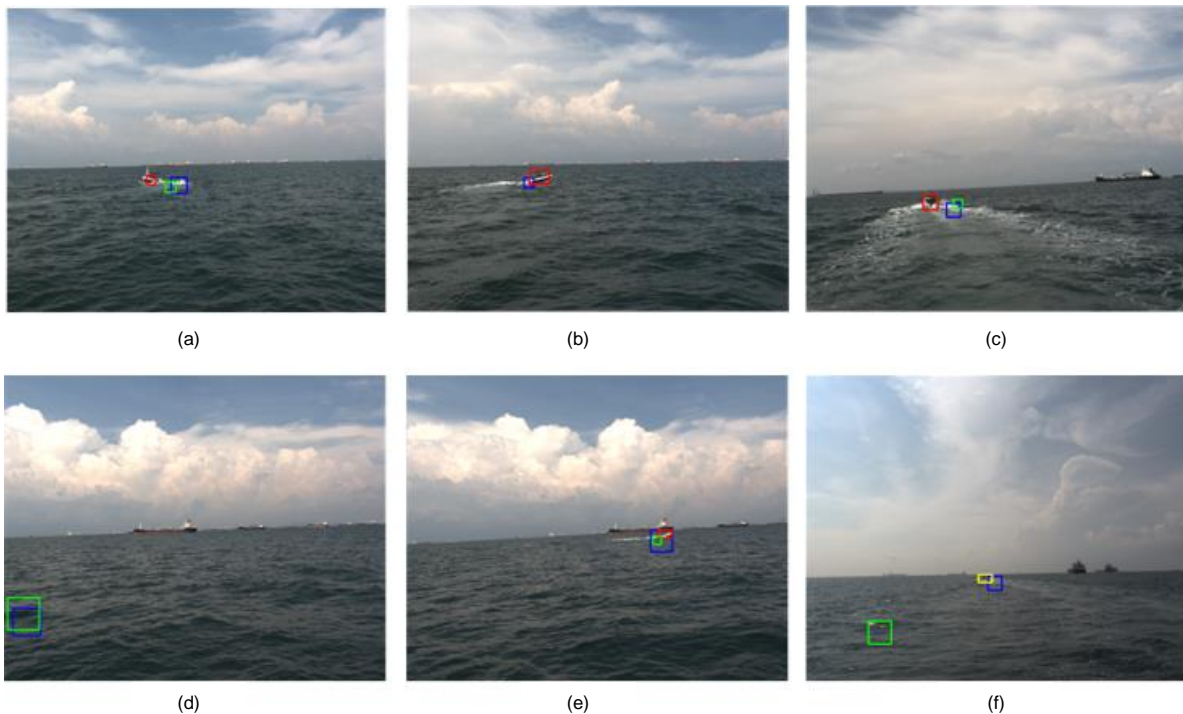


Fig. 4. Superior performance for maritime obstacle detection of the proposed algorithm (red) compared to that of the method of feature space reclustering [9] (green) and that of saliency detection VOCUS2 [10]. The yellow bounding box in (f) means that the red and the green bounding boxes are overlaid. (a) and (b) are from S_#1; (c) is from S_#2; (d) and (e) are from S_#3; and (f) is from S_#4. One can see that the performances of these three methods can be easily evaluated by human eyes.

To test the saliency detection method for our task, we implemented the work of [10] with our own dataset. As shown in Table 1, however, this state-of-the-art method for saliency detection does not perform well for our dataset. However, intuitively, it seems that the obstacles on the sea are more distinct and salient than the sea water. In fact, the saliency-based methods, which usually use the image local contrast information to detect distinct regions, are sensitive to the sea wave. For example, it can be seen in Fig. 4(a), (b), (c), and (f) that the white wake generated by the boat causes a big problem for this saliency detection method, which results in many false detections. Similarly, in Fig. 4(d), the dark region between the wave top and the wave bottom is detected as the saliency, but this region is not an obstacle. Although the detected saliency in Fig. 4(e) contains the obstacle, the bounding box is very big and is not well-fitted to the obstacle. Therefore, we can conclude that it may not be a wise choice to only apply saliency detection to solve the maritime obstacle detection task.

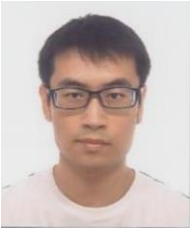
IV. CONCLUSION

In this paper, we introduced a new measure, “global sparsity potential (GSP)”, to capture the sparseness of an image patch throughout the sea area. Using GSP, we developed an accurate and robust approach for moving camera-based obstacle detection in maritime images. In this approach, image patches with a relatively small GSP value are considered the main cluster (i.e., sea surface), while their outliers, which have a relatively large GSP value and a relatively large Mahalanobis distance with respect to the mean feature of the sea surface), are considered the obstacles.

Although the proposed algorithm exhibits good performance, only the intensity image and the texture feature are explored. Further improvements can be expected by combining the color information and other discriminative features in a future work.

REFERENCES

- [1] T. Huntsberger, H. Aghazarian, A. Howard, and D. C. Trotz, “Stereo vision-based navigation for autonomous surface vessels,” *Journal of Field Robotics*, vol. 28, no. 1, pp. 3-18, 2011.
- [2] D. Bloisi, L. Iocchi, M. Fiorini, and G. Graziano, “Camera based target recognition for maritime awareness,” in *Proceeding of 15th International Conference of Information Fusion*, Singapore, pp. 1982-1987, 2012.
- [3] S. Fefilyatjev, and D. Goldgof, “Detection and tracking of marine vehicles in video,” in *Proceeding of 19th International Conference on Pattern Recognition (ICPR)*, Tampa, FL, pp. 1-4, 2008.
- [4] H. Wang, and Z. Wei, “Stereo vision based obstacle detection system for unmanned surface vehicle,” in *Proceeding of IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Shenzhen, China, pp. 917-921, 2013.
- [5] H. Wang, X. Mou, W. Mou, S. Yuan, S. Ulun, S. Yang, and B. S. Shin, “Vision based long range object detection and tracking for unmanned surface vehicle,” in *Proceeding of IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, Siem Reap, Cambodia, pp. 101-105, 2015.
- [6] M. Kristan, V. Sulic Kent, S. Kovacic, and J. Pers, “Fast image-based obstacle detection from unmanned surface vehicles,” *IEEE Transactions on Cybernetics*, vol. 43, no. 3, pp. 641-654, 2015.
- [7] H. Wang, Z. Wei, S. Wang, C. S. Ow, K. T. Ho, and B. Feng, “A vision-based obstacle detection system for unmanned surface vehicle,” in *Proceeding of IEEE 5th Conference on Robotics, Automation and Mechatronics (RAM)*, Qingdao, China, pp. 364-369, 2011.
- [8] N. Razavi, N. S. Alvar, J. Gall, and L. J. Van Gool, “Sparsity potentials for detecting objects with the Hough transform,” in *Proceeding of the British Machine Vision Conference (BMVC)*, Surrey, UK, pp. 1-10, 2012.
- [9] P. Voles, A. A. W. Smith, and M. K. Teal, “Nautical scene segmentation using variable size image windows and feature space reclustering,” in *Proceeding of the 6th European Conference on Computer Vision (ECCV)*, Dublin, Ireland, pp. 324-335, 2000.
- [10] S. Frintrop, T. Werner, and G. M. Garcia, “Traditional saliency reloaded: a good old model in new shape,” in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, pp. 82-90, 2015.
- [11] Y. Yan, B. S. Shin, X. Mou, W. Mou, and H. Wang, “Efficient horizon detection on complex sea for sea surveillance,” *International Journal of Electrical, Electronics and Data Communication*, vol. 3, no. 12, pp. 49-52, 2015.
- [12] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610-621, 1973.
- [13] T. Deselaers and V. Ferrari, “Global and efficient self-similarity for object classification and detection,” in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, pp. 1633-1640, 2010.
- [14] B. S. Shin, J. Tao, and R. Klette, “A superparticle filter for lane detection,” *Pattern Recognition*, vol. 48, no. 11, pp. 3333-3345, 2015.



Xiaozheng Mou

received his bachelor's degree in Automation (Application of Information Technology) from Donghua University, Shanghai, China, in 2010. He is now a project officer and is working towards his Ph.D. from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include visual object detection and tracking.



Han Wang

received his bachelor's degree in Computer Science from Northeast Heavy Machinery Institute (China), and his Ph.D. from University of Leeds (UK). He is currently an associate professor at School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include computer vision and robotics.