Are Object Detection Assessment Criteria Ready for Maritime Computer Vision?

Dilip K. Prasad, Huixu Dong, Deepu Rajan, and Chai Quek

***Abstract*— Maritime vessels equipped with visible and infrared cameras can complement other conventional sensors for object detection. However, application of computer vision techniques in maritime domain received attention only recently. The maritime environment offers its own unique requirements and challenges. Assessment of the quality of detections is a fundamental need in computer vision. However, the conventional assessment metrics suitable for usual object detection are deficient in the maritime setting. Thus, a large body of related work in computer vision appears inapplicable to the maritime setting at the first sight. We discuss the problem of defining assessment metrics suitable for maritime computer vision. We consider new bottom edge proximity metrics as assessment metrics for maritime computer vision. These metrics indicate that existing computer vision approaches are indeed promising for maritime computer vision and can play a foundational role in the emerging field of maritime computer vision.**

***Index Terms*— Object detection, marine vehicles, intelligent vehicles, performance evaluation.**

I. Introduction

MARITIME vessels (MV) are equipped with sensors such as radar, sonar and LIDAR for situational awareness. The automatic identification system (AIS) supports traffic data exchange over maritime communication channels, through which each MV with on-board AIS declares its position, speed, and intended path. The International Regulations for Preventing Collisions at Sea 1972 (COLREGs) impose that all cargo ships weighing more than 300 tonnes and all passenger ships are equipped with AIS. There is no such imposition on smaller MVs, including fishing boats and small-medium sized cargo MVs. Such MVs are invisible in traffic data. Moreover, the AIS channel may be inaccessible for several minutes to few hours at a time [1]. Cameras in the visible and infrared (IR) range now play a complementary role by overcoming disadvantages of traditional sensors like the minimum range associated with radar and sonar [2]. Thus, *computer vision (CV) techniques should play an important role in detecting objects in the maritime environment*, especially in detecting small and medium sized MVs that have weak radar or sonar signatures and lack on-board AIS.

Manuscript received November 13, 2018; revised May 13, 2019 and July 31, 2019; accepted November 6, 2019. Date of publication November 25, 2019; date of current version November 30, 2020. This work was supported by UiT The Arctic University of Norway. The Associate Editor for this article was K. Wang. *(Corresponding author: Dilip K. Prasad.)*

D. K. Prasad is with the Department of Computer Science, UiT The Arctic University of Norway, 9037 Tromsø, Norway (e-mail: dilipprasad@gmail.com).

H. Dong is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA.

D. Rajan and C. Quek are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798.

Digital Object Identifier 10.1109/TITS.2019.2954464

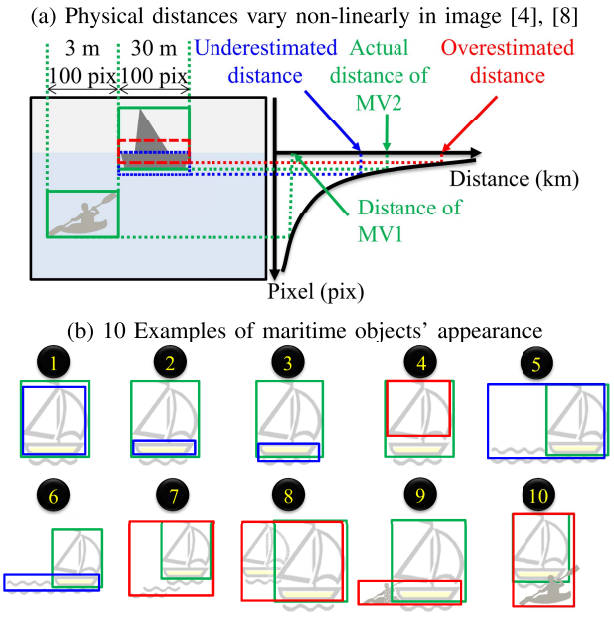


Fig. 1. What is an acceptable detection of a maritime vessel? (a) Collision avoidance requires accurate estimate of the distance, which is related to the bottom edge of the vessel, and the minimum span of a maritime object. (b) Green, blue, and red boxes denote ground truth, acceptable detection, and unacceptable detections, respectively.

Maritime CV for object detection faces several challenges. Maritime video streams are characterized by *scene flatness*, i.e. lack of landmarks and marked lanes as in roads. The maritime scene offers *difficult to model dynamic background* features because of challenges such as a semi-stochastic wave background, the sharp contrasts of wakes, possibilities of occlusion of MVs, and weather and illumination conditions such as rain, haze and glint [3]. Further, planning the manoeuver and deceleration for collision avoidance (CA) is challenging since the distance and span of the MVs in the scene is related non-linearly to the pixels along the $y-$axis [4], [8], see Fig. 1(a). There are other applications also, that face the same non-linearity between the physical space and image space. For example, a reviewer of this manuscript suggested terrestrial applications, where "obstacle detection by automative vehicle sensors (for automated braking for example) has the same bias, since the flat world assumption is usually used in this domain too." An appropriate maritime CV solution has to satisfy the following requirements:

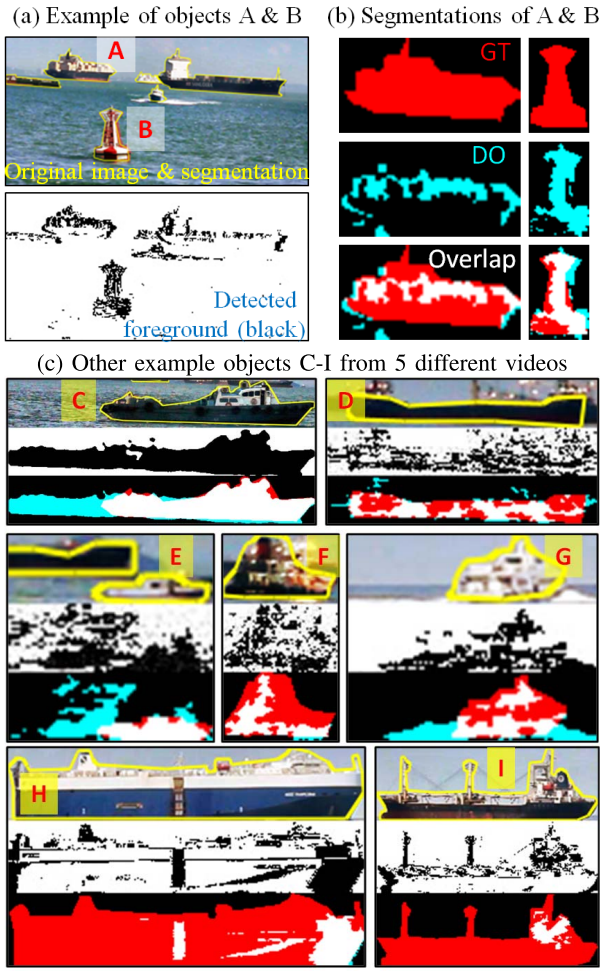
- detect and track MVs in the scene
- determine MVs' accurate spans, positions and tracks
- provide real-time results
- perform in all weather and illumination

Detection and tracking of MVs falls under the ensemble problem set of 'detection and tracking in a dynamic background', which has been extensively studied in computer vision. The existing CV solutions in this ensemble can provide a firm foundation for developing dedicated CV solutions for maritime object detection requirements. We note that the above identified goals of maritime CV comprise a broad topic and entail research for several years to come. In this paper, we choose a very specific problem within this broad scope and critical for the entailing research. The specific problem considered in this is paper as follows. Adoption of existing CV solutions for maritime CV encounters a set back. We show that traditional performance measures for object detection fail in the maritime environment and we discuss the following question. *How do we assess the quality of detection for maritime computer vision?*

We show that assessment metrics such as intersection over union (IOU, also called Jaccard index [5]) and intersection over ground truth (IOG, also called sensitivity [6]), most often used in object detection, are unsuitable for maritime CV. They are deficient in assessing the accuracy of span and distance of detected MVs. Either the detection method provides a very high IOU, say 90%, or customized assessment metric is needed to meet the requirements of maritime CV. *The aim of this paper is to design custom assessment metrics that provide good assessment of the quality of detected objects while not putting severe demands on detection algorithms.*

We discuss two new assessment metrics customized for maritime computer vision. We also study the performance of existing background subtraction (BGS) algorithms and regions with convolution neural network (R-CNN) features using conventional and proposed assessment metrics. We show that the conventional metrics indicate general unsuitability of BGS algorithms for maritime CV whereas the new metrics present hope of using them in maritime CV. We expect that this exercise shall provide useful cursors for developing maritime CV solutions.

The assessment requirements of maritime CV are discussed in section II. The deficiency of conventional metrics for maritime CV is discussed in section III. The proposed bottom edge proximity metrics are presented and compared with conventional metrics in section IV. Experimental results of existing BGS algorithms and R-CNN on a maritime dataset are presented in section V. Section VI concludes this paper with a discussion on the future outlook for maritime CV.



(d) Comparison of pixel-based and BB based segmentations

Object	IOU		IOG	
	Pixel-based	BB-based	Pixel-based	BB-based
A	0.27	0.64	0.29	0.83
B	0.55	0.85	0.64	0.85
C	0.68	0.65	0.95	0.96
D	0.45	0.60	0.49	0.97
E	**0.39**	**0.22**	0.87	1.00
F	0.39	0.89	0.40	0.99
G	0.36	0.42	0.47	0.91
H	0.15	0.47	0.15	0.48
I	0.09	0.10	0.09	0.10

Fig. 2. Pixel segmentations are more demanding than bounding boxes, see subfigures (a-c) for qualitative examples. The same metrics result into significantly lower values when computed for pixel segmentations, as noted in subfigure (d). The only exception in the examples considered in (a-c) is shown in bold in (d). In the subfigure (c), there are 3 panels for each object. The top, middle, and bottom panels respectively show GT in image, detected foreground pixels, and the overlap of DO and GT.

II. Requirements for Maritime CV

Before discussing the suitability of conventional metrics, or lack thereof, we consider the fundamental question: 'What is an acceptable detection of a maritime vessel?'. It is important to accurately estimate the location of the MV in a scene (given by the bottom edges of the MV) and its minimum span (determined by the width of the MV in pixels and its position in the image frame). See Fig. 1(a) for illustration. Consider the example cases 1-10 shown in Fig. 1(b). Example 1 is close to ideal, where the bounding box (BB) of the detected object (DO) is almost the same as the BB of the ground truth (GT). We restrict our discussion to bounding boxes because the pixel-segmentations are significantly more demanding than bounding boxes. We illustrate this point using Fig. 2. Fig. 2(a) shows two objects A and B in an image and also the foreground segmentation result obtained using a dynamic background subtraction method. Their pixel segmentations of GT and DO are shown in Fig. 2(b). Other examples are shown in Fig. 2(c). Fig. 2(d) shows values of IOU and IOG for pixel and BB segmentations. The small values

of IOU and IOG for the pixel segmentations of almost all the objects indicate that assessing the pixel segmentations is more demanding. Moreover, the pixel segmentations are not particularly more informative than BBs about the distance and span of the vessel anyway. Yet, at least for one example, i.e. object E in Fig. 2(c,d), the IOU for pixel based segmentation is larger than BB segmentation. Therefore, the importance of pixel segmentations in accurate detection of MVs cannot be discounted. It merits an elaborate study, which we relegate to the future work.

Although there is a large variety of MVs, in general, an MV is characterized by a hull and an optional super-structure, i.e. all parts above the hull, including masts. The existing CV solutions may detect hull and super-structure separately due to two reasons. First, super-structure is not an essential component and supervised learning approaches may undertrain for vehicles with super-structures. Second, stark differences in geometries, color, and other image features of the hull and the super-structure imply that the super-structure may appear as an independent object. The hull or the super-structure may even be left undetected, such as in the case of sailboats, due to a lack of contrast between the background and the super-structure. Consequently, the DO may appear as shown in examples 2-4. *For collision avoidance, accurate detection of the hull is important, irrespective of whether the super-structure is included in the DO with the hull (example 1), detected independently (example 4), or not detected at all (examples 2 and 3).* Furthermore, the physical distance between the MV and the sensor is mapped non-linearly in an image along a direction perpendicular to the horizon (see Fig. 1(a)). This means that the line in image corresponding to horizon is at infinity while the bottom most pixel is only a few meters away from the sensor. Thus, *incorrect estimation of bottom of hull may result in hugely incorrect estimation of the physical distance*. However, it is preferable to slightly underestimate the distance between the sensor and an MV for collision avoidance, rather than overestimate it. In this sense, DOs in examples 2 and 3 of Fig. 1(b) are acceptable.

Current BGS solutions for object detection struggle with the presence of wakes of maritime vessels [3]. Often wakes are detected as part of the MVs, such as shown in examples 5-7. Similar to the logic of underestimating the distance between the sensor and the detected MV, it is safer if the estimated width is not lesser than the actual span. Thus, horizontal wakes becoming a part of DO is acceptable, though not preferable. However, *large extension of the DO in the vertical direction below the hull may result in grossly incorrect estimate of distance, and is not preferred* (see example 7).

The condition of occlusion has a significant implication on collision avoidance. The extension of DO due to occlusion in any direction may mean that the MV with smaller pixel footprint is not detected (see examples 8-10). Though the DOs for all these examples are not preferred, the implications are much more severe for examples 9-10, which involve a small MV (kayak) with no on-board communication channel and poor detectability in radar and sonar. These situations call for a close to perfect overlap between the DO and the GT. However, even between examples 9 and 10, example 10 is the least preferred detection. In example 10, the DO leads to gross underestimation of the location of large MV and missed detection of a kayak that is much closer to sensor, much agile, and invisible in other sensor streams.

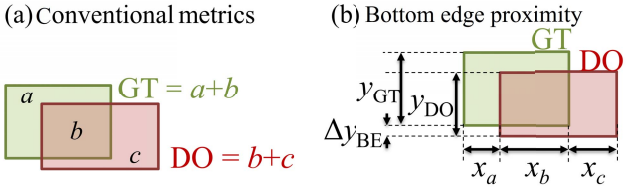


Fig. 3. The notations relevant to the conventional metrics and the proposed bottom edge proximity (BEP) metrics are shown here.

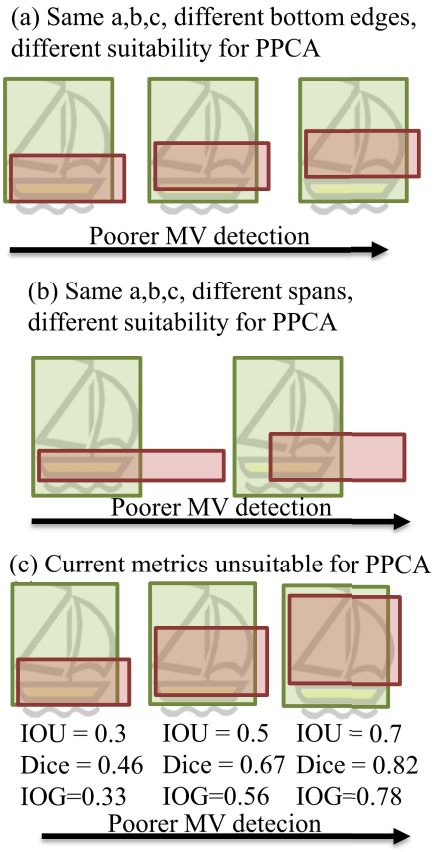


Fig. 4. The current metrics are unsuitable for assessing detected objects in maritime CV. For the same values of a, b, and c, one DO may be preferred over others (a,b). Increasing IOU, Dice Index, or IOG metrics need not indicate better detections (c).

III. Conventional Assessment Criteria Versus the Needs of Maritime CV

Assessment of the quality of detection is usually performed through similarity metrics, such as Jaccard index [5] (also called IOU) or Dice index [7]. Their generalized form is given by Twersky index [9], defined as follows:

$$S = \frac{b}{b + \alpha a + \beta c} \tag{1}$$

where a, b, c are the areas of (GT − DO), (GT ∩ DO), and (DO − GT), respectively (see Fig. 3(a)). The parameter

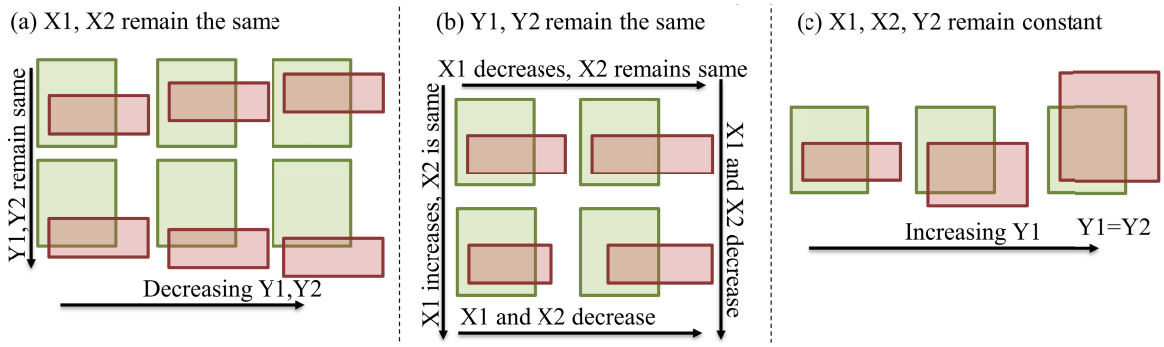


Fig. 5. BEP is sensitive to the bottom edges of the DO and GT (a). X_1 is more strict than X_2 (b). Y_1 is more strict than Y_2 (c). Thus BEP_1 is more strict than BEP_2.

α emphasizes the allegiance of the overlapped region with GT while the parameter β emphasizes the allegiance of the overlapped region with DO. Similarity metrics usually employ symmetry with respect to GT and DO, i.e. $\alpha = \beta$. Dice index corresponds to $\alpha = \beta = 0.5$ and widely used IOU corresponds to $\alpha = \beta = 1.0$. A detection is assessed as true positive if IOU $> c_0$. Similar threshold is employed if other similarity metrics are used. Usually in CV, IOU > 0.5 is considered sufficient. We consider an additional asymmetric metric with $\alpha = 1$, $\beta = 0$, which we refer to as intersection over ground truth (IOG). This metric assesses the intersection area b with respect to the area of GT $(a + b)$ only. Thus excess span detection due to wakes (examples 5-7 in Fig. 1(b)) or excess detection in vertical direction below the hull (example 3 in Fig. 1(b)) do not affect the assessment negatively if the metric IOG is used.

The essential problem with the above metrics is that two cases may have the same areas a, b, c, but one case may be a preferred detection over another. See Fig. 4(a,b) for examples. Also, the increasing value of the above mentioned metrics need not imply better detection, as shown in Fig. 4(c). New metrics that account specifically for the importance of the bottom edge of the hull are needed.

IV. Proposed Bottom Edge Proximity Criteria

We consider two new criteria that specifically judge the accuracy of detection of the bottom edge (BE) and the span of the DO. We call them bottom edge proximity 1 (BEP_1 appears here for the first time) and bottom edge proximity 2 (BEP_2, recently proposed in [10]). BEP_1 is symmetric with respect to DO and GT while BEP_2 is biased towards allegiance with GT. We use the notations in Fig. 3(b) for the definitions of BEP_1 and BEP_2 presented next.

A. Bottom Edge Proximity 1 (BEP_1)

We define $\text{BEP}_1 = X_1\ Y_1$ where

$$X_1 = \frac{x_b}{x_a + x_b + x_c}; \quad Y_1 = 1 - \frac{\Delta y_{\text{BE}}}{\min(y_{\text{GT}}, y_{\text{DO}})} \tag{2}$$

The smaller the distance between the edges of the GT and DO, the larger is Y_1. See Fig. 5(a) for an illustration of this point. However, if the DO is significantly smaller than GT, Y_1 becomes poorer. Thus, it indirectly embeds the vertical size of DO in comparison with GT. This is shown in Fig. 5(c).

B. Bottom Edge Proximity 2 (BEP_2)

We define $\text{BEP}_2 = X_2\ Y_2$ where

$$X_2 = \frac{x_b}{x_a + x_b}; \quad Y_2 = 1 - \frac{\Delta y_{\text{BE}}}{y_{\text{GT}}} \tag{3}$$

We note that BEP_1 is stricter than BEP_2. This is because X_1 is less tolerant to extended span of DO due to wakes as well as occlusions, as shown in Fig. 5(b). Further, Y_1 is sensitive to the size of DO if the DO is smaller than the GT, as shown in Fig. 5(c).

For convenience, we refer to X_1 and X_2 as X metrics. Similarly, we refer to Y_1 and Y_2 as Y metrics. An advantage of BEP metrics is that the threshold(s) for assessing a detection as a true positive can be chosen flexibly. Either a single threshold c_0 can be used for the net BEP score, or two thresholds x_0 and y_0 can be considered for X and Y metrics independently, and a TP can be assessed if both conditions $X > x_0$ and $Y > y_0$ are satisfied.

C. Qualitative Comparison for Examples in Fig. 1(b)

We perform a qualitative comparison of the metrics IOU, Dice index, IOG, BEP_1, and BEP_2 on the examples in Fig. 1(b), which were used to study acceptable and unacceptable detections for maritime CV. The results are shown in Table I. We briefly discuss the selection of the thresholds (given in parentheses) for the metrics. Since the threshold value of $c_0 = 0.5$ is conventionally used in object detection [11], we use this value for IOU. Similarly, we use $c_0 = 0.5$ as threshold for the Dice index and IOG as well. Since X_1 and X_2 are 1-dimensional analogues of the 2-dimensional IOU and IOG, we use a threshold value of $x_0 = \sqrt{0.5}$. Lastly, we use threshold value of $y_0 = 0.75$ because the accuracy of bottom edge is critical in collision avoidance.

As discussed before, conventional metrics that use a, b, c shown in Fig. 4 are not suitable for assessing detections in maritime CV. This is evident in Table I, where IOU, Dice index, and IOG have successes for less than half the number of examples. BEP_1 performs better, getting 6 successes out of 10 examples. BEP_2 performs the best, getting success in all the 10 examples. We further study the X and Y metrics, also provided in Table I. Notably, X_2 is less strict in assessing TPs, assessing all DOs as true positives. In BEP_2, Y_2 consequently

plays the role of suitable metric, providing correct assessment for all the 10 examples. Y_1 is only slightly poorer than Y_2, providing 8 correct assessments out of 10. Thus, the role of bottom edge in correct assessment is verified.

The general criteria of assessing the pixel-based semantic segmentation are the same for maritime CV, where distance and span of an MV are important considerations. The bottom most pixels in semantic segmentations, which also form the bottom edge in a bounding box, are the most important determinant of the distance. The widest span of the semantic segmentation, which also forms the width of a bounding box, is the determinant of the span of the vessel. Therefore, the concept of BEPs is generally applicable to semantic segmentation as well.

V. Experiments and Results

Detection of MVs in a maritime environment falls under the ensemble problem set of 'detection in dynamic background'. CV methods solve it by modeling and subtracting the dynamic background, followed by segmentation of the foreground [52], [53]. The dataset and the dynamic background subtraction methods used here are described below. We consider deep learning also for detection of MVs. These details are presented, followed by quantitative and qualitative results.

A. Dataset

We use on-shore (fixed camera) visible range maritime videos from the maritime dataset, namely Singapore maritime dataset, published with [3]. There are 34 high-definition videos taken from Canon 70D cameras, Canon EF 70-300mm f/4-5.6 IS USM. The dataset has been captured at different times, such as before sunrise, at sunrise, at mid day, in the afternoon, in the evening, and 2 hours after sunset. We excluded the videos taken in haze and rain to avoid additional challenges. BBs of objects in each frame of the video are provided along with the dataset. Each BB is labeled with one of the following class labels: boat, buoy, ferry, flying bird/plane, kayak, sailboat, speed boat, vessel/ship, and others. We have not included on-board videos for the reason we explain next. The motion of the vessel on which camera is mounted with respect to water and horizon presents additional challenges for dynamic background detection. The static background methods that use only current frame for background modeling are better candidates, but they have been shown to present extremely poor performance for maritime scenes [3].

B. Dynamic Background Subtraction (BGS) Methods Tested

We tested 22 BGS methods from the BGS library named bgslibrary [20], [46] and 14 BGS methods from the low rank and sparse (LRS) tools library name lrslibrary [47]. The methods in BGS library are implemented in C++. The methods in LRS library are implemented in Matlab. All the methods were executed on Intel i7 6500 U @2.5 GHz desktop with 16 GB RAM and Linux platform. Default parameters have been used for all the methods. Parameter tuning for achieving the best performance for each method is out of the scope of this work. Yet, we note that fine tuning the control parameters for each method is likely to have a positive effect on the quality of detections, and is likely to impact all the metrics positively. All detected BBs less than 20 pixels in any dimension are rejected as obviously spurious detections. We group the 36 methods into six broad categories based on their central concept. The groups and the methods in each of them are listed in Table II. Among the 36 methods, only IMBS has been developed specifically for maritime scenes.

C. Regions With Convolution Neural Network (R-CNN) Features for Detection Using Deep Learning

We conducted two experiments in deep learning. These experiments were executed in Matlab on NVidia DGX-1 graphics server and Linux platform. The standard procedure of applying non-max suppression has not been used for the reason explained next. Many overlapping objects may be present in a maritime scene. Consider Fig. 6(b) for example. The GT bounding box of the vessel A overlaps with the GT bounding boxes of the vessels F, G, H, and I. Even if the DOs corresponding to them might be accurate, applying non-max suppression will result into lower recall because it will suppress either the DO of object A and other objects. First, we randomly selected 20 videos from the dataset for training and trained R-CNN [48] with AlexNet architecture. The results for this experiment were extremely poor and are not reported here. We attribute the poor performance to the challenging nature of the maritime scene and consider that maritime scenes may require camera and illumination specific training. In the second experiment, we formed the training dataset using every fifth frame of all the videos. The objective was to test if R-CNN can detect the objects it has been trained for. R-CNN trained on CIFAR-10 [54] performed poorly but R-CNN trained on ImageNet [55] provided better results. We note that R-CNN experiments may be considered to have unfair advantage over the other methods tested in this paper because the R-CNN experiments use training on a subset of images drawn from the main set itself. We note also that use of R-CNN here [48] is a first attempt of deep learning for maritime CV. Better suited approaches may be identified in the future. Some options include faster R-CNN [49], long-term temporal convolution CNNs [50], networks on convolutional feature maps CNN [51].

D. Qualitative Examples

We consider four example frames, each taken from a different video of the dataset. The detection results of 10 BGS methods and R-CNN are shown in Fig. 6. The selected BGS methods are the ones that consistently outperform other methods in their groups either is precision or in recall. These methods are identified in Table III. All BGS methods are ineffective in subtracting the background. In Fig. 6, all BGS methods except SuBSENSE detect false positive objects in the water background. This problem is more severe in frames 3 and 4, which show relatively more turbulent waters.

TABLE I

QUALITATIVE COMPARISON OF METRICS FOR EXAMPLES IN FIG. 1(B) IS GIVEN HERE. THE THRESHOLDS USED FOR DETERMINING TPS ARE GIVEN IN PARENTHESES. FOR BEPS, (x_0, y_0) ARE GIVEN. THE NUMBER OF SUCCESSES IS THE NUMBER OF TIMES A METRIC ASSESSES THE EXAMPLE AS ACCEPTABLE FOR MARITIME CV (I.E. NUMBER OF MATCHES WITH THE MARITIME CV ROW)

Example	1	2	3	4	5	6	7	8	9	10	Number of Successes
Maritime CV	TP	TP	TP	FP	TP	TP	FP	FP	FP	FP	Not applicable
IOU (0.5)	TP	FP	FP	TP	FP	FP	FP	TP	FP	TP	3
Dice (0.5)	TP	FP	FP	TP	TP	FP	TP	TP	TP	TP	2
IOG (0.5)	TP	FP	FP	FP	TP	FP	TP	TP	FP	TP	4
BEP_1 (0.7,0.75)	TP	TP	TP	FP	FP	FP	FP	TP	TP	FP	6
BEP_2 (0.7,0.75)	TP	TP	TP	FP	TP	TP	FP	FP	FP	FP	10
X_1 (0.7)	TP	TP	TP	TP	FP	FP	TP	TP	TP	TP	3
X_2 (0.7)	TP	TP	TP	TP	TP	TP	TP	TP	TP	TP	5
Y_1 (0.75)	TP	TP	TP	FP	TP	TP	FP	TP	TP	FP	8
Y_2 (0.75)	TP	TP	TP	FP	TP	TP	FP	FP	FP	FP	10

TABLE II

LIST OF BACKGROUND SUBTRACTION METHODS IS PRESENTED HERE. THE METHODS ARE GROUPED ACCORDING TO THE CENTRAL CONCEPT BEHIND THEM. THE BEST RESULTS OF EACH GROUP APPEAR IN TABLE III. THE NUMBER OF METHODS IN EACH GROUP IS INDICATED IN {}

Group	**Methods in the group**
Spatio-temporal filters (STF) - {4}	Temporal mean (TM) [12], Prati's median (PM) [13], adaptive median (AM) [14], $\sigma - \Delta$ BGS [15]
Gaussian models (GM) - {8}	Simple Gaussian (SG) [16], Gaussian average (GA) [17], Grimson's Gaussian mixture model (GMM) [18], Zivkovic's adaptive GMM (AGMM) [19], mixture of Gaussians (MoG) [20], fuzzy Gaussian (FG) [21], type-2 fuzzy GMM - uncertain mean (T2FUM) [22], type-2 fuzzy GMM - uncertain variance (T2FUV) [22]
Kernel models (KM) - {2}	Kernel density estimation (KDE) [23], VuMeter [24]
Self organizing maps (SOM) - {2}	Adaptive self organizing maps (ASOM) [25], fuzzy ASOM (FASOM) [25]
Low rank and sparsity (LRS) - {15}	Eigen-background (EB) [26], active subspace (AS) robust principal component analysis (RPCA) [27], fast (F) principal component pursuit (PCP) [28], Reimanian robust (R2) PCP [29], MoG-RPCA [30], non-convex (NC) RPCA [31], Grassman average [32], greedy semi-soft go decomposition (GreGoDec) [33], orthogonal rank-one matrix pursuit (OR1MP) [34], Grassmannian rank-one update subspace estimation (GROUSE) [35], low-rank matrix completion by Riemannian optimization (LRGeomCG) [36], non-negative matrix factorization (NMF) with sparse matrix (LS2) [37], Deep semi NMF (DSNMF) [38], alternating direction method of multipliers (ADMM) [39], robust orthonormal subspace learning (ROSL) [40]
Texture, color, and regions (TCR) - {5}	Texture BGS (TBGS) [41], independent multimodal background subtraction (IMBS) [42], multicue [43], local binary similarity segmenter (LOBSTER) [44], self-balanced sensitivity segmenter (SuBSENSE) [45]

Consider fast moving objects in Fig. 6: E in frame 1, A in frame 2, and D in frame 4. Most methods generate phantom foreground for these objects, exceptions include Prati's median, SuBSENSE, and IMBS. Such phantoms may result into one wider detection or multiple individual detections, see KDE results for object A in frame 2 and object E in frame 1 for respective examples. These examples indicate a challenge not recognized in [3]. Dynamic BGS should incorporate large variations in the speeds of the vessels (both in the physical scales and the image scales) for avoiding phantom detections of fast vessels.

Wakes result in wider BBs in most methods for object D in frame 4. The detected spans of the fast moving objects and the objects with wakes are larger than the actual objects. For a fast moving object, information of minimum span and bottom edge is critically important for collision avoidance. It is acceptable, although not preferable, to interpret a larger span than the actual span. Thus, despite wider BBs, these detections are useful for collision avoidance. The BB of SuBSENSE corresponding to object A in frame 2 is comparatively less acceptable, since it underestimates the span of the vessel. IOU (0.5) estimates it as true positive, even though this detection indicates deficiency of SuBSENSE for collision avoidance. Also, note that fuzzy Gaussian BGS generates one significantly larger BB for each example frame, with the bottom edge of BB much below a GT's bottom edge. IOG detects it as a true positive, even though such detections are clearly deficient for collision avoidance.

Now, consider object A in frame 1 and objects B-D in frame 3. For these objects, several methods detect either the super-structure or the hull. Or, they break down the object into several smaller detections (note object A of frame 1). While the detected hulls indicate acceptable performance for collision avoidance, the detected super-structures or portions of the objects are unacceptable. BEPs are effective in assessing both these conditions appropriately.

Frame 2 presents an example of several occluded objects with small pixel foot prints. Different methods give varied results, several of them being useful for an initial estimate. This indicates potential for CV methods. However, suppression of false positive detections in water background is important for reasonable conclusion. At the same time, situations such as example 9 from Fig. 1(b) also occur in numerous places. See for example, the results of eigen-background and KDE for the example frame 2. Even with the BEP metrics, assessing them appropriately for collision avoidance in maritime CV is an open problem.

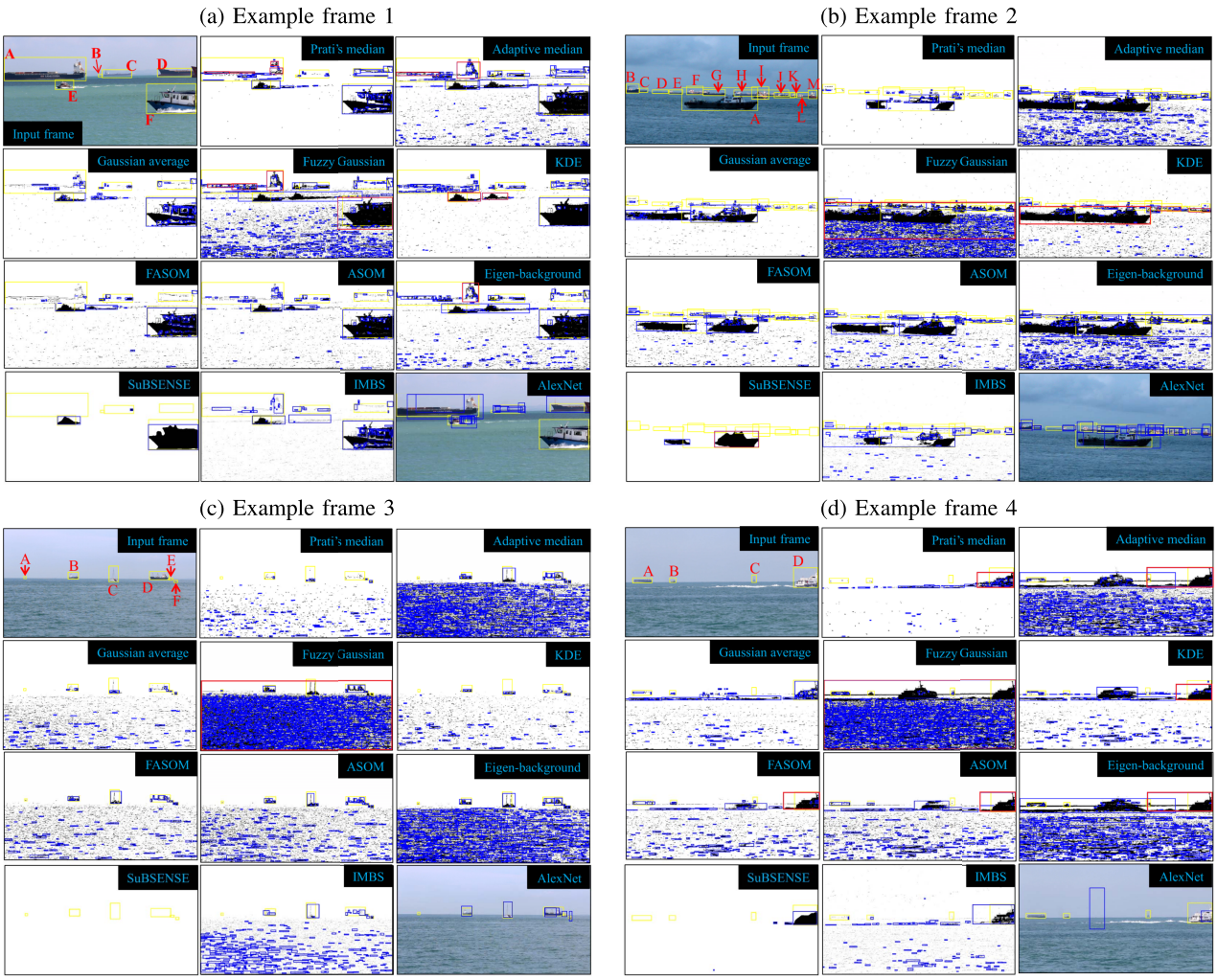


Fig. 6. Example results of CV methods for detection through dynamic BGS. The subtracted background appears white in the results of the methods. Ground truths: yellow BBs. Detected objects (foreground segmentations obtained after BGS): blue or red BBs. Red BBs: DOs referred in the text.

The results of R-CNN for the four example frames indicate that detections using R-CNN are better and less affected by wake. Moreover, DOs typically span both the hull and the super-structure. We note that the current implementation detects the same objects that it has been trained for, which is the reason for better quality of DOs. This approach is suitable only where environment specific training is feasible and practically useful.

E. Quantitative Results

We assess the true positive (TP) detections in all the frames of the all the videos in the dataset. The precision for the entire dataset is computed as the ratio of the total number of TPs to the total number of DOs. The recall is computed as the ratio of the total number of TPs to the total number of GTs. The assessment of TPs is performed using different assessment metrics and different threshold values for all of them. For IOU, Dice index, and IOG, we consider values 0.5, 0.7, and 0.9 for the threshold c_0. We note that IOU (0.5) is recommended in the well-known Pascal challenge [11]. The threshold x_0 for BEP_1 and BEP_2 is 1-dimensional analogue of c_0 for IOU and IOG, respectively. Thus, we use three values $\sqrt{0.5}$, $\sqrt{0.7}$, and $\sqrt{0.9}$ for x_0. We use three values 0.6, 0.75, and 0.9 for the threshold y_0. We include the results in which TPs are assessed using the Y metrics alone. The precision and recall values of the six BGS groups identified in Table II and the R-CNN are given in Table III. The precision and recall values are color coded for easy visual interpretation.

TCR methods are more effective at background subtraction than the other methods (see results of SuBSENSE in Fig. 6). So, false positive detections due to water background are very few, leading to better precision than other methods. Also, precision values of SuBSENSE for BEP_2 metric are not poor considering that it was not developed specifically for the maritime domain. On the other hand, IMBS does not provide the best precision or recall even though it was developed specifically for the maritime domain. A reason could be that IMBS was developed for high mounted cameras in urban maritime, a setting different from the current dataset. The precision and recall results for R-CNN are expectedly better than the other approaches. However, noting that the R-CNN here detects the objects it has been trained for, the precision

TABLE III

PRECISION AND RECALL OF CV METHODS FOR THE MARITIME DATASET. BEST RESULTS FOR EACH GROUP IDENTIFIED IN TABLE II ARE PRESENTED HERE. IN EACH GROUP, THE METHODS THAT CONSISTENTLY GIVE THE BEST PRECISION OR RECALL FOR MOST ASSESSMENT CRITERIA ARE INDICATED IN THE BOTTOM ROW

Legend									
Precision					Recall				
≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.4	≤ 0.5

	parameters		Precision							Recall						
	c_0 or x_0	y_0	STF	GM	KM	SOM	LRS	TCR	CNN	STF	GM	KM	SOM	LRS	TCR	CNN
IOU (c_0)	0.5	–	0.01	0.00	0.01	0.01	0.00	0.15	0.28	0.14	0.11	0.10	0.10	0.14	0.07	0.41
	0.7	–	0.00	0.00	0.00	0.00	0.00	0.05	0.12	0.05	0.04	0.04	0.03	0.05	0.02	0.18
	0.9	–	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00
Dice (c_0)	0.5	–	0.01	0.01	0.01	0.01	0.00	0.25	0.35	0.26	0.20	0.19	0.18	0.25	0.11	0.51
	0.7	–	0.00	0.00	0.00	0.01	0.00	0.14	0.25	0.12	0.09	0.08	0.08	0.11	0.07	0.37
	0.9	–	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.02	0.02	0.01	0.01	0.02	0.01	0.04
IOG (c_0)	0.5	–	0.01	0.01	0.01	0.01	0.07	0.43	0.40	0.32	0.30	0.20	0.19	0.32	0.19	0.58
	0.7	–	0.01	0.01	0.01	0.01	0.07	0.40	0.32	0.24	0.26	0.14	0.13	0.25	0.17	0.47
	0.9	–	0.00	0.00	0.00	0.00	0.07	0.36	0.17	0.15	0.19	0.09	0.07	0.17	0.16	0.24
BEP_1 (x_0, y_0)	$\sqrt{0.5}$	0.6	0.01	0.01	0.01	0.01	0.00	0.18	0.26	0.15	0.12	0.13	0.10	0.15	0.06	0.38
	$\sqrt{0.7}$	0.6	0.00	0.01	0.01	0.01	0.00	0.17	0.24	0.13	0.10	0.12	0.08	0.12	0.06	0.35
	$\sqrt{0.9}$	0.6	0.00	0.00	0.00	0.00	0.00	0.13	0.16	0.08	0.07	0.08	0.06	0.08	0.04	0.23
	$\sqrt{0.5}$	0.75	0.00	0.00	0.00	0.00	0.00	0.12	0.15	0.09	0.07	0.08	0.05	0.09	0.04	0.21
	$\sqrt{0.7}$	0.75	0.00	0.00	0.00	0.00	0.00	0.11	0.14	0.08	0.06	0.07	0.05	0.07	0.04	0.20
	$\sqrt{0.9}$	0.75	0.00	0.00	0.00	0.00	0.00	0.10	0.09	0.05	0.04	0.05	0.04	0.05	0.03	0.13
	$\sqrt{0.5}$	0.9	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.03	0.03	0.03	0.02	0.03	0.01	0.03
	$\sqrt{0.7}$	0.9	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.03	0.03	0.03	0.02	0.03	0.01	0.03
	$\sqrt{0.9}$	0.9	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.02	0.02	0.03	0.02	0.02	0.01	0.02
BEP_2 (x_0, y_0)	$\sqrt{0.5}$	0.6	0.01	0.01	0.02	0.01	0.00	0.21	0.33	0.38	0.31	0.27	0.23	0.38	0.12	0.49
	$\sqrt{0.7}$	0.6	0.01	0.01	0.01	0.01	0.00	0.21	0.31	0.35	0.28	0.25	0.21	0.32	0.12	0.45
	$\sqrt{0.9}$	0.6	0.01	0.01	0.01	0.01	0.00	0.17	0.21	0.25	0.20	0.20	0.16	0.25	0.09	0.31
	$\sqrt{0.5}$	0.75	0.01	0.01	0.01	0.01	0.00	0.16	0.26	0.33	0.26	0.23	0.19	0.33	0.10	0.38
	$\sqrt{0.7}$	0.75	0.01	0.01	0.01	0.01	0.00	0.16	0.23	0.30	0.24	0.21	0.17	0.30	0.10	0.34
	$\sqrt{0.9}$	0.75	0.01	0.01	0.01	0.01	0.00	0.13	0.15	0.23	0.18	0.18	0.14	0.23	0.08	0.23
	$\sqrt{0.5}$	0.9	0.01	0.01	0.01	0.01	0.00	0.09	0.16	0.26	0.21	0.18	0.14	0.27	0.08	0.23
	$\sqrt{0.7}$	0.9	0.01	0.01	0.01	0.01	0.00	0.09	0.14	0.24	0.20	0.17	0.14	0.25	0.08	0.20
	$\sqrt{0.9}$	0.9	0.00	0.00	0.01	0.01	0.00	0.07	0.09	0.19	0.15	0.15	0.11	0.20	0.07	0.13
Y_1 (y_0)	–	0.6	0.12	0.24	0.05	0.05	0.01	0.59	0.58	0.88	0.92	0.78	0.70	0.86	0.45	0.85
	–	0.75	0.09	0.17	0.04	0.04	0.01	0.53	0.55	0.81	0.87	0.72	0.63	0.80	0.37	0.81
	–	0.9	0.05	0.07	0.03	0.03	0.01	0.39	0.45	0.62	0.70	0.56	0.46	0.62	0.26	0.65
Y_2 (y_0)	–	0.6	0.01	0.01	0.02	0.01	0.00	0.23	0.34	0.38	0.31	0.28	0.24	0.38	0.13	0.49
	–	0.75	0.01	0.01	0.01	0.01	0.00	0.17	0.24	0.30	0.24	0.22	0.18	0.30	0.10	0.35
	–	0.9	0.00	0.01	0.01	0.01	0.00	0.08	0.09	0.20	0.15	0.15	0.11	0.20	0.07	0.13
Consistently best			PM	GA	KDE	FASOM	EB	SuBSENSE	AlexNet	AM	FG	KDE	ASOM	EB	IMBS	AlexNet

and recall should have been better. These clearly demonstrate the challenging nature of maritime CV.

The several false positives in most BGS methods (see Fig. 6) result in poor precision. Most methods have recall better than precision, with the exception of TCR methods. We also note that BEP_2 values are more encouraging than IOU, Dice Index, IOG, and BEP_2. The better suitability of BEP_2 was established in Table I. Moreover, it is noted in Table III that Y_1 is less selective about TPs. This puts the responsibility on X_1 for improving the selectivity of BEP_1. On the other hand, Y_2 is inherently more selective, as demonstrated by lower precision and recall values than Y_1. This directly helps in making BEP_2 selective.

We compare assessment metrics IOU(0.5) and $BEP_1(\sqrt{0.5}, 0.6)$, which correspond to most lenient threshold values. Recall values for $BEP_1(\sqrt{0.5}, 0.6)$ are better than IOU(0.5) in each group. For the most strict threshold values as well, recall values for $BEP_1(\sqrt{0.9}, 0.9)$ are better than IOU(0.9) in each group. The same can be inferred from the comparison of IOG and BEP_2, barring a few exceptions. Thus, although the conventional metrics indicate dismal performance of CV methods for maritime, the scene does not look so bleak when metrics designed for maritime domain are used. This highlights the need of both suitable metrics and dedicated CV solutions.

VI. DISCUSSION

We evaluated the existing metrics for assessing the quality of BB detections in the context of maritime CV. The unique needs of maritime CV imply that the current metrics are unsuitable. The proposed bottom edge proximity metrics, custom designed for maritime CV, provide a good starting point. However, there is a need to explore more options for assessing detections in maritime CV. Such assessment metrics would be strict in assessing the location of the bottom edges and minimum span of the BBs, suitable for assessing inaccurate detections due to occlusion, and tolerant for BB degradation in presence of wake or exclusion of super-structure in the detected BB. It is worth considering if the conventional BB labeling of GT is suitable for maritime CV. It should be explored if the GT of each vessel should comprise of GTs for hull, super-structure, and their union. An associated problem is to design assessment of detected BBs for such GT. Creating shape and pixel segmentations as ground truth for large videos needs

to be explored. Detections and their assessment in the form of shape and pixel segmentations can be explored for new maritime CV methods.

Our preliminary study of 36 background subtraction methods and two R-CNN experiments shows a gap in CV techniques for maritime applications. Appropriate modeling of maritime background can reduce false positives and improve precision. Modeling wakes as background as well may allow stricter assessment of span (larger x_0) and thus better assessment of occlusions as well. Large range of speeds and sizes of maritime objects may require innovative approaches for learning background with adaptive time scales in local regions. Deep learning also holds significant promise. Our current experiments assume the luxury of environment specific training. A more generalizable deep learning framework for maritime is needed for practical maritime computer vision.

We note that the maritime computer vision is in a nascent stage at present. It is too early to decide on a suitable metric. A better convergence on these topics will emerge with further engagement of the CV community. The engagement can be through new diverse maritime datasets and maritime CV challenges similar to the PASCAL challenge [11] with goal towards autonomous maritime vehicle technology.

REFERENCES

[1] E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, "Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1559–1582, May 2018.

[2] D. Bloisi and L. Iocchi, "ARGOS—A video surveillance system for boat traffic monitoring in venice," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 7, pp. 1477–1502, 2009.

[3] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 1993–2016, Aug. 2017.

[4] P. J. Withagen, K. Schutte, A. M. Vossepoel, and M. G. J. Breuers, "Automatic classification of ships from infrared (FLIR) images," *Proc. SPIE*, vol. 3720, pp. 180–188, Jul. 1999.

[5] M. Levandowsky and D. Winter, "Distance between sets," *Nature*, vol. 234, Nov. 1971, Art. no. 5323.

[6] D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity," *Brit. Med. J.*, vol. 308, p. 1552, Jun. 1994.

[7] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[8] A. Cuzzocrea, E. Mumolo, and G. M. Grasso, "Advanced pattern recognition from complex environments: A classification-based approach," *Soft Comput.*, vol. 22, no. 14, pp. 4763–4778, 2018.

[9] A. Tversky, "Features of similarity," *Psychol. Rev.*, vol. 84, no. 4, pp. 327–352, 1977.

[10] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Object detection in a maritime environment: Performance evaluation of background subtraction methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1787–1802, May 2019.

[11] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2014.

[12] A. H. S. Lai and N. H. C. Yung, "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 4, May/Jun. 1998, pp. 241–244.

[13] S. Calderara, R. Melli, A. Prati, and R. Cucchiara, "Reliable background suppression for complex scenes," in *Proc. ACM Int. Workshop Video Surveill. Sensor Netw.*, 2006, pp. 211–214.

[14] N. J. B. McFarlane and C. P. Schofield, "Segmentation and tracking of piglets in images," *Brit. Mach. Vis. Appl.*, vol. 8, no. 3, pp. 187–193, 1995.

[15] A. Manzanera and J. C. Richefeu, "A new motion detection algorithm based on $\Sigma-\Delta$ background estimation," *Pattern Recognit. Lett.*, vol. 28, no. 3, pp. 320–328, 2007.

[16] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[17] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[18] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 1999, pp. 246–252.

[19] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. 17th Int. Conf. Pattern Recognit.*, vol. 2, 2004, pp. 28–31.

[20] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vis. Image Understand.*, vol. 122, pp. 4–21, May 2014.

[21] M. H. Sigari, N. Mozayani, and H. R. Pourreza, "Fuzzy running average and fuzzy background subtraction: Concepts and application," *Int. J. Comput. Sci. Netw. Secur.*, vol. 8, no. 2, pp. 138–143, Feb. 2008.

[22] Z. Zhao, T. Bouwmans, X. Zhang, and Y. Fang, "A fuzzy background modeling approach for motion detection in dynamic backgrounds," in *Proc. Int. Conf. Multimedia Signal Process.* Springer, 2012, pp. 177–185.

[23] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 751–767.

[24] Y. Goya, T. Chateau, L. Malaterre, and L. Trassoudaine, "Vehicle trajectories evaluation by static video sensors," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2006, pp. 864–869.

[25] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 179–186, 2010.

[26] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.

[27] G. Liu and S. Yan, "Active subspace: Toward scalable low-rank learning," *Neural Comput.*, vol. 24, no. 12, pp. 3371–3394, 2012.

[28] P. Rodríguez and B. Wohlberg, "Fast principal component pursuit via alternating minimization," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 69–73.

[29] M. Hintermüller and T. Wu, "Robust principal component pursuit via inexact alternating minimization on matrix manifolds," *J. Math. Imag. Vis.*, vol. 51, no. 3, pp. 361–377, 2015.

[30] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang, "Robust principal component analysis with complex noise," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 55–63.

[31] Z. Kang, C. Peng, and Q. Cheng, "Robust PCA via nonconvex rank approximation," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2015, pp. 211–220.

[32] S. Hauberg, A. Feragen, and M. J. Black, "Grassmann averages for scalable robust PCA," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3810–3817.

[33] T. Zhou and D. Tao, "Greedy bilateral sketch, completion & smoothing," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2013, pp. 650–658.

[34] Z. Wang, M.-J. Lai, Z. Lu, W. Fan, H. Davulcu, and J. Ye, "Orthogonal rank-one matrix pursuit for low rank matrix completion," *SIAM J. Sci. Comput.*, vol. 37, no. 1, pp. A488–A514, 2015.

[35] L. Balzano, R. Nowak, and B. Recht, "Online identification and tracking of subspaces from highly incomplete information," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, Sep./Oct. 2010, pp. 704–711.

[36] B. Vandereycken, "Low-rank matrix completion by Riemannian optimization," *SIAM J. Optim.*, vol. 23, pp. 1214–1236, Jun. 2013.

[37] Y. Ji and J. Eisenstein, "Discriminative improvements to distributional sentence similarity," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 891–896.

[38] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 417–429, Mar. 2017.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[40] X. Shu, F. Porikli, and N. Ahuja, "Robust orthonormal subspace learning: Efficient recovery of corrupted low-rank matrices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3874–3881.
[41] M. Heikkila and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, Apr. 2006.
[42] D. Bloisi and L. Iocchi, "Independent multimodal background subtraction," in *Proc. Int. Conf. Comput. Modeling Objects Presented Images, Fundam., Methods Appl.*, 2012, pp. 39–44.
[43] S. Noh and M. Jeon, "A new framework for background subtraction using multiple cues," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 493–506.
[44] P.-L. St-Charles and G.-A. Bilodeau, "Improving background subtraction using local binary similarity patterns," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2014, pp. 509–515.
[45] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Flexible background subtraction with self-balanced local sensitivity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 414–419.
[46] A. Sobral. (2013). *BGSLibrary: An OpenCV C++ Background Subtraction Library*. [Online]. Available: https://github.com/andrewssobral/bgslibrary
[47] A. Sobral, T. Bouwmans, and E.-H. Zahzah, "LRSLibrary: Low-rank and sparse tools for background modeling and subtraction in videos," in *Handbook of Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. Boca Raton, FL, USA: CRC Press, 2016.
[48] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[49] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[50] G. Varol, I. Laptev, and C. Schmid, "Long-term temporal convolutions for action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1510–1517, Jun. 2018.
[51] S. Ren, K. He, R. Girshick, X. Zhang, and J. Sun, "Object detection networks on convolutional feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1476–1481, Jul. 2017.
[52] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Comput. Sci. Rev.*, vol. 11, pp. 31–66, May 2014.
[53] T. Bouwmans, F. Porikli, B. Höferlin, and A. Vacavant, *Background Modeling and Foreground Detection for Video Surveillance*. Boca Raton, FL, USA: CRC Press, 2014.
[54] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. TR-2009, 2009, vol. 1, no. 4.
[55] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

Dilip K. Prasad received the B.Tech. degree in computer science and engineering from the Indian Institute of Technology (ISM), Dhanbad, India, in 2003, and the Ph.D. degree in computer science and engineering from Nanyang Technological University, Singapore, in 2013. He is currently an Associate Professor with UiT The Arctic University of Norway. His current research interests include image processing, machine learning, and computer vision.

Huixu Dong received the B.Sc. degree in mechatronics engineering from the Harbin Institute of Technology in China, in 2013, and the Ph.D. degree from the Robotics Research Centre, Nanyang Technological University, Singapore, in 2018. He is currently a Post-Doctoral Fellow with the Robotics Institute, Carnegie Mellon University. His current research interests include robotic perception and grasp in unstructured environments, computer vision and robot-oriented artificial intelligence, and the navigation of mobile robot and optimal design of robotic gripper.

Deepu Rajan received the Bachelor of Engineering degree in electronics and communication engineering from the Birla Institute of Technology, Ranchi, India, the M.S. degree in electrical engineering from Clemson University, Clemson, SC, USA, and the Ph.D. degree from the Indian Institute of Technology, Mumbai, India. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His current research interests include image processing, computer vision, and multimedia signal processing.

Chai Quek received the B.Sc. and Ph.D. degrees from Heriot-Watt University, Edinburgh, U.K. He is currently with the School of Computer Engineering, Nanyang Technological University, Singapore. He has published over 250 international conference and journal articles. His research interests include neurocognitive informatics, biomedical engineering, and computational finance. He has been invited as a Program Committee Member and a Reviewer for several conferences and journals, including IEEE TNN and TEvC.