Object Detection in a Maritime Environment: Performance Evaluation of Background Subtraction Methods

Dilip K. Prasad, Chandrashekar Krishna Prasath, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek

***Abstract*— This paper provides a benchmark of the performance of 23 classical and state-of-the-art background subtraction (BS) algorithms on visible range and near infrared range videos in the Singapore Maritime dataset. Importantly, our study indicates the limitations of the conventional performance evaluation criteria for maritime vision and proposes new performance evaluation criteria that is better suited to this problem. This paper provides insight into the specific challenges of BS in maritime vision. We identify four open challenges that plague BS methods in maritime scenario. These include spurious dynamics of water, wakes, ghost effect, and multiple detections. Poor recall and extremely poor precision of all the 23 methods, which have been otherwise successful for other challenging BS situations, allude to the need for new BS methods custom designed for maritime vision.**

***Index Terms*— Maritime vehicles, autonomous automobiles, background subtraction, object detection, computer vision.**

I. Introduction

IT IS envisaged that autonomous ships will be equipped with numerous sensors, of which visible and infrared cameras would play an important role in collision avoidance [1], [2]. Object detection in a video stream captured from cameras is critical in this respect. The first step in a computer vision system that detects objects in videos is background subtraction (BS). The background information of the scene is estimated and subtracted from the original video frame, which results in the detection of foreground objects. The problem becomes challenging when the background is dynamic as opposed to being static.

It is clear that we are dealing with dynamic background in the maritime environment due to the presence of waves, wakes etc. However, as we show, even state-of-the-art BS algorithms that address dynamic backgrounds are unable to handle the dynamics present in the maritime situation. This is because conventional dynamic BS algorithms treat regions of high spatio-temporal correlation as foreground objects and waves present high spatio-temporal correlations [3]. See [1] and [4] for more extensive discussion on challenges of maritime vision problem, including BS.

In Table I, we compare the performance of 22 algorithms from Sobral's background subtraction library [5] on four datasets - the background models challenge (BMC) dataset [6], the change detection (CD) dataset [7] and the Singapore Maritime (SM) dataset [1] containing both visible (SMV) and infrared (SMIR) videos. The open source availability of the source codes and the previous benchmarking results [5], [7] were the reasons for the selection of these methods. The details of computing the precision and recall are discussed in sections IV and section V. We mention here that the true positive detections used for computing the precision and recall in Table I are determined using intersection over union (IOU) ratio being more than 0.5. The best performing methods for the BMC dataset with mean background (best precision, 0.935) and fuzzy Gaussian (best recall, 0.909) do not perform well for the SMV and SMIR dataset (precision < 0.133 and recall < 0.139). SuBSENSE, which performs the best for the CD dataset (precision 0.751 and recall 0.812), provides a precision of only 0.133 (for SMV) and recall < 0.139. **Interestingly, the *best* performances on SMV and SMIR datasets across all the methods are much poorer than even the *worst* performances on the BMC and CD datasets. Thus, we see that there is a pressing need to revisit the BS algorithms specifically for the maritime environment. As a first step, this paper provides a benchmark of current BS algorithms.**

In this study, we note that the challenges of applying BS in maritime environment begin with the lack of suitable metrics for evaluating the performance of object detection in maritime environment. We show that the conventional object detection metric in computer vision, IOU, is not suitable for several object detection possibilities in maritime computer vision. So, we present new metrics that may be better suited for maritime problem. Moreover, the study not only benchmarks several BS methods quantitatively, but also performs qualitative investigation of the reasons of poor performance. This leads to identification of specific BS challenges that plague object detection in maritime problem. Through numerous experiments and new performance metrics, we have clearly outlined methods that are ineffective and those that hold promise and need to be further investigated for adaption to the maritime domain. This

Manuscript received March 30, 2017; revised January 8, 2018; accepted April 30, 2018. Date of publication July 2, 2018; date of current version May 1, 2019. This work was supported by the National Research Foundation under the CorpLab@University Scheme. The Associate Editor for this paper was Q. Ji. *(Corresponding author: Dilip K. Prasad.)*

D. K. Prasad and C. K. Prasath are with the Rolls-Royce@ NTU Corporate Lab, Nanyang Technological University, Singapore 639798 (e-mail: dilipprasad@gmail.com).

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

L. Rachmawati is with Rolls Royce Pvt. Ltd., Singapore 797575.

E. Rajabally is with Rolls Royce plc, Derby DE24 7XX, U.K.

Digital Object Identifier 10.1109/TITS.2018.2836399

TABLE I
THE PERFORMANCE FOR BMC [5], CD [7], SMV AND SMIR DATASETS. THE DATA CORRESPONDS TO 22 METHODS IN SECTION III

	Best performance				Worst performance			
Dataset	Method	Precision	Method	Recall	Method	Precision	Method	Recall
BMC [5]	Mean background [8]	0.935	Fuzzy Gaussian [9]	0.909	Eigen-background [10]	0.658	Mean background [8]	0.597
CD [7]	SuBSENSE [11]	0.751	SuBSENSE [11]	0.812	KDE [12]	0.581	Zivkovic's GMM [13]	0.660
SMV	SuBSENSE [11]	0.133	Adaptive Median [14]	0.139	T2Fuzzy GMM-UM [15]	$<$0.001	T2Fuzzy GMM-UM [15]	$<$0.001
SMIR	Multicue	0.073	Texture	0.081	Mean background [8]	0.000	Mean background [8]	0.000

TABLE II
DETAILS OF THE SINGAPORE-MARITIME DATASET USED IN THIS PAPER

Datasets	Visible range	NIR range
Number of videos	33	22
Number of frames	16584	10846
Minimum number of objects in a frame	0	0
Maximum number of objects in a frame	20	21
Total number of objects in all the videos	192980	80543

exercise provides specific and useful pointers for the attention of researchers in this exciting and challenging problem with significant impact on the future of maritime technology.

The dataset and methods used in this study are described in sections II and III, respectively. Our methodology is presented in section IV. The metrics for analyzing the performance of object detection through BS are presented in section V. Results, both quantitative and qualitative, are presented in section VI. A discussion on the study is presented in VII. The study is concluded in section VIII.

II. DATASET

Singapore Maritime dataset [1] is composed of videos acquired in high definition format (1080 $\times$ 1920 pixels) using Canon 70D cameras around Singapore waters. The dataset has on-shore (camera fixed on a platform) and on-board (camera on-board a moving vessel) videos. The ground truths (GT) for the objects were annotated by volunteers unrelated to the project, using annotation tools developed in Matlab. The objects were enclosed in bounding boxes. The dataset and annotation files of the GTs are available on the project webpage https://sites.google.com/site/dilipprasad/home/singapore-maritime-dataset. We use only on-shore videos in this paper to focus on inherent problems of BS in maritime scenario that are unrelated to the motion of the ship on which the camera is mounted. The on-shore videos are further categorized into visible range and near infrared (NIR) range videos. The NIR videos are acquired using a modified Canon 70D camera with its hot filter removed and a visible range blocking filter added. The details of the dataset are given in Table II.

III. METHODS USED FOR BENCHMARKING

As mentioned earlier, we use 22 BS methods for whose source codes are available at https://github.com/andrewssobral/bgslibrary. We also include a recent method, namely generalized fused lasso foreground modeling [16], whose matlab source code is available. We performed parameter optimization to achieve reasonable performance since the default parameters were completely unsuitable for the maritime problem. We categorize the methods used in this study according to the survey in [5] and briefly review each method in this section. These categories are: basic methods that employ basic statistical techniques for modeling dynamics of background; methods that model background using Gaussian distributions, this is a large and expanding family of methods; methods that model background using complex statistical distributions other than Gaussian, also an expanding family; methods that use texture and color descriptors, comprising of family of the popular local binary and ternary patterns; and other machine learning based methods which employ mathematical models of background that can be solved as some form of optimization problem incorporating priors. Existing maritime background subtraction methods often ignore the temporal information and estimate background for each frame independently [1]. In a few cases, Gaussian models have been used with limited success, see [1]. We also note independent multimodal background subtraction method (IMBS) [17], which has been developed for maritime problem specifically. Default parameters have been used for all the methods being benchmarked, unless stated otherwise. For further details of the methods, please see [5], [16], and references therein.

We note that the BS methods considered in this paper are not exhaustive and it would be ideal to include all BS approaches. Interested readers are referred to review articles [5], [18]–[20], [21]–[24], references therein, and other recent articles [25]–[45]. These methods have not been tested for various reasons, such as unavailability of source code, insufficient information for faithful reproduction of algorithm, unsuitability to maritime environment, and small increment over the methods included in this study. In this sense, we deem that benchmarking on these 23 methods is sufficient to obtain insights about the effectiveness of BS methods in maritime domain and opens the field for richer evaluation by other researchers active in development of BS approaches in computer vision.

A. Basic BS Methods

1) Mean Background [8]: This method computes the background in the current frame as the mean image of the last T frames. Binary foreground mask is computed using a simple threshold of the difference between the current frame and the background. Due to the default small value of T, it can adapt fast but can only detect those objects which are subject to significant movement in these frames.

2) Adaptive Median Background [14]: This method initializes the background as the median of first T_0 frames. Then, after every T frames, the background is updated in selected regions such that the updated intensities in these regions is

closer to the corresponding intensities in the current frame. The background regions to update are determined adaptively in each update interval. This method is suitable for slowly varying background and relatively fast moving objects.

3) Prati's Median Background [46]: This method is different from [14] in three major ways. First, it updates its background model using a temporal median filter over the last T frames and the last estimate of the background. Second, it computes the thresholds for determining background and foreground adaptively. And third, it uses a fast bootstrapping method for background initialization which involves partitioning the image region into blocks, adaptively adjusting blocks that demonstrate high difference in consecutive frames and quickly converging to stable blocks such that a stable initialization of background is obtained. Through these features, it is expected to perform better in vacillating background such as encountered in maritime videos.

B. Methods That Use Gaussian Background Models

1) Grimson's Gaussian Mixture Model (GMM) [47]: Grimson's GMM models the temporal distribution of the intensities (or color values) at a pixel as a mixture of Gaussian distributions, which are adaptively updated using online K-means approximation approach. Then the spatial distribution of the parameters and weights of the fitted GMMs are used to adaptively estimate the background as the GMMs update after each frame. Large deviations from the current background model indicate foreground pixels. This approach allows for a variety of backgrounds with large standard deviations (such as rustle of leaves or water ripples) and long term temporal changes to be effectively modeled.

2) Wren's Gaussian Average (GA) [48]: Conceptually, Wren's GA is similar to Grimson's GMM, but it fits only one Gaussian distribution per pixel. The mean values of the Gaussian distributions at the pixels are used to model the background. Online learning of background is made possible through the update of the Gaussian distributions for all the pixels in each frame. Such method accommodates long term slow temporal changes in the background but allows only one form of background distribution per pixel. Thus, in comparison to Grimson's GMM, it is expected to be less effective in dealing with wakes as background.

3) Simple Gaussian [49]: This implementation is conceptually similar to Wren's GA [48], but it utilizes likelihood maximization for updating the parameters of the Gaussian distribution at each pixel.

4) Zivkovic's Adaptive GMM (AGMM) [13]: Zivkovic's AGMM models both background and foreground as GMMs. Thus, although the GMMs can be fit for each pixel, Gaussian distributions with small weights would indicate the foreground. The need to include a new Gaussian distribution in the GMM is also an indicator of the foreground. In addition to the online learning of background through this concept, Zivkovic's AGMM incorporates forgetting of the old background model through the use of suitable constant weight for updating the parameters and weights of the GMMs.

5) Mixture of Gaussian (MoG) [5]: This is a hybrid of Grimson's GMM [47] and Zivkovic's [13], where the GMM model and update is derived from [47] but the number of Gaussian distributions used to represent the background is determined through [13].

6) Fuzzy Gaussian [9]: This method models background as a Gaussian low pass filter and adaptively updates the background using a fuzzy running average scheme. The decision on foreground pixels is also taken using fuzzy logic.

7) T2 Fuzzy GMM (T2FGMM) [15]: It models the background as GMM with either uncertain means (UM) or uncertain variances (UV) specified by fuzzy logic. The background is made adaptive by updates of the fuzzy functions for mean or variance.

C. Methods That Use Other Statistical Background Models

1) Kernel Density Estimation (KDE) [12]: KDE models background as composed of non-parametric kernels, thus allowing non-Gaussian distributions of the background to be modeled. The adaptation of the background occurs as online tweak of the probability density function. The intensities with larger probabilities correspond to background. In the current implementation, Gaussian kernel of zero mean is used.

2) VuMeter [50]: VuMeter is a KDE based non-parametric background model. However, instead of using Gaussian kernels [12], it uses Kronecker delta kernels. A simple temporal update of the model is used and foreground detection is further cast as likelihood maximization of Gaussian particle filters representing the foreground segmentations.

3) Independent Multimodal Background Subtraction (IMBS) [17]: IMBS has three components. The first component is an on-line clustering algorithm. The RGB values observed at a pixel is represented by histograms of variable bin size. This allows for modeling of non-Gaussian and irregular intensity patterns. The second component is a region-level understanding of the background for updating the background model. The regions with persistent foreground for a certain number of frames are included as background in the updated background model. The third component is a noise removal module that helps to filter out false detections due to shadows, reflections, and boat wakes. It models wakes as outliers of the foreground, forming a second background model specifically for such outliers. It is the only method in this list of methods which has been developed for maritime computer vision problem specifically.

D. Methods That Use Texture and Color Descriptors

1) Texture as Background [51]: Background is modeled as composed of textures described using histogram vectors (HVs) of local binary patterns (LBPs). At each pixel, the weights of LBPs are computed. The weight vector is compared against the stored HVs of stored background. Small intersection implies foreground pixel. Large intersection prompts the update of HVs of LBPs of background.

2) Multicue [52]: It uses multiple cues such as texture, color, and region appearance. Texture model with scene adaptive LBP descriptors is used to detect the initial foreground regions. Color value statistics are integrated with

texture descriptors to refine the results of texture based foreground segmentations. Lastly, Hausdroff distance between the edgemap and the segmented foreground regions is used to refine the foreground. As already seen in Fig. 3, Multicue may not perform well for maritime problem if the texture is not appropriately modeled.

3) Local Binary Similarity Segmenter (LOBSTER) [53]: LOBSTER uses local binary patters with self-similarity throughout the images. It is adapted from ViBe [54] and replaces ViBe's pixel-intensity level description of the background with spatiotemporal similarity of local binary patterns. The model is updated using simple rules based on distance between the similarity patterns.

4) Self-Balanced Sensitivity Segmenter (SuBSENSE) [11]: SuBSENSE employs pixel level change detection using comparisons of colors and local binary similarity patterns. It uses an adaptive feedback mechanism for updating the background features as the background dynamics change with time.

E. Other Machine Learning Based Methods

1) Eigen-Background [10]: In this technique, M eigenimages of last N frames are incrementally computed with each frame. These eigenimages encode static and large scale dynamic features, including long term temporal changes such as in illumination conditions. Thus, they are suitable for small foreground objects and are expected to deal with wakes which persist for a long duration.

2) Adaptive Self Organizing Maps (SOM) [55]: In order to mimic human neural systems, it uses hue-saturation-value color space and trains a neural networks-based self-organizing map in which the background at each pixels is represented by weights of multiple neurons in an intermediate layer. It is shown in [55] to be amenable to moving backgrounds, variations in gradual illumination and camouflage, and shadows.

3) Fuzzy Adaptive SOM (ASOM) [56]: It has two major enhancement over adaptive SOM. First, it introduces a spatial coherence variant to its neural network. Second, it uses fuzzy logic to make soft decisions about classifying pixels as background or foreground. Consequently, this method has lower false positives and better robustness in comparison to Adaptive SOM.

4) Sigma Delta [57]: Sigma-Delta background employs a recursive non-linear operator called $\Sigma - \Delta$ filter. This filter emulates a simple counter that adds one to the output if the current signal is more than previous signal and deducts one from the output if the current signal is less than the previous signal [58]. The method is made robust using a spatio-temporal regularizer. Multiple filters with different time constants are used to accommodate complex background dynamics.

5) Generalized Fused Lasso Foreground Modeling [16]: This method is similar to eigen-background in representing the background as low-ranked sub-space. Foreground estimation is made robust through generalized fused lasso regularization and incorporation of penalty for total variation. The Matlab code of the method, provided on the project website of [16] is used for generating the results. Computation time of GFLFM is prohibitively long. Thus, we have provided results for visible range videos only.

IV. Methodology

All videos are down-sampled in spatial dimensions by a factor of 2 for methods in BGS library and by a factor of 4 for GFLFM. Each BS method generates a binary image in which background is assigned 0 and the foreground region is assigned 1. All the results are generated on an Intel Xeon CPU with E5-1650@3.2 GHz and 64 GB RAM system. The holes in the foreground binary mask are filled and closed regions are detected. The bounding boxes of closed regions are determined. Bounding boxes of dimensions smaller than 10 pixels are considered as spurious detections due to noise and water dynamics and are removed from further processing. Each of the remaining bounding boxes is considered as a detected object.

For each pair of detected object (DO) and ground truth object (referred to as GT), we compute a performance metric C (see section V for the metrics considered). We note that it is preferable to use the shape segmentation of the GT and DO. However, obtaining shape segmentations manually for all the frames of maritime videos is tedious and resource demanding, and ultimately a heuristic exercise. Thus, we use the bounding boxes of the GT and DO objects in maritime videos for performance evaluation. The metrics have been defined such that if the DO-GT pair does not qualify certain thresholds, the value of the metrics is assigned to be $-\infty$. Then, we use the Hungarian algorithm [59] to find one-to-one correspondences (i.e. only one DO should be considered as representing a GT) between the detected objects and ground truth objects. We use $1-C$ as the input argument for the Hungarian method. Indeed, the pairs of DO-GT with the value of metric equals to infinity do not result in any correspondences due to the value $-\infty$ being assigned to the metrics when the thresholds are not met. The other possible correspondences are ranked to minimize the value of $1 - C$ after all correspondences are complete. The total-number of correspondences is used as the number of true positives (TP). The number of detected objects for which correspondences could not be determined corresponds to false positives (FP). Similarly, the number of ground truth objects for which correspondences could be found is used as false negatives (FN).

Then precision, recall, F-score are determined as follows:

$$\text{Precision} = \frac{\sum_{t=1}^{T} \text{TP}_t}{\sum_{t=1}^{T} (\text{TP}_t + \text{FP}_t)}; \quad \text{Recall} = \frac{\sum_{\text{t}=1}^{\text{T}} \text{TP}_\text{t}}{\sum_{\text{t}=1}^{\text{T}} (\text{TP}_\text{t} + \text{FN}_\text{t})} \tag{1}$$

$$\text{F-score} = \frac{2 \text{ Precision Recall}}{\text{Precision} + \text{Recall}} \tag{2}$$

where t denotes the frame and T is the total number of frames.

V. Metrics for Performance Evaluation

In computer vision, object detection is usually evaluated by the metric known as Intersection over Union (IOU), which is defined as

$$\text{IOU} = \frac{\text{Area}(\text{GT} \cap \text{DO})}{\text{Area}(\text{GT} \cup \text{DO})} \tag{3}$$

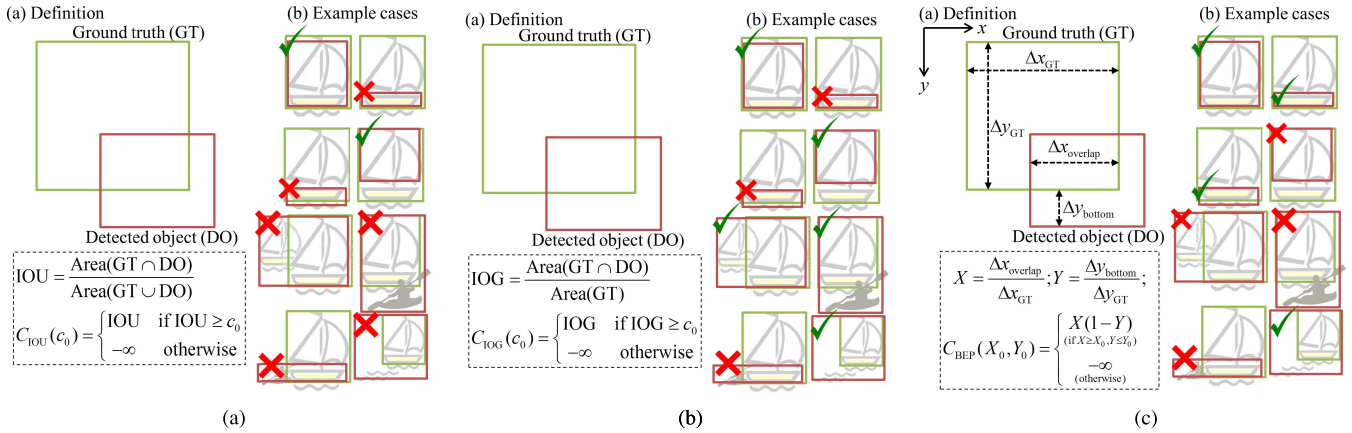

Fig. 1. The ticks indicate true detections while the crosses indicate false detections. (a) Metric C_{IOU} is illustrated here. (b) Metric C_{IOG} is illustrated here. (c) Metric C_{BEP} is illustrated here.

In order to discard poor DO-GT pairs before finding one-to-one DO-GT correspondences, we use a threshold c_0 and modify the IOU metric as follows:

$$C_{IOU}(c_0) = \begin{cases} IOU & \text{if } IOU \geq c_0 \\ -\infty & \text{otherwise} \end{cases} \tag{4}$$

Illustration of the metric and examples are given in Fig. 1(a).

The presence of wakes behind the vessels often result in detected objects wider than the GT. It is important in maritime vision to detect objects for collision avoidance. Consequently, even though a detected object is wider than the GT due to the presence of wake, it is preferable to consider it as a true positive. However, IOU may reject such object as false positive. Thus, we introduce a slight modification to the IOU metric by changing the denominator to comprise the area of the GT only. We call it Intersection over Ground truth (IOG) and define it as follows:

$$IOG = \frac{Area(GT \cap DO)}{Area(GT)} \tag{5}$$

and

$$C_{IOG}(c_0) = \begin{cases} IOG & \text{if } IOG \geq c_0 \\ -\infty & \text{otherwise} \end{cases} \tag{6}$$

The illustration of the metric and examples are shown in Fig. 1(b). In comparison with Fig. 1(a), it shows that such metric may allow true positives corresponding to objects with wake but can be prone to false positives in the case of occlusions.

Maritime objects may be characterized by a solid dense hull having larger possibility of detection and sparse mast region, which may not be amenable to detection. In order to handle this case, we propose a new metric called Bottom Edge Proximity (BEP) to test if the bottom edges of DO and GT are close. We define it as follows:

$$C_{BEP}(X_0, Y_0) = \begin{cases} X(1-Y) & \text{if } X \geq X_0, Y \leq Y_0 \\ -\infty & \text{otherwise} \end{cases} \tag{7}$$

where,

$$X = \frac{\Delta x_{overlap}}{\Delta x_{GT}}; \quad Y = \frac{\Delta y_{bottom}}{\Delta y_{GT}} \tag{8}$$

$\Delta x_{overlap}$ is the overlap of the DO and GT in the horizontal direction, Δx_{GT} is the width of the GT, Δy_{bottom} is the distance between the bottom edges of the DO and GT, and Δy_{GT} is the height of the GT, as illustrated in Fig. 1(c). This metric incorporates the salient property of C_{IOG} along the horizontal direction, allowing wider detections due to wakes. At the same time, it assess true positives for the detections that focus on hull and cannot detect the mast region.

We select three methods, LOBSTER [53], Multicue [52], and SubSENSE [11], for comparing the metrics and their relative merits and demerits in practical maritime scenario. These methods are chosen for their superior performance in comparison to others (as shall be evident in section VI).

The precision-vs-recall map of these methods are given in Fig. 2. For different metrics and their thresholds, the top row plots median precision and median recall of these methods on the visible range videos, while the 90th percentile values of precision and recall are plotted in the bottom row. We find that $C_{IOU}(0.5)$, the IOU metric with threshold of 0.5 conventionally used for object detection, generates quite poor precision and recall. This is because the IOU metric is quite stringent for the maritime problem due to the possibility of detection of hulls only or wider detections due to wakes. Thus, in our evaluations, we use relaxed threshold of 0.3 for IOU.

In comparison, the IOG and BEP metrics provide better values of recall for the same value of c_0 and X_0. Further, BEP provides better value of precision and recall in general, as seen here for LOBSTER and SubSENSE. We note that IOU penalizes false positive portion of the DO. Through this, large IOU supports better performance in comparison to IOG. For example, a DO that covers the entire frame has a perfect IOG. On the other hand, BEP does not suffer from this shortcoming of IOG and retains the property of penalizing the false positive regions in the DO.

In addition to the quantitative analysis, we also provide qualitative results in Figs. 3, 4, and 5, which highlight the strengths and drawbacks of these metrics more explicitly.

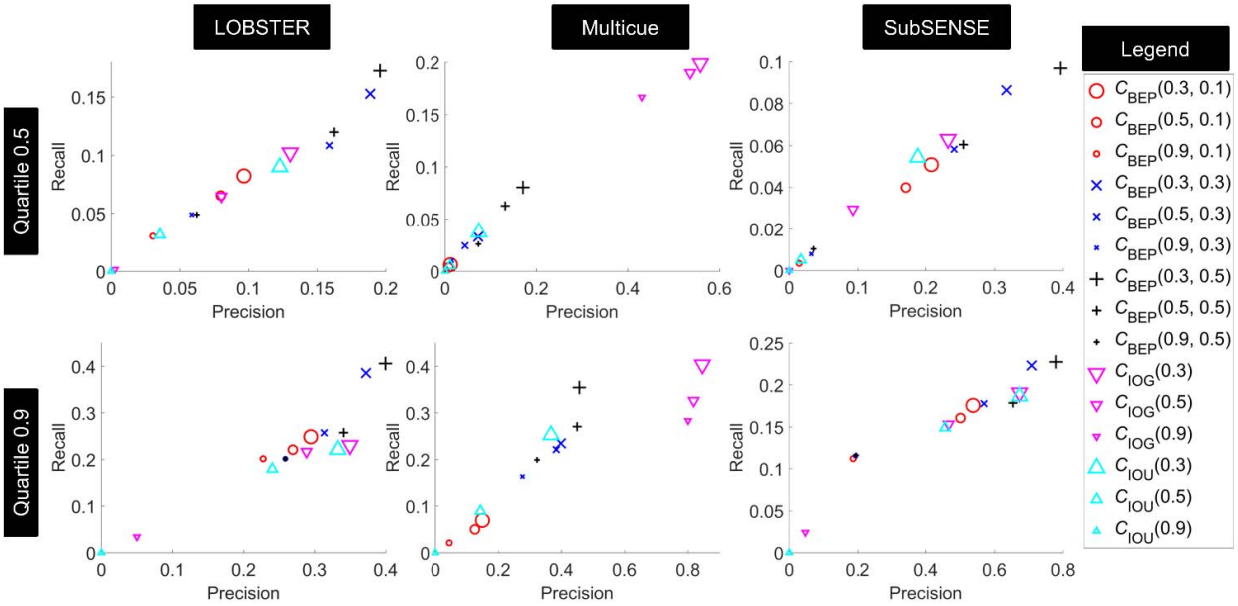

Fig. 2. Comparison of metrics for three methods providing superior quantitative results with balanced precision and recall on visible range videos.

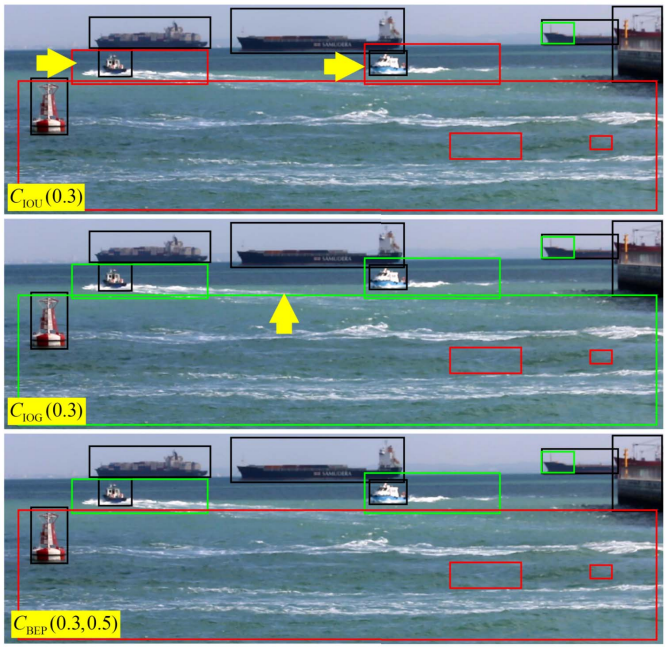

Fig. 3. Illustration of why metrics other than the standard Intersection over union (IOU) may be needed for maritime videos. Black boxes indicate the ground truth. Multicue method is used to detect the objects (red and green boxes). Green and red boxes indicates the DOs identified as the true positive and false positive, respectively.

The yellow arrows in the top row of Fig. 3 indicate the DOs which are wider than the corresponding ground truths due to the presence of wakes. IOU clearly fails in such cases, especially with a long or wide trail of wake. This is due to its property of penalizing the false positive regions of the DO, as discussed before. It is seen in the second and third rows that IOG and BEP deal with such cases better. On the other hand, the yellow arrow in the second row of Fig. 3 shows

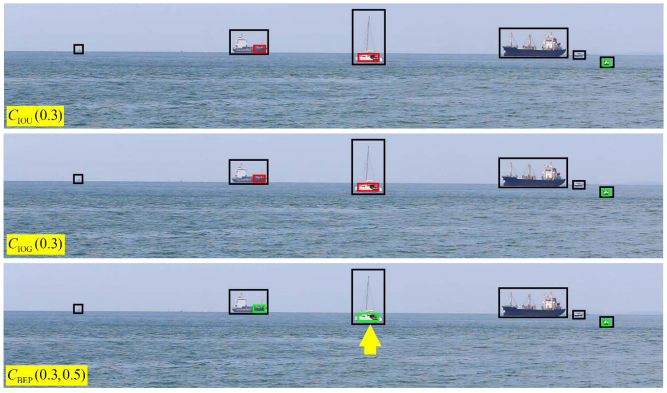

Fig. 4. Illustration of why metric C_{bottom} may be really needed for some cases. Black boxes indicate the ground truth. LOBSTER method is used to detect the objects (red and green boxes). Green and red boxes indicates the DOs identified as the true positive and false positive, respectively.

the inherent limitation of IOG, where IOG incorrectly detects a true positive if a very large DO is falsely detected due to large dynamics of water. The yellow arrow in the third row of Fig. 4 shows a sail-boat for which most methods may detect only the base and only BEP can correctly identify the corresponding DO as true positive. However, the DO to the left of this object is an example in which the BEP enables classification of the DO as true positive. It might be argued that this is a poor detection and that a metric should be more stringent in classifying such DO as true positive. Nevertheless, we highlight that most BS methods only allow for partial detection of the object. This motivated our choice of small value (0.3) for the thresholds c_0 and X_0. The last qualitative example in Fig. 5 shows two objects in close vicinity with one object of very low contrast. The BS method yields only one DO for both the objects, as highlighted with the yellow arrow

Fig. 5. Illustration of why we should not discard IOU. Black boxes indicate the ground truth. SuBSENSE method is used to detect the objects (red and green boxes). Green and red boxes indicates the DOs identified as the true positive and false positive, respectively.

in the top row of Fig. 5. Only IOU is effective in assigning this DO as a false positive. Yet, it may be argued that in maritime, it is beneficial to assign this as a true positive and should be considered in collision avoidance and navigation planning. Thus, we consider that the problem of defining suitable metrics for maritime scenario is an important one.

We also highlight that segmentations other than bounding boxes, such as boundary regions, or metrics based on background pixel classification may be used. Examples include PSNR, SSIM, and D-score. However, these form of performance evaluations may not be always effective. For example, using boundary regions with a pixel based metric will be ineffective for objects such as sail boats, objects from which only a small part is detected as DO, and objects with long wakes. Additionally, forming region based ground truth is very difficult in maritime videos and requires man hours which are inordinately large in comparison to the bounding box annotations of the ground truth. Hence, pixel or region based metrics are not used in this paper.

VI. Results

We present the results of the 22 methods for SMV and SMIR datasets using IOU(0.3), IOG(0.3), and BEP(0.3, 0.5), based on our analysis in section V. Results for SMV and SMIR datasets are discussed in sections VI-A and VI-B, respectively.

A. Visible Range Videos

1) Quantitative Evaluation: The performance of various methods are shown in Fig. 6 and listed in Table III. It is noted that the precision of all the methods is quite poor, indicating huge number of false positives, irrespective of the metric used. Recall is significantly better for most methods. Nevertheless, recall for all methods is quite low in comparison to the recall observed for the datasets and problems considered in [5] and [7]. We provide a visual summary of Table III in Fig. 7 for the metric $C_{\text{BEP}}(0.3, 0.5)$. In Fig. 7, we divide the precision-recall plot in 6 regions, namely, poor-precision-poor-recall, poor-precision-moderate-recall, poor-precision-good-recall, moderate-precision-poor-recall, and other two unpopulated regions. In lieu of the generically poor precision, moderate precision here is being defined as precision >0.1.

It is notable that adaptive-median method among the basic methods, simple Gaussian method and fuzzy Gaussian method among Gaussian background methods, and eigen-background method among machine learning methods provide good recall, although precision of all of these methods is very poor. The reason of poor precision is evident in qualitative results presented in Fig. 8. These methods are ineffective at modeling the water as background, causing dynamic speckles and variations in water to be detected as foreground. We call these effects as spurious dynamics of water (SDOW).

In general, texture based methods and GFLFM provide a balance between precision and recall, as noted in Fig. 7. Notably, they provide significantly better precision than other methods, however with poor recall. The reason for this is evident in the qualitative results presented in Fig. 8(d,e). These methods show great effectiveness in modeling and subtracting SDOW and thus increasing the precision. However, they suppress slow moving and almost stationary foreground objects as background. In essence, they are effective for detecting maritime objects with patterned fast motion.

2) Qualitative Evaluation: Now, we discuss some interesting characteristics of maritime vision problem observed consistently across a variety of methods. We list the qualitative capabilities of the methods in Table IV. This is done through qualitative examples presented for all the methods in Fig. 8. We consider results of the 200th frame of the videos since the learning would be mature for almost all methods by that frame. Then, we selected those videos whose 200th frames illustrate multiple effects typical of maritime scenario. In these figures, the methods in blue colored text show 'ghost effect' and the methods in magenta colored text show false positives corresponding to 'wakes'. Lastly, we note that the detected foreground pixels may form disjoint subsets of actual foreground pixels, for example in adaptive median method (Fig. 8(a)). Such effect may result into multiple small detections for a single object. For the convenience of reference, we refer to this effect as multiple detections (MD). Most methods studied in this paper, with the exception of multicue, demonstrate MD, as noted in Table IV.

a) Basic BS methods: Among the basic methods (Fig. 8(a)), adaptive median method shows false positives due to both SDOW as well as ghost effect. The ghost effect is prominently evident for moving objects. The ship in the foreground entering from the right hand side has just started entering a few frames before and its ghost does not appear. This implies that the ghost effect is related to the history of the object. Of course, for the objects that are stationary or moving very slowly, the presence of historical learning does not result in a spatially well-separated ghost. However, this effect becomes a challenge for the fast moving objects. Its presence appears to be related to two aspects of the algorithm. The first is that the methods that use regional correspondences may have a better self-correction as the object moves away to newer spatial regions in subsequent frames. This is confirmed through the absence of ghost effect in Prati's median method, where concept of image blocks is used. The second is the importance

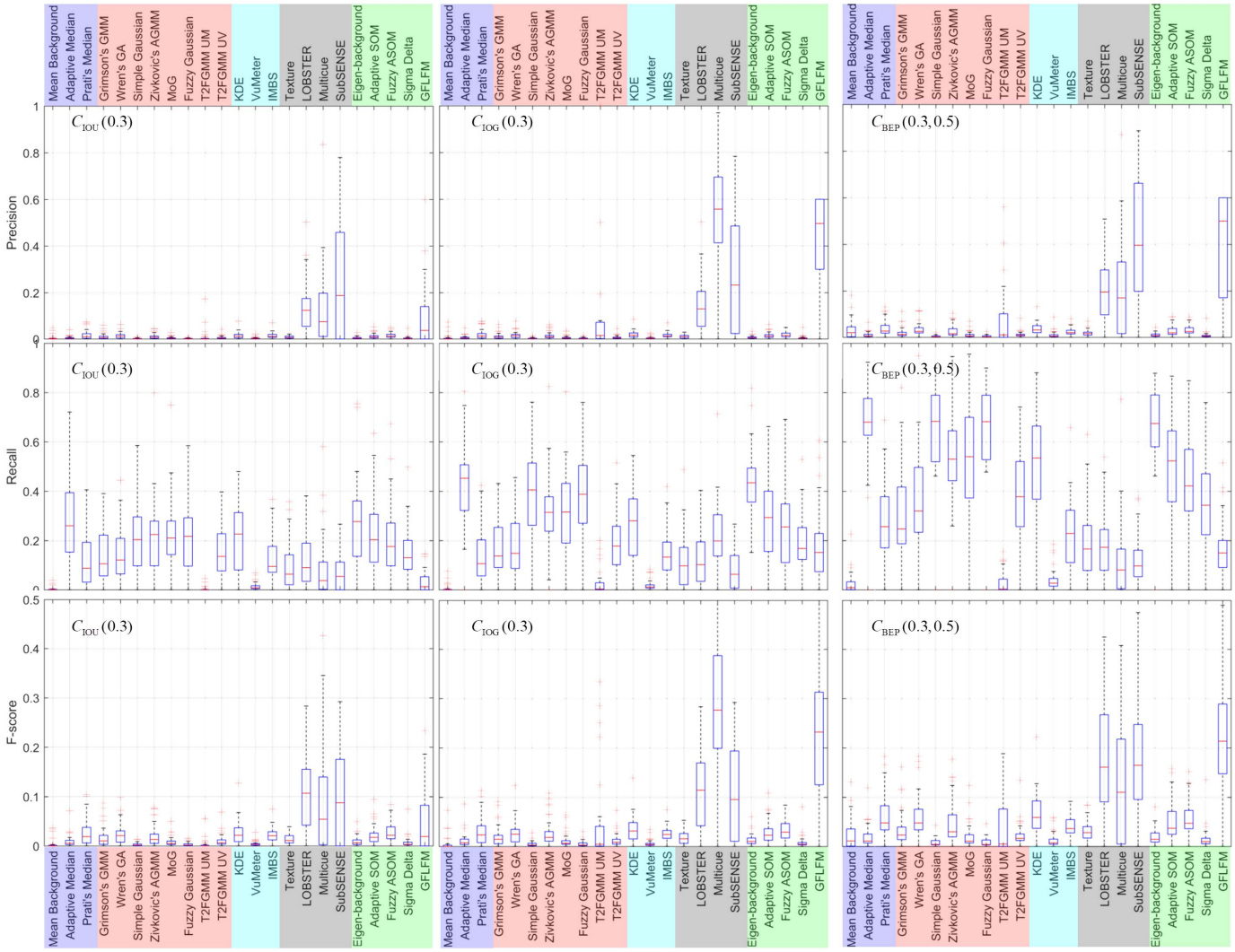

Fig. 6. Statistics of the performance metrics for visible range videos. The colors denote the different types of methods and match the colors used in Table III.

of forgetting old memories or having smaller temporal scales for learning. For example, the small value of T in mean background method is the reason of lack of history of the moving objects. However, temporal scales of learning are also directly related to the robustness of background models. Thus, there appear to be a conflicting requirements for the temporal scales.

b) Methods that use Gaussian background models: Among these methods (Figs. 8(c)), simple Gaussian method, fuzzy Gaussian method, and MoG suffer from false detections due to SDOW. Grimson's GMM, Wren's GA, MoG, and T2FuzzyGMM-UV are ineffective at modeling wakes as background. Simple Gaussian, Zivkovic's GMM, and Fuzzy Gaussian suffer from the ghost effect. While T2FuzzyGMM-UM is free of all these effects, it appears to be more stringent on the foreground as well, delegating a large portion of the foreground to the estimated background. Nevertheless, its effectiveness in modeling maritime background is commendable and post-processing such as foreground region growing etc. may aid in improving the foreground detection. We also note that T2FuzzyGMM-UM suppresses stationary or slowly moving foreground objects as background.

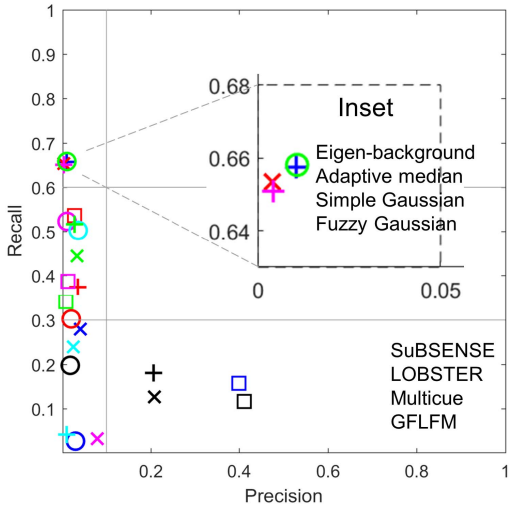

Fig. 7. Visual summary of Table III for $C_{BEP}(0.3, 0.5)$ is provided here. The markers corresponding to the methods match the markers in Table III.

On one hand, it is not expected for methods performing dynamic BS to detect stationary objects as foreground, on the other hand, the detection of these objects is important in

TABLE III
MEAN VALUES OF PRECISION, RECALL, AND F-SCORE OF DIFFERENT METHODS FOR VISIBLE RANGE VIDEOS

Methods	$C_{\mathrm{IOU}}(0.3)$			$C_{\mathrm{IOG}}(0.3)$			$C_{\mathrm{BEP}}(0.3, 0.5)$			Time
	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score	(ms/frame)
Basic BS methods										
Mean [8] ◯	0.004	0.004	0.004	0.006	0.006	0.006	0.030	0.027	0.024	**63.13**
Adaptive Median [14] +	0.005	**0.286**	0.010	0.007	**0.422**	0.013	0.010	**0.657**	0.020	109.98
Prati's median [46] ×	**0.019**	0.119	**0.029**	**0.021**	0.137	**0.033**	**0.041**	0.280	**0.064**	149.52
Methods that use Gaussian background models										
Grimson's GMM [47] ◯	0.010	0.141	0.018	0.012	0.172	0.021	0.021	0.303	0.036	125.19
Wren's GA [48] +	**0.014**	0.141	**0.024**	0.017	0.177	0.029	0.036	0.374	**0.061**	**71.73**
Simple Gaussian [49] ×	0.001	0.220	0.003	0.002	**0.401**	0.005	0.004	**0.653**	0.008	170.00
Zivkovic's AGMM [13] □	0.012	0.217	0.022	0.016	0.316	0.029	0.028	0.536	0.051	85.41
MoG [47], [13] ◯	0.005	**0.228**	0.009	0.006	0.319	0.012	0.011	0.523	0.021	138.96
Fuzzy Gaussian [9] +	0.002	0.222	0.003	0.003	0.399	0.005	0.004	0.651	0.008	187.94
T2Fuzzy GMM-UM [15] ×	0.010	0.004	0.005	**0.153**	0.028	0.045	**0.079**	0.032	0.039	84.19
T2Fuzzy GMM-UV [15] □	0.006	0.158	0.011	0.007	0.199	0.013	0.013	0.387	0.024	194.47
Methods that use other statistical background models										
KDE [12] ◯	**0.015**	**0.208**	**0.027**	**0.019**	**0.270**	**0.034**	**0.036**	**0.503**	**0.066**	102.52
VuMeter [50] +	0.003	0.014	0.005	0.004	0.020	0.006	0.011	0.042	0.014	**55.55**
IMBS [17] ×	0.014	0.125	0.023	0.016	0.148	0.026	0.026	0.239	0.043	237.60
Methods that use texture and color descriptors for background										
Texture [51] ◯	0.008	0.095	0.014	0.010	0.116	0.017	0.018	**0.198**	0.031	1431.67
Multicue [52] +	0.144	0.087	0.101	**0.552**	**0.228**	**0.291**	0.208	0.126	0.146	**186.06**
LOBSTER [53] ×	0.137	**0.112**	**0.109**	0.146	0.120	0.118	0.206	0.181	**0.172**	695.68
SuBSENSE [11] □	**0.258**	0.073	0.102	0.275	0.078	0.110	**0.411**	0.116	0.164	1181.73
Machine learning based methods										
Eigen-background [10] ◯	0.005	**0.268**	0.009	0.007	**0.425**	0.013	0.011	**0.658**	0.021	228.78
Adaptive SOM [55] +	0.012	0.230	0.022	0.015	0.307	0.028	0.028	0.517	0.052	126.04
Fuzzy ASOM [56] ×	0.015	0.203	0.027	0.018	0.254	0.033	0.034	0.445	0.061	130.38
Sigma-Delta [57] □	0.004	0.150	0.008	0.005	0.191	0.009	0.009	0.342	0.016	**104.24**
GFLFM [16] □	**0.089**	0.032	**0.044**	**0.409**	0.180	**0.231**	**0.399**	0.157	**0.210**	363 sec/frame

maritime problem. Here again, the role of temporal scales of learning becomes relevant. Stationary or slow moving objects correspond to almost no additional learning over a long period of time except due to illumination changes or occlusion. This property may be useful in detecting stationary objects in maritime background.

c) Methods that use other statistical background models: Among these methods (Fig. 8(b)), KDE is more sensitive to SDOW, although not as sensitive as simple Gaussian, fuzzy Gaussian, etc. KDE also suffers from the ghost effect. While IMBS is better among these methods for detecting foreground pixels, even related to stationary or slow moving foreground objects, it will benefit from some form of region growing techniques as post-processing. We note that IMBS is ineffective for modeling wakes as foreground, as seen in the second example in Fig. 8(b). It suffers from a very alleviated form of the ghost effect, where it is sensitive to recent motion but not to prolonged history, as seen in the first example in Fig. 8(b).

d) Methods that use texture and color descriptors for background: Except the basic texture method, the other three methods are very effective in suppressing SDOW and wakes. Multicue and LOBSTER suffer from ghost effect, although LOBSTER is less severely effected in comparison to other methods that are afflicted by ghost effect. Multicue has a tendency of forming foreground blobs much larger than the foreground. In fact, Multicue may easily bleed the foreground blobs into the water region and incorrectly detect a large portion of water as foreground. This is evident in the second example of Fig. 8(d) as well as Fig. 1(c). On one hand, Multicue method ensures that the suppressed background is unlikely to have any foreground objects. On the other hand, Multicue makes processing of the detected foreground region for finer detection of foreground objects mandatory. Lastly, SuBSENSE is effective in detecting fast moving foreground objects, and suppressing all forms of maritime background. It however suppresses the stationary and slow moving objects as background, which is undesirable.

e) Machine learning based methods: All machine learning based methods, except sigma-delta and GFLFM methods suffer from ghost effect. Sigma-delta method suffers from false positives due to wakes. GFLFM is effective in detecting fast moving foreground objects and suppressing SDOW. It, however, suppresses the stationary and slow moving objects as background. It also suffers from MD and inability to model wakes as background. All the machine learning methods, except GFLFM, are affected by SDOW to some extent.

3) Summary of Results in Visible Range: The quantitative and qualitative results appear conflicting at a cursory glance at Tables III and IV. However, they are actually consistent and quite insightful, as we discuss here. A major observation from Table III is that all methods have poor precision, i.e. they generate a large number of false positive detections. The largest number of false detections is due to the incapability of the current methods to deal with SDOW. The second largest contributor in false detections is wakes and the third largest contributor is the MD. Among all the methods used in this dataset, only LOBSTER is effective in dealing with these three causes of false detections. In general, methods that use texture

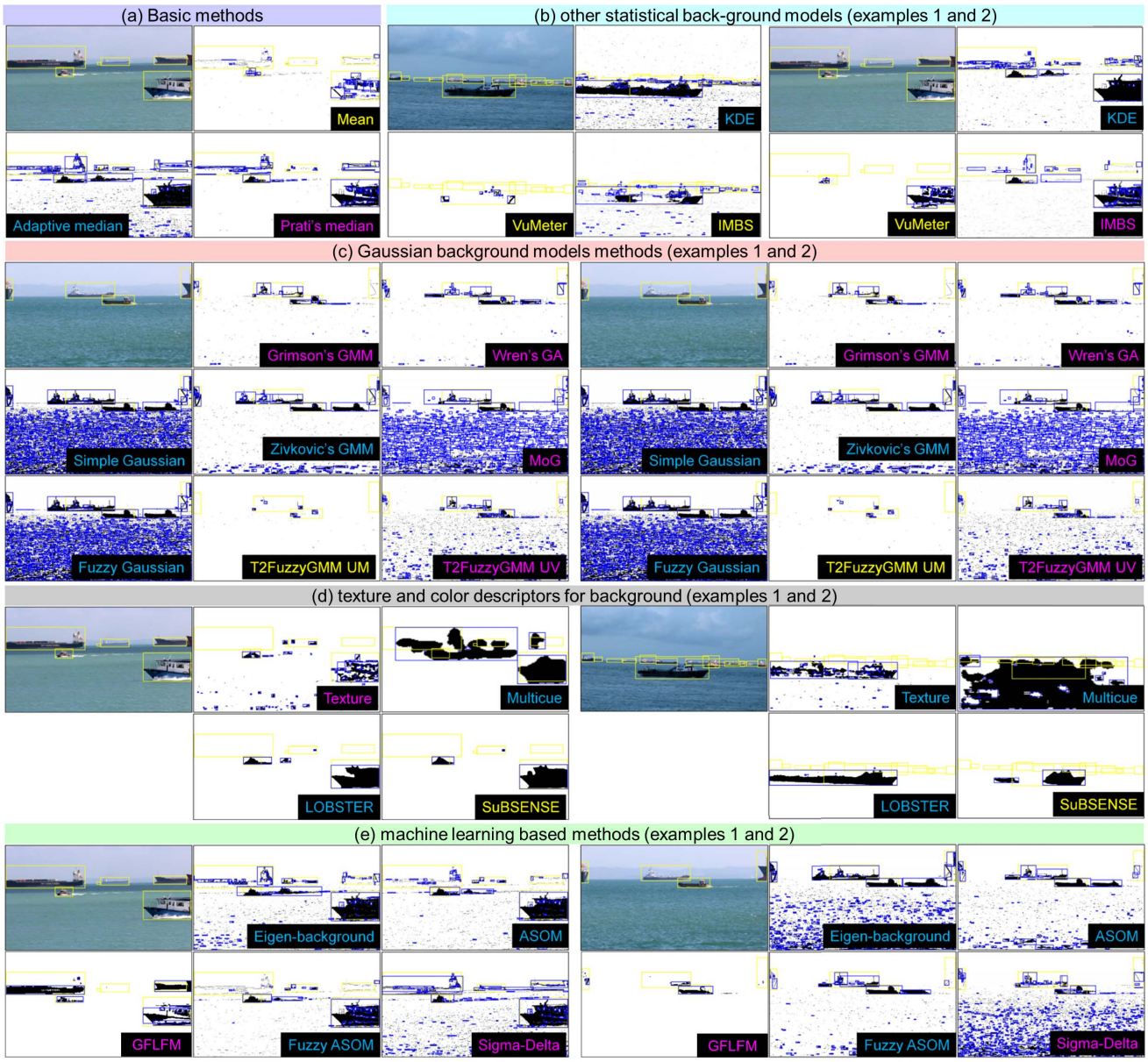

Fig. 8. Qualitative comparison of the methods using different examples. Yellow boxes indicate the ground truth while the blue boxes indicate the detected objects. Magenta colored text indicates failure to deal with wakes and blue colored text indicates ghost effect.

and color descriptors (Multicue, LOBSTER, and SuBSENSE) are better at dealing with these issues. Thus, it is not surprising that their precision is superior to other methods.

MD is potentially an easier problem to tackle among all these issues. If MD is not significant, some form of region growing or graph cut methods may provide simple solutions to the MD problem. However, modeling SDOW as background is more difficult; only 9 methods out of 22 can model SDOW as background. Modeling wakes as background is even tougher; only 6 methods out of 22 are capable of dealing with wakes. We do note that there have been some works on detecting water regions that are considered as background [60], [61].

Lastly, ghost effect is related to temporal scales of learning and the motion characteristics of the vessels. Potentially, multiple temporal scales of learning and incorporation of additional regional cues may help in suppressing ghost effect. Four inferences can be drawn from these analysis. First, no non-regional pixel-only methods can alone be useful in maritime vision problem. Incorporation of additional background descriptors potentially with regional characteristics is essential. Second, background modeling at multiple temporal scales may be quite critical for suppression of ghost effect. Third, background suppression for maritime vision problems need additional region growing or region cleaning procedures to effectively deal with false detections due to SDOW, wakes, and MD. Fourth, new background modeling approaches for SDOW and wakes should be developed.

B. NIR Range Videos

1) Quantitative Evaluation: The performance of the methods for the NIR range videos is compared in Fig. 9 and Table V. Similar to visible range videos, the precision

TABLE IV

QUALITATIVE CHARACTERISTICS OF METHODS IN THE CONTEXT OF MARITIME VISION PROBLEM ARE LISTED HERE. ✓ INDICATES EFFECTIVENESS IN DEALING WITH THE ISSUE, × INDICATES INEFFECTIVENESS, AND − INDICATES INCONCLUSIVENESS

	Visible range				NIR range			
Methods	SDOW	Wakes	Ghost	MD	SDOW	Wakes	Ghost	MD
Basic BS methods								
Mean	✓	×	✓	×	✓	✓	✓	×
Adaptive Median	×	−	×	×	×	×	×	−
Prati's median	✓	×	✓	×	✓	×	✓	×
Methods that use Gaussian background models								
Grimson's GMM	✓	×	✓	×	×	×	✓	✓
Wren's GA	✓	×	✓	×	×	×	✓	×
Simple Gaussian	×	×	×	−	×	×	×	−
Zivkovic's GMM	×	−	×	×	×	×	×	−
MoG	×	×	✓	−	×	×	✓	×
Fuzzy Gaussian	×	×	×	−	×	×	×	−
T2FuzzyGMM UM	✓	✓	✓	×	✓	✓	✓	×
T2FuzzyGMM UV	×	×	×	×	×	×	✓	✓
Methods that use other statistical background models								
KDE	×	−	×	×	✓	✓	✓	×
VuMeter	✓	✓	✓	×	✓	✓	✓	×
IMBS	−	×	×	×	×	×	✓	×
Methods that use texture and color descriptors for background								
Texture	×	×	✓	×	×	×	×	×
Multicue	✓	−	×	✓	×	×	✓	✓
LOBSTER	✓	✓	×	✓	✓	×	×	×
SuBSENSE	✓	✓	✓	−	✓	✓	✓	×
Machine learning based methods								
Eigen-background	×	×	×	×	×	×	×	✓
ASOM	×	✓	×	×	×	×	×	✓
Fuzzy ASOM	×	×	✓	✓	×	×	×	✓
Sigma-Delta	×	✓	✓	×	×	×	✓	×
GFLFM	✓	×	✓	×	not tested			

of all methods is quite poor in the NIR range. Interestingly, the recall is poorer for NIR range videos than the visible range videos. We attribute this to the presence of only one intensity channel in NIR range as compared to the three color channels in the visible range.

We provide a visual summary of Table V in Fig. 10 for the metric $C_{\text{BEP}}(0.3, 0.5)$. We follow the partitioning used in Fig. 7. For the NIR range videos, only Simple Gaussian method and Fuzzy Gaussian methods provide good recall while Eigen-background barely misses the zone of good-recall.

Similar to the visible range videos, methods that use texture and color cues (Multicue, LOBSTER, and SuBSENSE) provide a balance between precision and recall. Prati's median also surprisingly provides similar balance for NIR range videos only but not for visible range videos.

2) Qualitative Evaluation: The qualitative analysis for NIR range videos is performed along the same lines as visible range videos. We present qualitative examples for NIR range videos in Fig. 11. We discuss the qualitative results for each category of methods below, while the general qualitative characteristics are listed in Table IV.

a) Basic BS methods: As shown in Fig. 11(a) using two examples, Mean method suppresses almost all the background, including SDOW and wakes, but MD is quite severe for the Mean method in NIR range videos. Adaptive median is ineffective for SDOW and wakes. It also generates ghost effect. Prati's median suffers from MD and inability to model wakes as background. Among the three, although Adaptive median generates a large number of false positives, it provides better recall than the other two methods, as evident qualitatively in Fig. 11(a) and quantitatively in Table V.

b) Methods that use Gaussian background models: An example of qualitative results for these methods is given in Fig. 11(d). It is seen that among all the methods, Grimson's GMM and T2FuzzyGMM-UV perform better despite being unable to deal with wakes and ineffective for SDOW. SDOW severely afflicts Simple Gaussian, MoG, and Fuzzy Gaussian. MD severely affects Wren's GA and T2FuzzyGMM-UM. Among these two, although Wren's GA is ineffective for wakes, it is more effective than T2FuzzyGMM-UM for detecting the true objects.

c) Methods that use texture and color descriptors for background: It is not surprising that the absence of color cues causes deterioration in the performance of Multicue, often leading to false detections and foreground bleeding (opposite of the reason behind MD) in regions with SDOW even if no actual foreground object is present there. However, ghost effect is completely absent in Multicue and quite diminished in LOBSTER for NIR range videos. The absence of color affects LOBSTER by making it ineffective in modeling wakes as background. Lastly, the lack of color increases MD in SuBSENSE.

d) Machine learning based methods: All methods is this group are ineffective for SDOW and wakes. Ghost effect is a prominent feature of Eigen-background. It afflicts ASOM and Fuzzy ASOM as well, although to a lesser severity. Among these methods, only Sigma-Delta suffers from MD.

e) Methods that use other statistical background models: Although KDE is more effective in suppressing background in NIR range videos than visible range videos, it is also more severely afflicted by MD in NIR range videos. VuMeter performs similar in either datasets and is not promising. As compared to visible range videos, IMBS performs slightly poorer in NIR range videos in suppressing wakes and SDOW.

3) Summary of Results in NIR Range: It is conclusively evident that NIR range videos pose more challenge in comparison to the visible range videos, mainly due to the lack of color cues. The other observations for visible range videos generically apply to the NIR range videos as well.

VII. DISCUSSION

Due to the specific characteristics of the maritime environment, conventional BS algorithms, including state of the art ones, have been rendered largely ineffective. In this section, we focus on those characteristics and offer solutions that could prove effective in countering the problems in BS due to them.

A. SDOW and MD

Dynamic background methods used in this study are quite sophisticated and have previously demonstrated ability to deal with complex backgrounds. Some of them have been applied to dynamic water backgrounds as well [60], [61], although the region of view is the same order as the foreground object both in lateral span as well as in focal depth in these methods. Practical maritime vision involves large view region and large depths of view. Consequently, the physical scales vary greatly

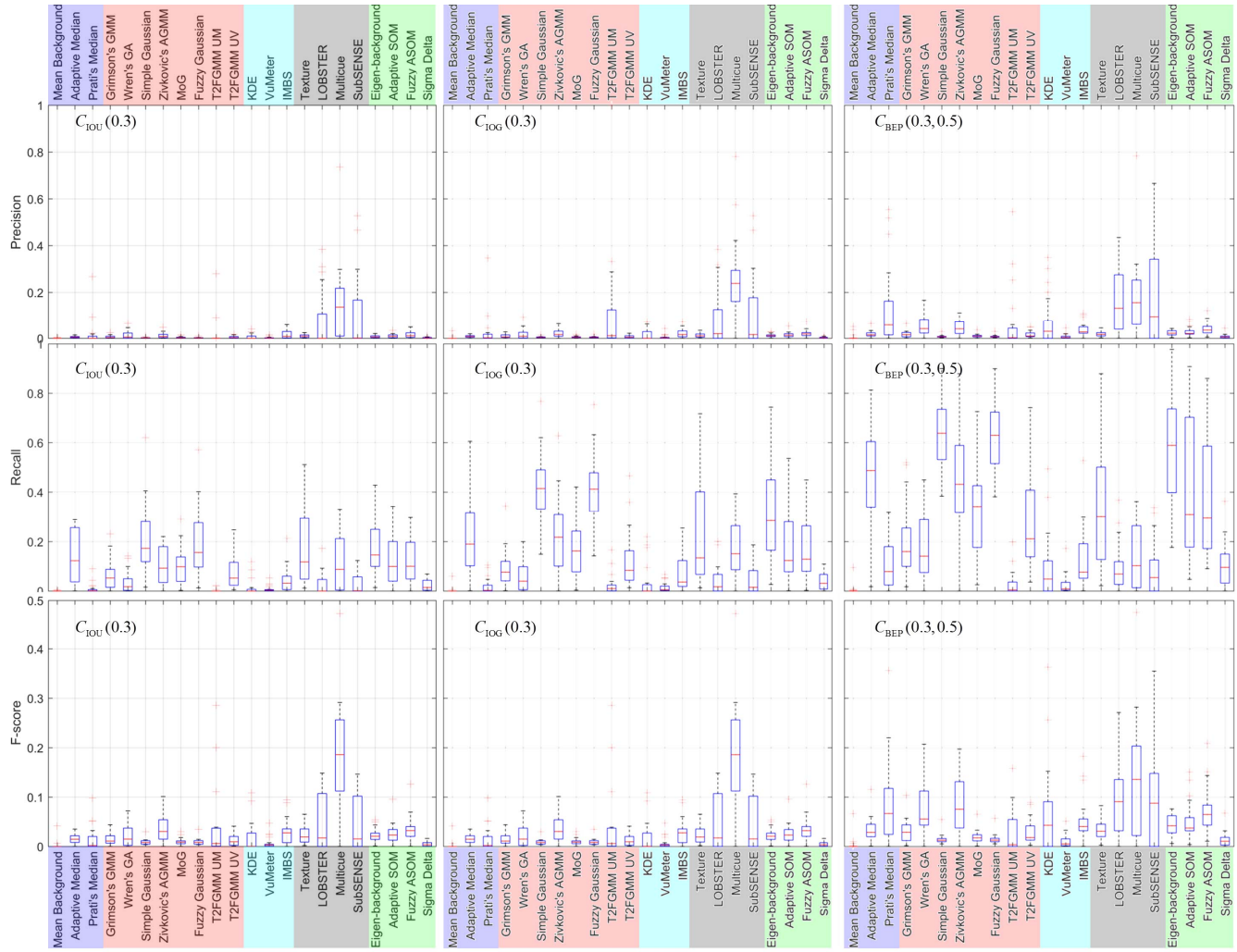

Fig. 9. Statistics of the performance metrics for NIR range videos. The colors denote the different types of methods and match the colors used in Table V.

across the scene due to non-linear mapping of the world space coordinates to the sensor coordinates. Further, glint, speckle, color variations induced due to illumination conditions, and underwater topography together form spurious dynamics of water with large variation in the spatial scales in the image. The methods that are able to model SDOW well (for example SuBSENSE [11]) usually intensifies the issue of MD. MD is often a result of the edge regions of foreground being assigned to background. In most situations, this happens because the method models drastic variations as background in order to model SDOW and the edge regions of the foreground objects also demonstrate such drastic variations.

B. Modeling Wakes

Our study shows that modeling wakes as background is difficult for existing BS methods. Even IMBS [17], developed specifically for maritime vision, shows failure in modeling wakes as background. As with SDOW, methods that model wakes as background suffer from MD.

C. Ghost Effect

The speeds of foreground objects in maritime vision problems vary greatly in both the physical units as well as image units. As a consequence, the temporal scales of learning which are suitable for certain speeds are unsuitable for other speeds. This causes ghost effect related to the historical trail of the foreground objects for which the temporal scales of learning were unsuitable.

D. Intensity Only Data and Infrared Spectrum

Our results show deteriorated performance of most methods when applied to near infrared (NIR) range videos instead of visible range videos. Poor performance of most methods even in visible range, and worse performance in NIR range indicates the need for the design of new BS algorithms that can provide better performance for maritime vision in the absence of color cues.

E. Potential Solutions

Improvement in modeling SDOW and wakes involves trade-off with MD. A potential solution is to choose methods that are effective in modeling SDOW and wakes as the core and address the issue of MD in the post-processing. For instance, growing regions of multiple detection and merging regions with local homogeneous properties after region growing could be used to reduce multiple detections.

TABLE V
MEAN VALUES OF PRECISION, RECALL, AND F-SCORE OF DIFFERENT METHODS FOR NIR RANGE VIDEOS

Methods	$C_{\mathrm{IOU}}(0.3)$			$C_{\mathrm{IOG}}(0.3)$			$C_{\mathrm{BEP}}(0.3, 0.5)$			Time
	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score	(ms/frame)
Basic BS methods										
Mean [8] ◯	0.000	0.000	0.000	0.002	0.003	0.002	0.005	0.005	0.004	**39.92**
Adaptive Median [14] +	0.005	**0.136**	**0.010**	0.009	**0.228**	**0.016**	0.018	**0.469**	0.035	57.76
Prati's median [46] ×	**0.020**	0.009	0.009	**0.027**	0.018	0.014	**0.135**	0.102	**0.086**	78.48
Methods that use Gaussian background models										
Grimson's GMM [47] ◯	0.006	0.062	0.010	0.009	0.096	0.016	0.018	0.194	0.031	85.36
Wren's GA [48] +	0.013	0.037	0.016	0.019	0.058	0.023	0.058	0.192	0.079	**40.07**
Simple Gaussian [49] ×	0.002	**0.207**	0.004	0.004	**0.415**	0.009	0.007	**0.644**	0.014	98.93
Zivkovic's AGMM [13] □	0.011	0.102	**0.020**	0.022	0.215	0.038	0.050	0.457	**0.085**	50.71
MoG [47], [13] ◯	0.003	0.099	0.006	0.005	0.170	0.010	0.011	0.349	0.020	82.36
Fuzzy Gaussian [9] +	0.002	0.199	0.004	0.005	0.409	0.009	0.007	0.637	0.015	114.57
T2Fuzzy GMM-UM [15] ×	**0.017**	0.002	0.003	**0.183**	0.029	**0.046**	**0.067**	0.029	0.029	61.30
T2Fuzzy GMM-UV [15] □	0.005	0.084	0.009	0.007	0.130	0.013	0.016	0.291	0.030	133.36
Methods that use other statistical background models										
KDE [12] ◯	0.006	0.013	0.007	0.016	0.036	0.019	**0.075**	0.093	**0.069**	**41.37**
VuMeter [50] +	0.002	0.008	0.003	0.005	0.015	0.007	0.008	0.028	0.011	51.99
IMBS [17] ×	**0.016**	**0.045**	**0.022**	**0.021**	**0.068**	**0.029**	0.038	**0.127**	0.054	108.63
Methods that use texture and color descriptors for background										
Texture [51] ◯	0.009	**0.163**	0.018	0.012	**0.205**	0.023	0.020	**0.319**	0.037	563.47
Multicue [52] +	**0.145**	0.109	**0.115**	**0.254**	0.170	**0.184**	0.175	0.134	**0.139**	**145.86**
LOBSTER [53] ×	0.076	0.028	0.035	0.088	0.043	0.046	0.160	0.095	0.094	492.00
SuBSENSE [11] □	0.094	0.027	0.036	0.107	0.042	0.048	**0.193**	0.086	0.099	827.93
Machine learning based methods										
Eigen-background [10] ◯	0.007	**0.169**	0.013	0.012	**0.325**	0.024	0.023	**0.585**	0.044	183.30
Adaptive SOM [55] +	0.010	0.119	0.018	0.014	0.190	0.026	0.029	0.380	0.053	93.76
Fuzzy ASOM [56] ×	**0.015**	0.117	**0.025**	**0.020**	0.177	**0.035**	**0.042**	0.378	**0.074**	95.78
Sigma-Delta [57] □	0.002	0.024	0.003	0.003	0.039	0.005	0.007	0.108	0.013	**59.94**

Tackling the problem of ghost effect may require exploitation of regional characteristics instead of only pixel-level background learning, use of multiple time scales of learning, and mechanisms to forget historical knowledge. Lastly, background subtraction methods for NIR videos may require development of newer more complex background descriptors with regional features to counter the absence of color cues. It may be interesting to use multi-scale spatio-temporal features for this purpose. We expect that spectral feature such as wavelets or level sets may be useful for this purpose.

Further, we note that the present study does not cover the influence of learning rates on the performance. The various control parameters of each method determine background modeling and learning rates. In turn, they can provide trade-offs in terms of various challenges and influence the ghost effect versus false negatives. However, such study requires method-specific attention and deserves a dedicated study for better performing methods. Such study is currently out of the scope of this paper. We invite the attention of other active researchers working on background suppression to explore the trade-offs in terms of the above identified challenges.

Lastly, we discuss the computation times of the methods. We note that for reasonable real-time processing, it is preferred to have computation rate of >25 frames per second, i.e. <40 ms per frame. However, all benchmarked methods require an average of more than 50 ms per frame for visible range videos (see Table III). Some methods take as much as >1000 ms per frame and GFLFM takes few hundred seconds per frame. The average computation time per frame is significantly lesser for each method in NIR videos (see Table V). However, we note that none of these methods have been tested with GPUs and thus there is a large scope for reduction in computation time. Further, once accurate object detection can be performed, the computation time can be reduced by optimization of algorithmic implementation and using parallel processing. Thus, while we acknowledge the importance of practically small computation time, we do not consider computation time as the currently critical requirement for maritime vision.

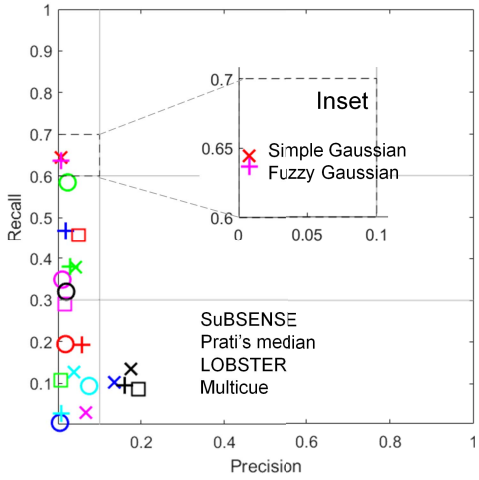

Fig. 10. Visual summary of Table V for $C_{\mathrm{BEP}}(0.3, 0.5)$ is provided here. The markers corresponding to the methods match the markers in Table V.

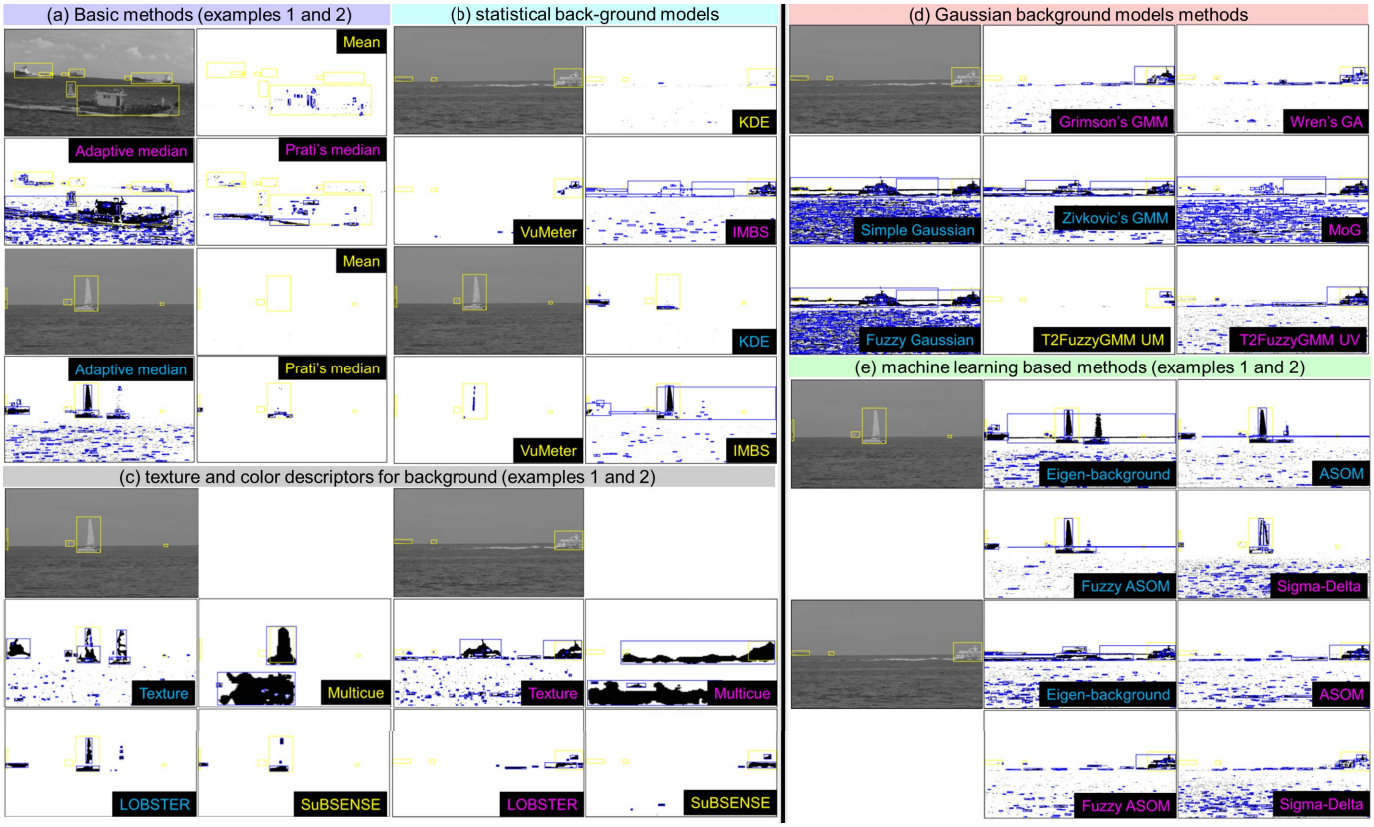

Fig. 11. Qualitative comparison of various methods for example frames of NIR range videos. Yellow boxes indicate the ground truth while the blue boxes indicate the detected objects. Magenta colored text indicates failure to deal with wakes and blue colored text indicates ghost effect.

VIII. Conclusion

The presented study provides an in-depth insight into the challenges of background subtraction in maritime computer vision. The results in this study demonstrate that maritime background subtraction is inordinately challenging for the state-of-the-art background subtraction methods. While we determine the limiting issues for each method, we also indicate the windows of opportunity for the design of algorithms better suited for maritime vision.

An in-depth qualitative analysis reveals that maritime background subtraction is plagued with four major effects, namely spurious dynamics of water (speckle, glint, color variations), wakes, ghost effect, and multiple detections. Underlying reasons point out that practical solutions may be incorporated in the existing maritime methods for ghost effect and multiple detections through incorporation of regional cues, multiple temporal scales, and foreground region growing. However, SDOW and wakes may need new background modeling approaches. Incorporation of forgetting mechanisms may help the background models in dealing with not only ghost detections but also in SDOW and wakes due to their transient nature. Our study also indicates the need of new metrics that can incorporate wakes, occlusion, and partial detections typical of maritime vision in the performance evaluation.

We hope that the challenges and potential solutions identified in this work will lead to the development of new robust maritime background suppression methods which will be a big leap towards the realization of autonomous maritime vehicles.

Acknowledgement

The authors acknowledge Rolls Royce@NTU Corporate Lab, where the work was conducted.

References

[1] 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.

[2] T. Waterborne, "Implementating waterborne strategic research agenda," Waterborne TP, Brussels, Belgium, Tech. Rep., 2011.

[3] V. Ablavsky, "Background models for tracking objects in water," in *Proc. Int. Conf. Image Process.*, vol. 2, Sep. 2003, pp. III-125–III-128.

[4] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabaly, and C. Quek, "Challenges in video based object detection in maritime scenario using computer vision," in *Proc. 19th Int. Conf. Connected Vehicles*, 2017, pp. 1–6.

[5] 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.

[6] A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequièvre, "A benchmark dataset for outdoor foreground/background extraction," in *Proc. Asian Conf. Comput. Vis.*, Springer, 2012, pp. 291–300.

[7] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 393–400.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

[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] Z. Zhao, T. Bouwmans, X. Zhang, and Y. Fang, "A fuzzy background modeling approach for motion detection in dynamic backgrounds," in *Multimedia and Signal Processing* (Communications in Computer and Information Science), vol. 346, F. L. Wang, J. Lei, R. W. H. Lau, and J. Zhang, Eds. Berlin, Germany: Springer, 2012.

[16] B. Xin, Y. Tian, Y. Wang, and W. Gao, "Background Subtraction via generalized fused lasso foreground modeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4676–4684.

[17] D. Bloisi and L. Iocchi, "Independent multimodal background subtraction," in *Proc. Int. Conf. Comput. Modeling Objects Presented Images, Fundamentals, Methods Appl.*, 2012, pp. 39–44.

[18] A. Vacavant, L. Tougne, L. Robinault, and T. Chateau, "Special section on background models comparison," *Comput. Vis. Image Understand.*, vol. 122, pp. 1–3, May 2014.

[19] T. Bouwmans, J. Gonzàlez, C. Shan, M. Piccardi, and L. Davis, "Special issue on background modeling for foreground detection in real-world dynamic scenes," *Mach. Vis. Appl.*, vol. 25, no. 5, pp. 1101–1103, 2014.

[20] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Comput. Sci. Rev.*, vols. 11–12, pp. 31–66, May 2014.

[21] W. Kim and C. Jung, "Illumination-invariant background subtraction: Comparative review, models, and prospects," *IEEE Access*, vol. 5, pp. 8369–8384, 2017.

[22] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur, "A survey of vision-based traffic monitoring of road intersection," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2681–2698, Oct. 2016.

[23] S. K. Choudhury, P. K. Sa, S. Bakshi, and B. Majhi, "An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios," *IEEE Access*, vol. 4, pp. 6133–6150, 2017.

[24] Z. Chen, T. Ellis, and S. A. Velastin, "Vision-based traffic surveys in urban environments," *J. Electron. Imag.*, vol. 25, no. 5, p. 051206, 2016.

[25] S. Varadarajan, H. Wang, P. Miller, and H. Zhou, "Fast convergence of regularised region-based mixture of Gaussians for dynamic background modelling," *Comput. Vis. Image Understand.*, vol. 136, pp. 45–58, Jul. 2015.

[26] K. L. Chan, "Detection of foreground in dynamic scene via two-step background subtraction," *Mach. Vis. Appl.*, vol. 26, no. 6, pp. 723–740, 2015.

[27] A. K. S. Kushwaha and R. Srivastava, "Maritime object segmentation using dynamic background modeling and shadow suppression," *Comput. J.*, vol. 59, no. 9, pp. 1303–1329, Sep. 2016.

[28] G. Gemignani and A. Rozza, "A robust approach for the background subtraction based on multi-layered self-organizing maps," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5239–5251, Nov. 2016.

[29] Z. Zeng, J. Jia, Z. Zhu, and D. Yu, "Adaptive maintenance scheme for codebook-based dynamic background subtraction," *Comput. Vis. Image Understand.*, vol. 152, pp. 58–66, Nov. 2016.

[30] Z. Qu and X.-L. Huang, "The foreground detection algorithm combined the temporal–spatial information and adaptive visual background extraction," *Imag. Sci. J.*, vol. 65, no. 1, pp. 49–61, 2017.

[31] C. Cuevas, R. Martínez, D. Berjón, and N. García, "Detection of stationary foreground objects using multiple nonparametric background-foreground models on a finite state machine," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1127–1142, Mar. 2017.

[32] Z. Zhong, B. Zhang, G. Lu, Y. Zhao, and Y. Xu, "An adaptive background modeling method for foreground segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1109–1121, May 2017.

[33] G. Ramirez-Alonso, J. Ramirez-Quintana, and M. I. Chacon-Murguia, "Temporal weighted learning model for background estimation with an automatic re-initialization stage and adaptive parameters update," *Pattern Recognit. Lett.*, vol. 96, pp. 34–44, Sep. 2017.

[34] G. Han, J. Wang, and X. Cai, "Background subtraction based on modified online robust principal component analysis," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 6, pp. 1839–1852, 2017.

[35] D. Berjón, C. Cuevas, F. Morán, and N. García, "Real-time nonparametric background subtraction with tracking-based foreground update," *Pattern Recognit.*, vol. 74, pp. 156–170, Feb. 2018.

[36] M. Balcilar and A. Sonmez, "Background estimation method with incremental iterative Re-weighted least squares," *Signal, Image Video Process.*, vol. 10, no. 1, pp. 85–92, 2016.

[37] M. Dkhil, A. Wali, and A. M. Alimi, "An intelligent system for road moving object detection," in *Proc. Int. Conf. Hybrid Intell. Syst.*, vol. 420, 2016, pp. 189–198.

[38] S. Javed, S. H. Oh, A. Sobral, T. Bouwmans, and S. K. Jung, "Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop*, Dec. 2015, pp. 930–938.

[39] D.-H. Kim and J.-H. Kim, "Effective background model-based RGB-D dense visual odometry in a dynamic environment," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1565–1573, Dec. 2016.

[40] J. Son, S. Kim, and K. Sohn, "Fast illumination-robust foreground detection using hierarchical distribution map for real-time video surveillance system," *Expert Syst. Appl.*, vol. 66, pp. 32–41, Dec. 2016.

[41] Y. Yang, Q. Zhang, P. Wang, X. Hu, and N. Wu, "Moving object detection for dynamic background scenes based on spatiotemporal model," *Adv. Multimedia*, vol. 2017, May 2017, Art. no. 5179013.

[42] H. Sajid and S.-C. S. Cheung, "Universal multimode background subtraction," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3249–3260, Jul. 2017.

[43] D. D. Bloisi, A. Pennisi, and L. Iocchi, "Parallel multi-modal background modeling," *Pattern Recognit. Lett.*, vol. 96, pp. 45–54, Sep. 2016.

[44] R. Trabelsi, I. Jabri, F. Smach, and A. Bouallegue, "Efficient and fast multi-modal foreground-background segmentation using RGBD data," *Pattern Recognit. Lett.*, vol. 97, pp. 13–20, Oct. 2017.

[45] T. Akilan, Q. M. J. Wu, and Y. Yang, "Fusion-based foreground enhancement for background subtraction using multivariate multi-model Gaussian distribution," *Inf. Sci.*, vols. 430–431, pp. 414–431, Mar. 2018.

[46] 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.

[47] 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, p. 252.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] S. J. Noh and M. Jeon, "A new framework for background subtraction using multiple cues," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 493–506.

[53] 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.

[54] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[55] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008.

[56] 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.

[57] A. Manzanera and J. C. Richefeu, "A new motion detection algorithm based on Σ−Δ background estimation," *Pattern Recognit. Lett.*, vol. 28, no. 3, pp. 320–328, 2007.
[58] S. R. Norsworthy, R. Schreier, and G. C. Temes, *Delta-Sigma Data Converters: Theory, Design, and Simulation*. Hoboken, NJ, USA: Wiley, 1996.
[59] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistics Quart.*, vol. 2, nos. 1–2, pp. 83–97, Mar. 1955.
[60] D. Gao, V. Mahadevan, and N. Vasconcelos, "On the plausibility of the discriminant center-surround hypothesis for visual saliency," *J. Vis.*, vol. 8, no. 7, p. 13, 2008.
[61] A. Mumtaz, W. Zhang, and A. B. Chan, "Joint motion segmentation and background estimation in dynamic scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 368–375.

Dilip K. Prasad received the B.Tech. degree in computer science and engineering from IIT (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 a Senior Research Fellow with the Rolls-Royce@ NTU Corporate Lab, Nanyang Technological University, Singapore. He has authored over 65 internationally peer-reviewed research articles. His current research interests include image processing, machine learning, and computer vision.

Chandrashekar Krishna Prasath received the M.S. degree from Coventry University, U.K., in 2009. He was a Research Associate with the Rolls-Royce@ NTU Corporate Lab, Nanyang Technological University, Singapore. His current research interests include image processing, autonomous vehicles, and computer vision.

Deepu Rajan received the B.E. degree in electronics and communication engineering from 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 IIT, 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.

Lily Rachmawati received the B.Eng. and Ph.D. degrees from National University of Singapore in 2004 and 2009, respectively. She has seven years of research experience in maritime technology. She is currently a Staff Technologist with Rolls-Royce, Singapore.

Eshan Rajabally received the M.Eng. degree from University of Newcastle-Upon-Tyne in 1999 and the Ph.D. degree from University of Bath in 2006. He has five years of maritime technology research experience. He is currently a Technologist with Rolls-Royce, Derby, U.K.

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 authored over 250 international conference and journal papers. 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 TRANSACTIONS ON NEURAL NETWORK and IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION. He is a member of the IEEE Technical Committee on Computational Finance and Economics.