International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:11, No:1, 2017

Challenges in Video Based Object Detection in Maritime Scenario Using Computer Vision

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

***Abstract*—**This paper discusses the technical challenges in maritime image processing and machine vision problems for video streams generated by cameras. Even well documented problems of horizon detection and registration of frames in a video are very challenging in maritime scenarios. More advanced problems of background subtraction and object detection in video streams are very challenging. Challenges arising from the dynamic nature of the background, unavailability of static cues, presence of small objects at distant backgrounds, illumination effects, all contribute to the challenges as discussed here.

***Keywords*—**Autonomous maritime vehicle, object detection, situation awareness, tracking.

I. Introduction

WHILE computer vision techniques have advanced video processing and intelligence generation for several challenging dynamic scenarios, research in computer vision for maritime is still in nascent state and several challenges remain open in this field [1]. This paper presents some of the challenges unique to the maritime domain.

A simple block diagram for processing of maritime videos is given in Fig. 1, where the objective is to track foreground objects and generate intelligence and situation-awareness. Foreground objects are the objects anchored, floating, or navigating in water, including sea vessels, small personal boats and kayaks, buoys, debris, etc. Air vehicles, birds, and fixed structures, such as in ports, qualify as outliers or background. Also, wakes, foams, clouds, water speckle, etc. qualify as background. The first four blocks form the core of video processing and the performance of these blocks directly affect the attainment of the objective. The challenges specific to these four blocks are discussed in Sections II to V, respectively. The challenges due to weather are discussed in Section VI.

We use 3 datasets from three different sources to illustrate the challenges. Two datasets are from the external sources, buoy dataset [2] and Mar-DCT dataset [3]. The camera in the buoy dataset is mounted on a floating buoy which is subject to significant amount of motion from one frame to another. The camera used in Mar-DCT dataset is mounted on a stationary platform on-shore. Sometimes, zoom operations are used while capturing the videos. The third dataset Singapore-Marine-dataset is created by the authors using

Dilip K. Prasad is with the Rolls-Royce@NTU Corporate Lab, Singapore (e-mail: dilipprasad@gmail.com).
C. Krishna Prasath is with the Rolls-Royce@NTU Corporate Lab, Singapore.
Deepu Rajan and Chai Quek are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore.
Lily Rachmawati is with the Rolls-Royce plc, Singapore.
Eshan Rajabally is with the Rolls-Royce Derby, United Kingdom.

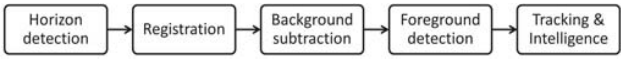

Fig. 1 Simple block diagram for maritime video processing

Canon 70D camera. Videos are acquired in two scenarios, namely at sea (videos captured on-board a vessel in motion) and on-shore (videos captured with camera on a stationary platform on-shore). The details of the datasets are presented in Table I.

II. Horizon detection

We represent horizon using two parameters, the vertical position Y of the center of the horizon from the upper edge of the image, and the angular position α made by the horizon with the horizontal axis. This is illustrated in Fig. 2. In the case of cameras mounted on mobile platform, the vertical and angular position is subject to large amount of motion, as noted in Table I. In Table I, E(Y) and E(α) represent the mean values of Y and α for a video. The ground truth for horizon is generated for each frame of these videos manually using independent volunteers [4].

We discuss two state-of-the-art methods [2], [5], which we succinctly refer to as FGSL (abbreviation derived from the first alphabets of the authors' names) [2] and ENIW (abbreviation derived from the first alphabets of the authors' names) [5], in the context of the present datasets. They are chosen as

TABLE I
Details of the Datasets Used in This Paper

Camera	At sea		On-shore	
Datasets	Buoy	Singapore-Marine-Dataset		Mar-DCT
Number of videos	10	11	28	9
Number of frames	998	2772	12604	7410
		Horizon related		
min(Y-E(Y)) (pixels)	-281.68	-436.30	-13.54	-52.32
max(Y-E(Y)) (pixels)	307.82	467.86	9.95	35.69
Std. dev. of Y (pixels)	107.98	145.10	1.52	9.98
min(α-E(α)) (degree)	-15.72	-26.34	-0.99	-1.25
max(α-E(α)) (degree)	20.72	12.99	0.51	1.75
Std. dev. of α (degree)	4.40	1.11	0.04	0.22
		Objects related		
Min number of objects	0	0	0	1
Max. number of objects	3	10	20	2

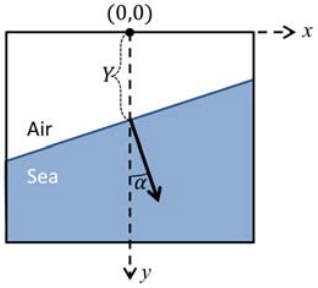

Fig. 2 Representation of horizon using Y and α

they both use a combination of two main approaches used for horizon detection, as discussed next.

One popular approach is to detect the most prominent line feature through parametric projection of edges in the image space to the parametric space of line features, such as Hough transform (HT). This approach assumes that horizon appears as a long line feature in the image. We note that this approach uses projective mappings and parametric space and is different from another line of research on line fitting on edge maps [6]-[8]. Although we do not exclude the utility of dominant point detection and line fitting [9], [10] for horizon detection in the on-board maritime detection problems, we note that no research work on horizon detection has so far employed these techniques.

The second popular approach is to select a candidate horizon solution that maximizes the statistical distances between the color distributions [11] of the two regions created by the candidate solution. This approach assumes that sea and sky regions have color distributions with large statistical distance between them and that the candidate solution separates the regions into sea and sky regions. While they are similar in using statistical distribution as the main criterion and using prominent linear features as candidate solutions, they are different in the choice of statistical distance measures.

The performance of these methods is presented in Table II. It is seen that the methods perform extremely well for Buoy dataset but perform poorly for the other datasets in terms of the vertical position of the horizon. In Fig. 3, we show that the assumption behind the statistical approach used by both methods may not apply. We present one image from each dataset (3rd row), the horizon ground truth (red solid line), the most prominent HT candidate (green dashed line), and the color distributions of the regions created by them in Fig. 3. For the first image, it is seen that the HT candidate for the horizon matches the ground truth and indeed the color distributions corresponding to the sea and sky regions match well. However, for the other three images, the Hough transform candidates do not match with the ground truth. Let us first consider the upper regions created by the ground truth and the Hough candidates. For the Singapore-Marine dataset, the upper region created by the Hough candidate includes the sky region and part of the sea region. This causes some change in the color distribution at lower color values. Nevertheless, the distribution is clearly dominated by sky and statistical distance metrics may not be effective in distinguishing sea and sky regions effectively. For example, the mean values (shown using vertical lines in the color distribution plots) of the distributions corresponding to the incorrect horizon are not significantly different from the mean values of the distributions corresponding to the ground truth. Numerically, the means show the same shift for all the color channels between the incorrect horizon and the ground truth. The maximum shift of 25 value (between 0 to 256 digital values) is observed for the third image for the sky region. The shift is caused by the inclusion of part of the sea in the upper region. In the other cases, the typical shift is 0 value to 5 values. The same observation applies to the example from Mar-DCT dataset as well, however with the shift observed in the bottom region.

TABLE II
STATISTICS OF ERRORS IN Y FOR DIFFERENT METHODS

	Buoy	On-board	On-shore	Mar-DCT
Error in Y				
25th percentile (1st quartile)				
ENIW	0.92	71.82	15.30	1.38
FGSL	0.72	72.06	7.30	4.29
50th percentile (median)				
ENIW	1.93	117.81	115.25	37.43
FGSL	1.59	118.14	115.25	198.58
Error in α				
25th percentile (1st quartile)				
ENIW	0.24	0.47	0.18	0.26
FGSL	0.20	0.49	0.18	0.64
50th percentile (median)				
ENIW	0.46	1.10	0.38	1.18
FGSL	0.38	1.19	0.35	1.00

Further, we note some frames from Singapore-Marine-dataset in Fig. 4, which are challenging due to reasons such as absence of line features of horizon, presence of competing line features (such as through ships and vegetation), adverse effects of conditions such as haze and glint, etc. For all these images, we show below them their edge maps where red edges are long edges and green edges are the edges of medium length. The dearth of line features representing horizon is evident in these edge maps. Also notable is that in conditions such as haze, the color distributions of sea and sky regions may be practically inseparable.

The statistical distance between sea and sky distributions may be increased by adding extra spectral channels [12] and abstract statistical distance metrics may be used through machine learning techniques [13], [14]. However, these approaches require sensor modification or their performance depends upon the diversity of the training dataset.

III. REGISTRATION

Registration refers to the situation where different frames in a scene correspond to the same physical scene with matching world coordinates. In marine scenario, especially for sensors mounted on sea vessels and buoys, the unpredictable motion of the sensors often result in a complicated registration problem where even the consecutive frames are not registered and may have a large angular and positional shift, as noted in Table I.

The angular difference between the two consecutive frames may have all the three angular components, viz. yaw, roll, and pitch. If the horizon is present, roll and pitch can be

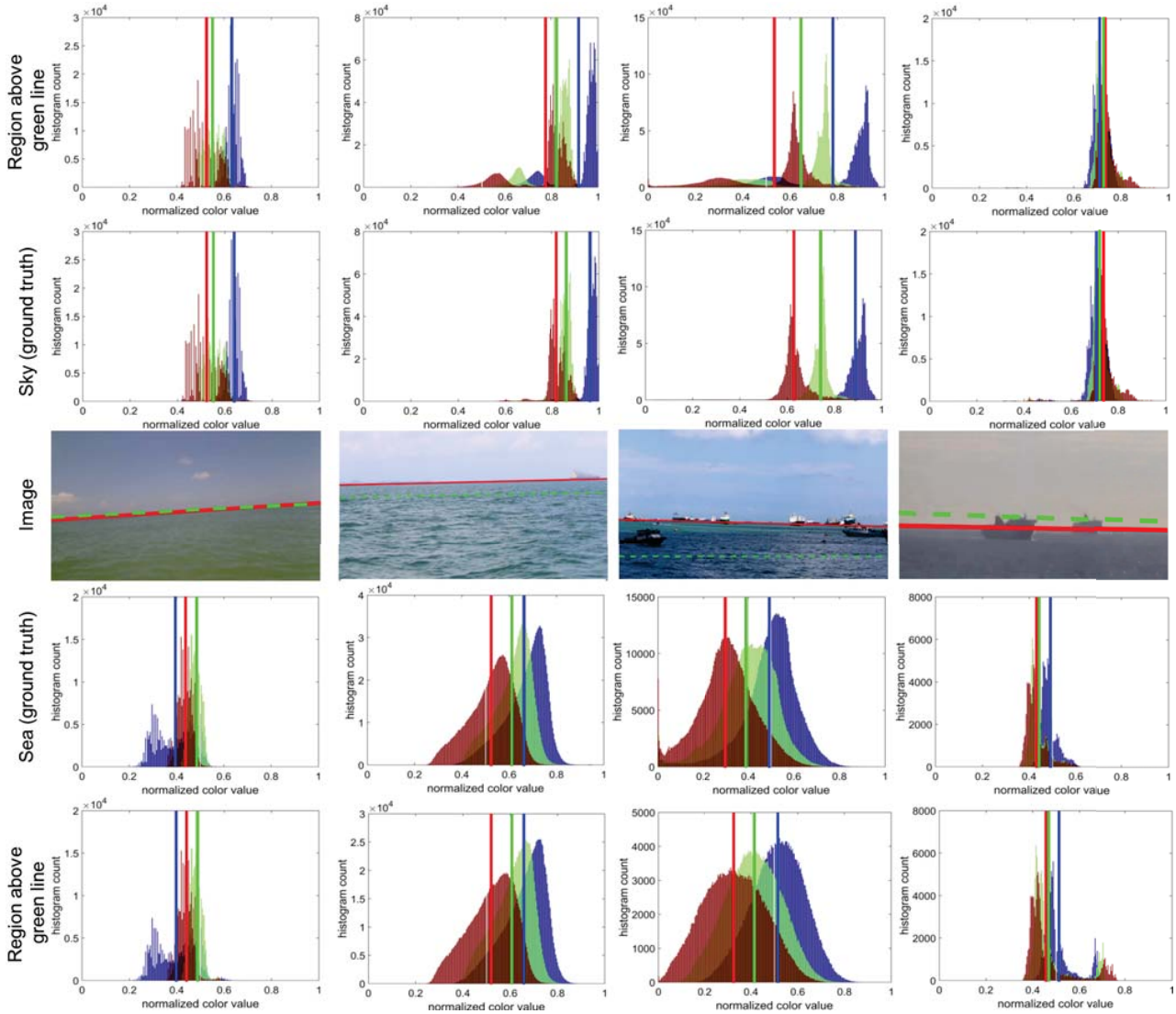

Fig. 3 Statistical distribution of the sea and sky regions determined by the horizon ground truth (solid red line) and the upper and lower regions determined by the most prominent HT candidate (green dashed lines)

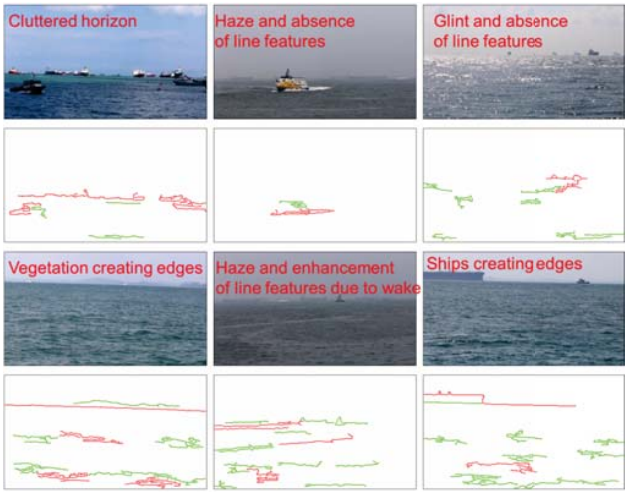

Fig. 4 Challenging situations in which horizon detection is challenging

significantly corrected for since they result in the change of angle and position of the horizon, respectively. However, yaw cannot be corrected for. This is illustrated using two consecutive frames from a video in the buoy dataset are used in Fig. 5. It is seen that horizon based registration does reduce the differences (see middle row, 3rd image) but the zoom-ins shown in the bottom row clearly indicate that the boat and cloud have unequal horizontal difference between them. In this scenario, it is impossible to say if the cloud was stationary and the boat moved, or the boat was stationary and the cloud moved, or both of them moved.

In order to correct for the yaw, we need some additional features that allow the detection of the horizontal staggering between two consecutive frames. The availability and possibility to detecting the stationary features is important for yaw correction. Buildings, landmarks, and terrain features may serve this purpose [15], if they are present in the scene. For

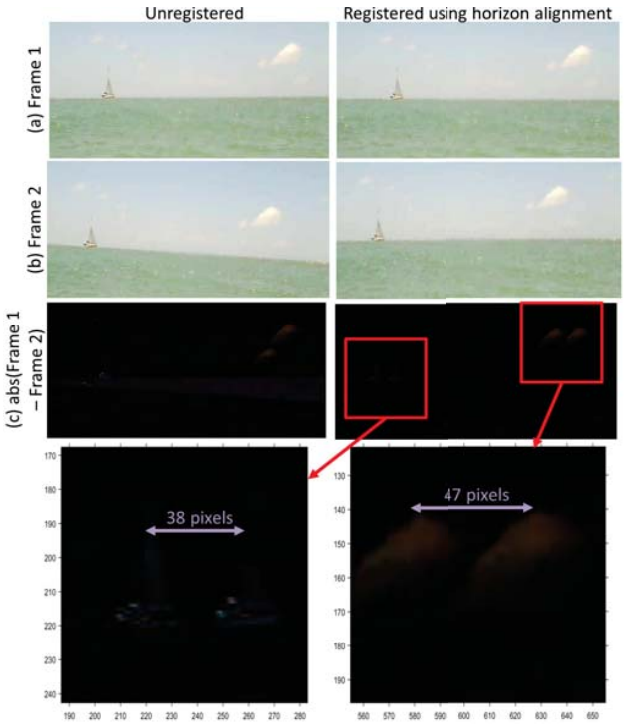

Fig. 5 The top row shows the original consecutive frames and their difference from a video. Results of registration using horizon are shown in the second row. The third row shows two insets from difference image of the registered image

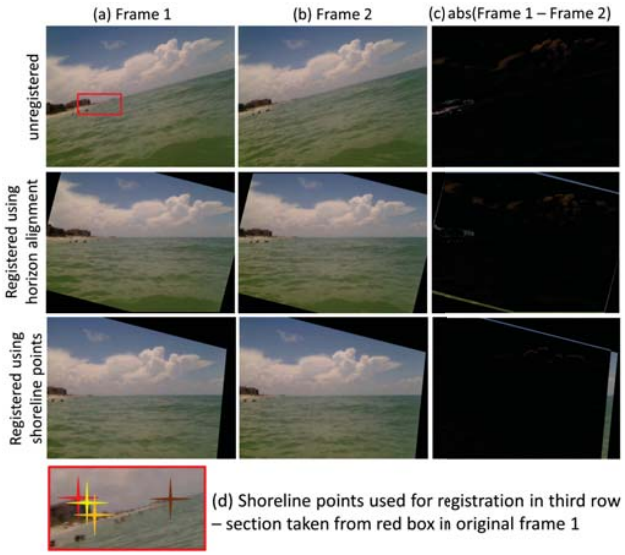

Fig. 6 The top row shows the original consecutive frames and their difference from a video. Results of registration using horizon are shown in the second row. Registration results using just four fixed points on the shoreline are shown in the third row. The four points used for registration are shown in the last row

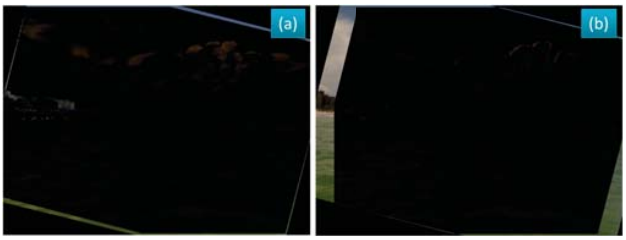

Fig. 7 Registration using cross-correlation of strip around the horizon. (a) The difference image obtained by registration using horizon only, reproduced from Fig. 6. (b) The difference image after horizontal shift of 48 pixels, identified as the peak of the cross-correlation function

example, we consider two consecutive frames in Fig. 6 taken from another video in buoy dataset which does have stationary features. The result of registration using horizon only is shown in the middle row. However, using just a few manually selected stationary points on the shoreline, accuracy in registration is significantly enhanced, as seen in the third row.

Notably, although a ship may be stationary and can be easily detected, it is difficult to conclude whether the ship is stationary or not. Also, it is discussed in [16] that the line features in a scene with moving vessels and absence of stationary cues may enable registration only if the vessels in the scene are not rotating. Thus, for a general maritime scenario, registration of frames is still a challenge. Strictly speaking, the best possible way of dealing with this scenario is the use of the ship's motion sensors and gyro sensors. Nevertheless, some help can be derived from texture-based features for registration across frames, assuming that the generalized shapes of texture boundaries might not change significantly over few consecutive frames [17]. Another related approach is used in [18] for registration, where a narrow horizontal strip is taken around the horizon in both the images and the shift at which the two images have maximum correlation is determined. This shift is used for registration. An example is shown in Fig. 7. Optical flows may also be useful [19], although at significant computation cost.

IV. Background Subtraction

There are several useful surveys on the topic of background suppression in video sequences [20]. Water background is more difficult than other stationary as well as dynamic backgrounds because of several reasons. One reason is that water background is continuously dynamic both in spatial and temporal dimensions due to waves, whereas the background subtraction methods typically address dynamic backgrounds that where dynamics are either spatially restricted (such as rustle of trees) or temporally restricted (such as a parked car). Second reason is that waves have a high spatio-temporal correlations [21] while the dynamic background subtraction methods implicitly infer high spatio-temporal correlations as patterned (i.e. non-random) movement of foreground objects. An associated difficulty in marine background detection is that the electro-optical sensor mounted on a mobile platform is subject to a lot of motion. Most background learning methods learn background by assuming that a pixel remains background or foreground for at least a certain period of time. Thus, background modelling depends upon the accuracy

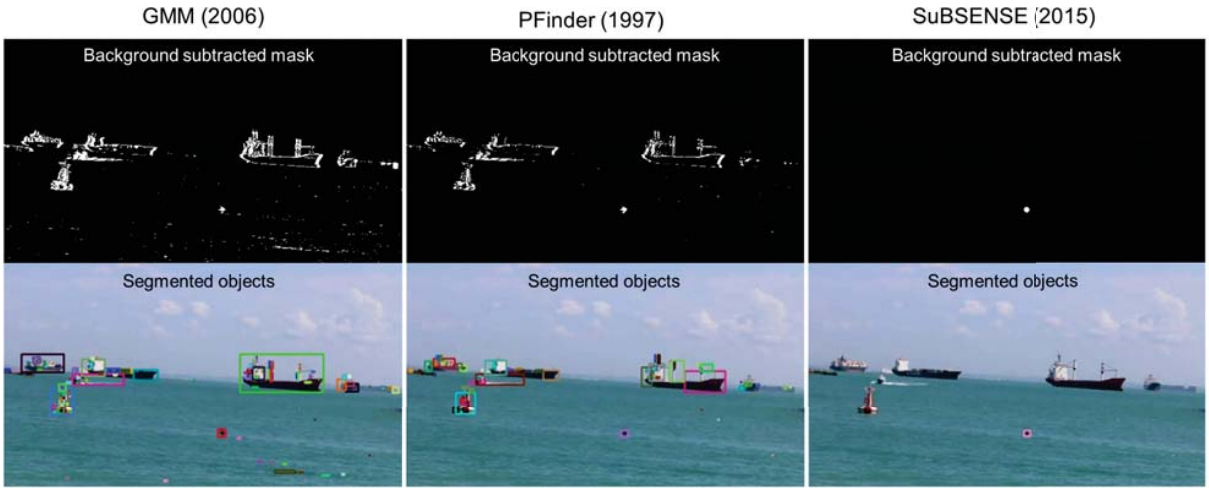

Fig. 8 Results of three methods from change detection competition that perform the best or fastest

of registration, which is a challenging problem as discussed in the previous section. Third reason is that wakes, foams, and speckle in water are inferred as foreground by typical background detection method whereas they are background in the context of maritime object detection problem.

To illustrate the need for new algorithms addressing maritime background, we applied the 34 algorithms that participated in the change detection competition [22]. This competition was conducted in 2014 as a part of a change detection workshop at a prestigious computer vision conference [23]. It used a dataset of 51 videos comprising of about 140,000 frames separated into 11 categories of background challenges such as dynamic background, camera jitter, intermittent object motion, shadows, infrared videos, snow, storm, fog, low frame rate, night videos, videos from pan-tilt-zoom camera, and air air turbulence. Since the dataset addressed several background challenges encountered in maritime videos as well and the submitted algorithms represented the state-of-the-art for these challenges, we tested their performance on Singapore-Marine-dataset.

Here, we show in Fig. 8 the result of three methods for one frame of a video from on-shore Singapore-Marine dataset. The three methods are Gaussian mixture model (GMM) [24], which models background's color distribution as mixture of Gaussian distributions, Gaussian background model of PFinder [25], which models the intensity at each background pixel as a single Gaussian function and then clusters these Gaussian functions as representing the background, and the self-balancing sensitivity segmenter (SuBSENSE) [26], which uses local binary similarity patterns at pixel levels for modeling background. It is seen that these methods are ineffective through producing false positives in the water region or through producing false negatives while suppressing water background.

V. Foreground Object Detection

Even with proper dynamic background subtraction, such that wakes, foams, clouds, etc. are suppressed, it is notable that further foreground segmentation can result in detection of mobile objects only. However, as noted in Table I, there are several stationary objects as well in the videos. In Table I, the ground truth for stationary and dynamic objects have been generated for each video manually by independent volunteers. The segmented background has to be further analysed for detecting the static foreground objects. Since the general dynamic background subtraction and foreground tracking problems do not require the detection of static objects, no integrated approaches exist that can simultaneously detect the stationary and mobile foreground objects. This is an open challenge for the maritime scenario. Research for the problem of object detection in images may be applied for detection of objects in individual images, thus catering for both static and mobile objects. However, the complicated maritime environment with potential of occlusion, orientation, scale, and variety of objects make it computationally challenging [27]. Further, complicated motion patterns imply that frame to frame matching of objects for tracking is challenging if detection is performed independently for each frame.

VI. Weather and Illumination Conditions

A maritime scene is subjected to a vast variety of weather and illumination conditions such as bright sunlight, twilight conditions, night, haze, rain, fog, etc. Further, the solar angles induce different speckle and glint conditions in the water. Tides also influence the dynamicity of water. The situations that affect the visibility influence the contrast, statistical distribution of sea and water, and visibility of far located objects. Effects such as speckle and glint create non-uniform background statistics which need extremely complicated modelling such that foreground is not detected as the background and vice versa. Also, the color gamuts for illumination conditions such as night (dominantly dark), sunset (dominantly yellow and red), and bright daylight (dominantly blue), and hazy conditions (dominantly gray) also vary significantly. As a consequence, the suitable methods and models for one weather and illumination condition is

not effective for other conditions. Seamless selection of approaches and transition between one approach to another with varying conditions is important for making maritime processing practically useful.

VII. Conclusion

As discussed above, maritime video processing problem poses challenges that are absent or less severe in other video processing applications. It needs unique solutions that address these challenges. It also needs algorithms with better adaptability to the various conditions encountered in maritime scenario. Thus, the field is rich with possibilities of innovation in maritime video processing technology. We hope that the discussion here motivates the researchers to pursue maritime video processing challenges with enthusiasm and vigour.

VIII. Funding Information

This work was conducted within Rolls Royce@NTU Corporate Lab with the support of National Research Foundation under the CorpLab@University scheme.

References

[1] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabaly, and C. Quek, "Video processing from electro-optical sensors for object detection and tracking in maritime environment: A Survey," *Intelligent Transportation Systems, IEEE Transactions on*, 2017.

[2] S. Fefilatyev, D. Goldgof, M. Shreve, and C. Lembke, "Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system," *Ocean Engineering*, vol. 54, pp. 1–12, 2012.

[3] D. D. Bloisi, L. Iocchi, A. Pennisi, and L. Tombolini, "ARGOS-Venice boat classification," in *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*, 2015, pp. 1–6.

[4] D. K. Prasad, D. Rajan, C. Prasath, L. Rachmawati, E. Rajabaly, and C. Quek, "MSCM-LiFe: multi-scale cross modal linear feature for horizon detection in maritime images," in *IEEE TENCON*, 2016.

[5] S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M. Waszak, "Vision-guided flight stability and control for micro air vehicles," *Advanced Robotics*, vol. 17, no. 7, pp. 617–640, 2003.

[6] D. K. Prasad, M. K. Leung, C. Quek, and S.-Y. Cho, "A novel framework for making dominant point detection methods non-parametric," *Image and Vision Computing*, vol. 30, no. 11, pp. 843–859, 2012.

[7] D. K. Prasad and M. K. Leung, *Polygonal representation of digital curves*. INTECH Open Access Publisher, 2012.

[8] D. K. Prasad, M. K. Leung, C. Quek, and M. S. Brown, "DEB: Definite error bounded tangent estimator for digital curves," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4297–4310, 2014.

[9] D. K. Prasad and M. S. Brown, "Online tracking of deformable objects under occlusion using dominant points," *JOSA A*, vol. 30, no. 8, pp. 1484–1491, 2013.

[10] D. K. Prasad, "Fabrication imperfection analysis and statistics generation using precision and reliability optimization method," *Optics express*, vol. 21, no. 15, pp. 17 602–17 614, 2013.

[11] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," *JOSA A*, vol. 31, no. 5, pp. 1049–1058, 2014.

[12] D. K. Prasad and L. Wenhe, "Metrics and statistics of frequency of occurrence of metamerism in consumer cameras for natural scenes," *JOSA A*, vol. 32, no. 7, pp. 1390–1402, 2015.

[13] S. Fefilatyev, V. Smarodzinava, L. O. Hall, and D. B. Goldgof, "Horizon detection using machine learning techniques," in *International Conference on Machine Learning and Applications*, 2006, pp. 17–21.

[14] D. K. Prasad and K. Agarwal, "Classification of hyperspectral or trichromatic measurements of ocean color data into spectral classes," *Sensors*, vol. 16, no. 3, p. 413, 2016.

[15] R. Behringer, "Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors," in *Virtual Reality*, 1999, pp. 244–251.

[16] X. Cao, Z. Rasheed, H. Liu, and N. Haering, "Automatic geo-registration of maritime video feeds," in *International Conference on Pattern Recognition*, 2008, pp. 1–4.

[17] A. Criminisi and A. Zisserman, "Shape from Texture: Homogeneity Revisited," in *British Machine Vision Conference*, 2000, pp. 1–10.

[18] S. Fefilatyev, "Algorithms for visual maritime surveillance with rapidly moving camera," Ph.D. dissertation, University of South Florida, 2012.

[19] D. Dusha, W. Boles, and R. Walker, "Attitude estimation for a fixed-wing aircraft using horizon detection and optical flow," in *Digital Image Computing Techniques and Applications*, 2007, pp. 485–492.

[20] S. Y. Elhabian, K. M. El-Sayed, and S. H. Ahmed, "Moving object detection in spatial domain using background removal techniques-state-of-art," *Recent patents on computer science*, vol. 1, no. 1, pp. 32–54, 2008.

[21] V. Ablavsky, "Background models for tracking objects in water," in *International Conference on Image Processing*, vol. 3, 2003, pp. III–125.

[22] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.

[23] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "Cdnet 2014: an expanded change detection benchmark dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 387–394.

[24] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.

[25] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.

[26] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Subsense: A universal change detection method with local adaptive sensitivity," *Image Processing, IEEE Transactions on*, vol. 24, no. 1, pp. 359–373, 2015.

[27] D. Bloisi, L. Iocchi, M. Fiorini, and G. Graziano, "Automatic maritime surveillance with visual target detection," in *Proc. of the International Defense and Homeland Security Simulation Workshop*, 2011, pp. 141–145.