# Video Processing From Electro-Optical Sensors for Object Detection and Tracking in a Maritime Environment: A Survey

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

***Abstract*— We present a survey on maritime object detection and tracking approaches, which are essential for the development of a navigational system for autonomous ships. The electro-optical (EO) sensor considered here is a video camera that operates in the visible or the infrared spectra, which conventionally complements radar and sonar for situational awareness at sea and has demonstrated its effectiveness over the last few years. This paper provides a comprehensive overview of various approaches of video processing for object detection and tracking in the maritime environment. We follow an approach-based taxonomy wherein the advantages and limitations of each approach are compared. The object detection system consists of the following modules: horizon detection, static background subtraction, and foreground segmentation. Each of these has been studied extensively in maritime situations and has been shown to be challenging due to the presence of background motion especially due to waves and wakes. The key processes involved in object tracking include video frame registration, dynamic background subtraction, and the object tracking algorithm itself. The challenges for robust tracking arise due to camera motion, dynamic background, and low contrast of tracked object, possibly due to environmental degradation. The survey also discusses multisensor approaches and commercial maritime systems that use EO sensors. The survey also highlights methods from computer vision research, which hold promise to perform well in maritime EO data processing. Performance of several maritime and computer vision techniques is evaluated on Singapore Maritime Dataset.**

***Index Terms*— Maritime vehicles, maritime navigation, autonomous automobiles, video signal processing, computer vision.**

## I. Introduction

MARITIME surveillance is a critical part of law enforcement and environment protection for littoral nations. However, with the growth of commercial ocean liners and other seafaring vessels such as cruise ships, technologies that have been traditionally deployed for military purposes, e.g. radars and sonars, are found to be of immense utility in providing support for navigation as well. The International Regulations for Preventing Collisions at Sea 1972 (COLREGs) requires all ships to be equipped with radars for proper lookout to provide early warning of potential collision. However, radar measurements are sensitive to the meteorological condition and the shape, size, and material of the targets. Thus, radar data has to be supplemented by other situational awareness sensors for better collision avoidance and navigation.

Situational awareness at sea would undergo a paradigm shift with future development of the autonomous ship equipped with numerous sensors to support advanced decision and remote operation [1]. Autonomy in ship navigation would lead to reduction in crew numbers as a result of re-skilling and relocation of crew to the shore, potentially resulting in less vigilant look-out. It is imperative that ranging devices are augmented with other sensors so that fail-safe decisions can be rapidly taken with high level of confidence.

Electro-optical (EO) sensors are primed to complement ranging devices. In this paper, EO sensors imply video cameras operating in the visible and infrared portions of the electromagnetic spectrum. Some works [12], [43] even recommend them as a replacement to ranging devices in special circumstances such as populated urban maritime scenario. EO sensors are of interest for two major reasons. Firstly, the image streams generated by them are directly interpretable and intuitive for human operators, alleviating the need for specialized training. Secondly, the image streams from them are amenable to image processing and computer vision such that advanced intelligence can be generated computationally without significant human intervention. Visible range EO sensors benefit from the availability of color data and high quality optics. On the other hand, infrared EO sensors benefit from night time visibility and suppression of highly dynamic regions in the video. This helps in the development of robust video processing algorithms [44].

However, there are some disadvantages associated with EO data [45]. Although the atmospheric propagation characteristics for long wave infrared spectrum are superior to other visible and infrared frequencies [46], in general, the atmospheric propagation losses restrict the range of the EO sensors to only a few kilometers. Further, EO data processing for automatic intelligence generation is quite challenging for maritime environment. Some of the challenges are:

Manuscript received April 20, 2016; revised September 22, 2016; accepted November 16, 2016. Date of publication January 10, 2017; date of current version July 31, 2017. This work was supported by the Rolls Royce@NTU Corporate Laboratory within the National Research Foundation under the CorpLab@University scheme. The Associate Editor for this paper was K. Wang.

D. K. Prasad is with the Rolls-Royce@ NTU Corporate Laboratory, 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.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2634580

TABLE I
COMPARISON OF SENSORS USED IN MARITIME SCENARIO FOR SITUATION AWARENESS

| Sensor | Distance | Advantages/Characteristics | Disadvantages |
|---|---|---|---|
| Sonar [2]−[5] | ∼ 1 km to few 100 km | ⊙ Long range sensing ability<br>⊙ Underwater detection<br>⊙ Detects objects with large acoustic signatures (ex. whales and icebergs) | ⊗ Needs separate systems for small range detections<br>⊗ Performs poorly for objects with small acoustic signatures (ex. growler, small boats, and debris)<br>⊗ Requires specialized user training |
| Radar [6]−[10] | ∼ 1 km to few 100 kms | ⊙ Long range sensing ability<br>⊙ Detects objects with high radar cross-sections (mostly metallic)<br>⊙ Large on-board power supply requirement | ⊗ Suffers from minimum range<br>⊗ Cannot penetrate water<br>⊗ Cannot detect big objects with small radar cross-section [11]<br>⊗ Requires specialized user training |
| Visible range electro-optical [11]− [30] | ∼ m to ∼ km | ⊙ Processes color information<br>⊙ High resolution, advanced optics available<br>⊙ Adaptive to new technology<br>⊙ Uses image processing/computer vision algorithms<br>⊙ Naturally intuitive, no need of user training | ⊗ Sensitive to illumination and weather changes<br>⊗ Not suitable for night vision<br>⊗ Computation intensive<br>⊗ Low range sensing due to atmospheric attenuation<br>⊗ Difficult to detect far objects and predict their size and distance<br>⊗ Difficult to model water dynamics, wakes, and foam |
| Infrared range electro-optical [28]−[42] | ∼ m to ∼ km | ⊙ Longer range than visible range EO<br>⊙ Allows night vision<br>⊙ Water appears less dynamic<br>⊙ Intuitive, no need of user training<br>⊙ Adaptive to new technology<br>⊙ Uses image processing/computer vision algorithms | ⊗ Significantly poorer optics available<br>⊗ Saturated images in day time<br>⊗ Sensitive to illumination and weather changes<br>⊗ Computation intensive<br>⊗ Difficult to detect far objects and predict their size and distance<br>⊗ Horizon not-well defined in IR images |

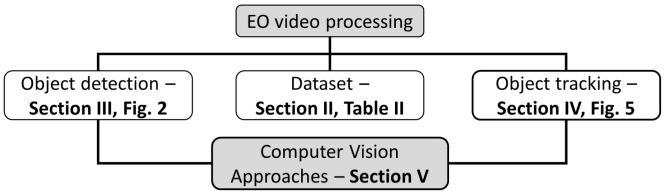


Fig. 1. Organization of the survey.

- the difficulty in modeling the dynamics of water (including waves, wakes and foams) for background subtraction and detection of foreground objects,
- variations in object appearances due to distance and angle of viewing, and
- changes in illumination and weather conditions, such as due to clouds, sunshine, rain, glint, etc.

This paper presents a taxonomic survey of the approaches for processing EO data acquired from maritime environment. The organization of this survey is given in Fig. 1. Table I outlines the advantages and disadvantages of sonar, radar, and EO sensors. The survey focuses on maritime object detection and tracking using EO data to fulfil the navigational needs of an autonomous ship. The EO data is assumed to be available in the form of a video, either in the visible spectrum or in the infrared range. We exclude special cameras such as for monocular or stereovision from this survey. Survey on monocular and stereovision can be found in [47], [48]. Further, we exclude device-level signal processing and high-level intelligence generation (such as vehicle behavior [49]). We discuss post processing of the tracking data, maritime multi-sensor approaches, and commercial maritime systems that use EO sensors in Appendix.

## II. MARITIME DATASET FOR COMPARATIVE EVALUATION

Works in maritime image processing typically use military owned or proprietary datasets which are not made available for research purposes. The authors are aware of only one dataset MarDCT [1] that is available online for academic and research purposes. Although this dataset does have images and videos acquired from both visible range and infrared range sensors, they are either in urban navigation scenario atypical of the usual maritime scenario or consider very simple scenarios with only one or two maritime vessels close to horizon. There is a pressing need for a benchmark dataset of maritime videos so that quantitative comparison of various algorithms can be performed. To this end, we have created Singapore Maritime Dataset, using Canon 70D cameras around Singapore waters. All the videos are acquired in high definition (1080 × 1920 pixels). We divide the dataset into parts, on-shore videos (visible and near-infra red) and on-board videos, which are acquired by camera placed on-shore on fixed platform and camera placed on-board a moving vessel, respectively. Annotation tools developed in Matlab were used by volunteers not related to the project for annotation of ground truths (GTs) of horizon and objects in each frame. The dataset and annotation files of the GTs for horizon, objects, and tracks are available at the project webpage.[2] Details of the dataset are given in Table II.

## III. OBJECT DETECTION

For object detection in maritime EO data processing, each frame of the EO video stream is considered independently without taking temporal information into account. The general framework of object detection approaches in maritime scenarios is shown in Fig. 2. It consists of three main steps, viz., horizon detection, background subtraction, and foreground segmentation, discussed in the following subsections.

### A. Horizon Detection

There are three main approaches for horizon detection − projection based, region based, and hybrid approach.

[1] http://www.dis.uniroma1.it/ labrococo/MAR/
[2] https://sites.google.com/site/dilipprasad/home/singapore-maritime-dataset

TABLE II
DETAILS OF THE SINGAPORE MARITIME DATASET

| | On-board videos | On-shore videos |
|---|---|---|
| Number of videos | 4 | 32 |
| Total number of frames | 1196 | 16254 |
| Number of frames in a video | 299 | $\in [206, 995]$ |
| Size of frames (pixels) | $1920 \times 1080$ | $1920 \times 1080$ |
| **Horizon and registration related** | | |
| $Y$ (pixels) | $\in [190.6, 1077.1]$ | $\in [283.2, 925.6]$ |
| Mean($Y$) $\pm$ standard deviation($Y$) (pixels) | $552.5 \pm 183.9$ | $530.8 \pm 107.1$ |
| $\alpha$ ($^\circ$) | $\in [-27.13, 0.40]$ | $\in [3.36, 8.43]$ |
| Mean($\alpha$) $\pm$ standard deviation($\alpha$) ($^\circ$) | $-7.18 \pm 5.80$ | $6.34 \pm 1.00$ |
| **Object detection and tracking related** | | |
| Number of objects per frame | | $\in [2, 20]$ |
| Number of stationary objects per frame | | $\in [0, 14]$ |
| Number of moving objects per frame | | $\in [0, 10]$ |
| Total number of object annotations in a video | | 192980 |
| Total number of stationary objects in a video | | 137485 |
| Total number of moving objects in a video | | 55495 |
| Number of tracks in a video | | $\in [4, 19]$ |
| Temporal length of tracks (in frames) | | $\in [19, 600]$ |

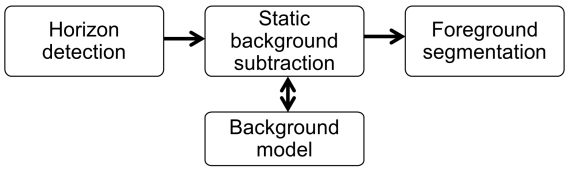


Fig. 2. General pipeline of maritime EO data processing for object detection.

Fig. 3(a) shows three examples of maritime images with horizon. Image 1 has two very low contrast targets close to a blurry horizon. Image 2 has horizon characterized by good contrast between the sky and water. Image 3 does not have a well defined horizon although the presence of the skyline may be a useful cue for its detection. However, the image suffers from false horizontal line created by the wakes of two targets, which is likely to be confused as horizon. Due to a variety of available cues as well as challenges, we use these images as examples to demonstrate the strengths and weaknesses of different horizon detection methods.

*1) Projections From Edge Map:* In these methods, first the edge map of the image is computed using edge detectors [50]. It is then projected to another space where prominent line features in the edge map can be identified easily. Typically, the Hough and Radon transforms are used for such projections. Given the equation of a line:

$$x\cos(\theta) + y\sin(\theta) = \rho \tag{1}$$

Each edge pixel with coordinates $(x, y)$ is transformed into a curve in the Hough space $(\theta, \rho)$ using the projection [50]:

$$H(\theta, \rho) = \int\int_{x,y} \Big(1 - \delta\big(I(x, y)\big)\Big)\delta(x\cos\theta + y\sin\theta - \rho)dxdy \tag{2}$$

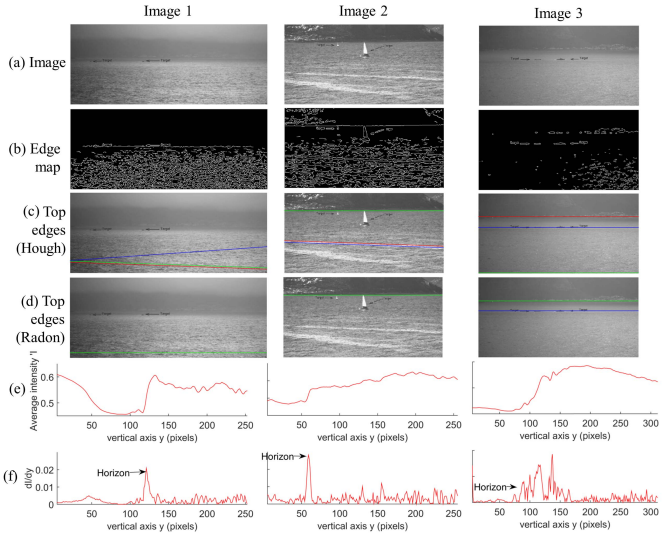


Fig. 3. Three example images (row a) from [11] and their edge maps (row b) are used for studying the problem of horizon detection. The top 3 candidates, with largest strengths in Hough and Radon spaces are shown in rows (c,d) using colored lines. Average intensity profiles in the vertical direction are shown in (e). The gradients of intensity profiles in (e) are shown in (f).

where $\delta$ represents the Dirac delta function and $I(x, y)$ is the edge map. This is analogous to computing the 2D histograms of $(\theta, \rho)$. Cells in the $(\theta, \rho)$ histogram corresponding to few largest values of $H(\theta, \rho)$ determines the parameters of the line.

The transformation into Radon space is achieved by [50]:

$$R(\theta, \rho) = \int\int_{x,y} I(x, y)\delta(x\cos(\theta) + y\sin(\theta) - \rho)dxdy \tag{3}$$

Similar to the Hough space, cells in $(\theta, \rho)$ with the highest number of entries in $R(\theta, \rho)$ are the parameters of the line.

While the simplicity of these approaches makes them popular, projective transforms are sensitive to preprocessing such as histogram equalization and filtering before the extraction of the edge map [50], [51]. Further, they can detect the horizon only if it appears as a prominent line feature in the edge map. Thus, as shown in Fig. 3(c,d), both Hough and Radon transforms perform poorly for image 1 but detect the horizon in image 2. Although the edge map of image 3 (Fig. 3(b)) does not have significant features corresponding to the horizon, the city skyline provides sufficient number of edge pixels parallel and close to the horizon, enabling a rough detection of horizon. Notably, the wake creates a dark horizontal stripe close to the targets in image 3 which causes detection of the line corresponding to the wakes as well.

*2) Region Based Horizon Detection:* The intensity variations in the region of the horizon are higher compared to sky or sea regions alone. In Fig. 3(e,f), the mean intensity along the vertical axis and its gradient are plotted for images 1−3. The regions of horizon are characterized by significantly large intensity changes in each of the three images, although the intensity gradient itself is not sufficiently conclusive of the horizon in image 3. Such localized intensity characteristics

TABLE III
HORIZON DETECTION APPROACHES

| | Methods | Advantages | Disadvantages |
|---|---|---|---|
| Projection from edge map [14], [26], [38] | Radon transform, Hough transform | ⊙ Simple<br>⊙ Mathematically well-defined | ⊗ Sensitive to preprocessing<br>⊗ Work for prominent well-defined linear horizon only<br>⊗ Horizon may not be the most prominent |
| Region based [17], [29], [31], [41] | Median, correlation, covariance | ⊙ Work for blurred horizon as well<br>⊙ Suitable for IR images | ⊗ Requires statistical apriori knowledge<br>⊗ Based on statistics |
| Hybrid [15], [31], [59] | | ⊙ More accurate<br>⊙ More robust to low-contrast images | ⊗ More complex<br>⊗ More computation intensive |

are used for detecting horizon, especially in unmanned aerial vehicles [52]–[54]. Quite often, the pixels in an image are classified as belonging to sky and sea (or ground) [29].

This is a three step procedure. The first step is to use a local smoothing operator, such as top-hat filter [27], median filter [28], mean filter [41], Gaussian filter [55], or standard deviation filter [41]. The second step is to approximate these local statistics with sum of Gaussian functions or polynomial functions [17], [32], where each function represents distribution of one region, such as the sea region or the sky region. More complex representations of the regions, such as linear discriminant analysis [56], textures [56], covariances [28], [57], and eigenvalues [57], may be used. In the last step, the boundary of two classified regions is identified as horizon. We note that region based techniques inherently assume apriori information, such as suitable statistical representations or machine learning of the trend of intensities at the horizon.

Instead of the second and third steps, Bouma et. al [31] used high intensity gradient to conclude the common boundary of sea-sky regions and used it as horizon. A more robust version of this approach employed multi-scale approach [58].

*3) Hybrid Methods:* The above methods are ineffective for Image 3 in Fig. 3. In such cases, hybrid methods are useful. In the hybrid approach of [60], for each candidate generated by a projection-based method, the regions above and below the candidate line were considered as hypothetical sky and sea regions, and their statistical distributions were computed. The candidate that gives maximum value of the Mahalanobis distance between the distributions of the hypothetical sea and sky regions is chosen as horizon. In [15], local statistical features were used, but explicit representations of sea and sky regions were not employed. Then, using a set of training images and machine learning techniques, features representing horizon were learnt directly. Recent algorithms have combined multi-scale filtering and projection based approaches for providing state-of-the-art results [61], [62].

*4) Comparison of Methods for Horizon Detection:* A qualitative comparison of the methods is provided in Table III. For quantitative comparison on Singapore Maritime dataset, we use the representation of horizon as shown in Fig. 4(a). $Y$ is the distance between the center of the horizon and the upper edge of the frame. $\alpha$ is the angle between the normal to the horizon and the vertical axis of the frame. The ranges and standard deviations of $Y$ and $\alpha$ are quite large for on-board videos (see Table II), making them significantly more challenging than on-shore videos. On-board videos are challenging due to presence of land features very close to horizon, while on-shore videos are also challenging because of the occlusion of horizon by vessels and the presence of wakes in the foreground. Frames from both types of videos are shown in Fig. 4(b,c).

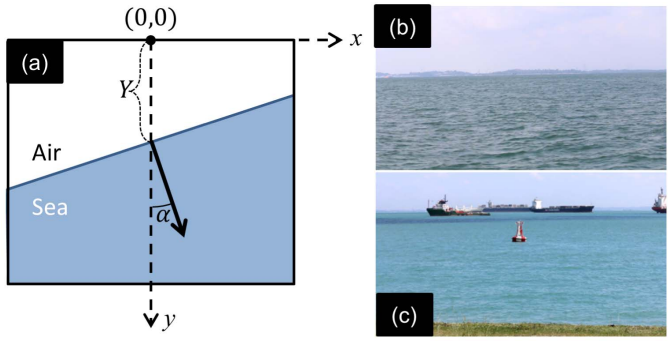


Fig. 4. Representation of horizon for quantitative comparison of horizon detection approaches is shown in (a). Representative frames from on-board and on-shore videos are shown in (b,c) respectively.

TABLE IV
QUANTITATIVE COMPARISON OF METHODS FOR HORIZON DETECTION. THE SMALLEST ERROR IN EACH COLUMN IS INDICATED IN BOLD

| | Position error (pixels) $\lvert Y_{GT} - Y_{est} \rvert$ | | | | Angular error (°) $\lvert \alpha_{GT} - \alpha_{est} \rvert$ | | | | Time/ frame (s) |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Q25 | Q50 | Q75 | Mean | Q25 | Q50 | Q75 | |
| On-board videos | | | | | | | | | |
| Hough | 219 | 131 | 220 | 295 | 2.6 | 0.6 | 1.7 | 3.4 | **0.3** |
| Radon | 372 | 213 | 362 | 517 | 40.6 | 1.5 | 3.4 | 87.7 | 2.7 |
| MuSMF | 269 | 156 | 283 | 379 | **1.8** | **0.5** | **1.2** | **2.5** | 0.9 |
| ENIW | **120** | **63** | **116** | 166 | 1.9 | **0.5** | **1.2** | **2.5** | hours |
| FGSL | **120** | **63** | 117 | **165** | **1.8** | **0.5** | **1.2** | **2.5** | 12.8 |
| On-shore videos | | | | | | | | | |
| Hough | 208 | 26 | 194 | 354 | **1.2** | 0.2 | 0.7 | 1.5 | **0.2** |
| Radon | 313 | 28 | 359 | 549 | 32.9 | 0.2 | **0.4** | 88.1 | 2.0 |
| MuSMF | **60** | 25 | **49** | **85** | **1.2** | 0.2 | **0.4** | **1.1** | 0.9 |
| ENIW | 121 | 15 | 94 | 163 | **1.2** | 0.2 | **0.4** | 1.3 | hours |
| FGSL | 112 | **12** | 91 | 162 | **1.2** | 0.2 | **0.4** | **1.1** | 12.3 |

We use position error $|Y_{GT} - Y_{est}|$ and angular error $|\alpha_{GT} - \alpha_{est}|$ as performance metrics for horizon detection. We provide comparison of Hough transform [28] (referred to as Hough), Radon transform [50] (Radon), mulit-scale median filter [31] (MuSMF), Ettinger et. al's method [53] (ENIW), and Fefilatyev et. al's method [60] (FGSL). Hough and Radon are projection-based, MuSMF and ENIW are region-based, and FGSL is a hybrid method. We have implemented MuSMF, ENIW, and FGSL since their codes are not available.

The comparison results are given in Table IV. It is notable that the error in $Y$ is more severe for all the methods as

compared to the error in $\alpha$. Projection based methods show poorest performance and statistical methods perform better. MuSMF performs the best for the on-shore videos , FGSL which uses both Hough transform and statistical distance measure for identifying the horizon performs the best for the on-board videos. MuSMF is parallelizable, with a possibility of making it about 10 times faster.

### *B. Static Background Subtraction*

There is a large corpus of works related to background subtraction that originated from the computer vision community (see [63] for a review). Maritime background subtraction can be considered under two scenarios: open seas and close to port/harbor. In the former, the challenge arises due to the dynamic nature of the water background in the form of waves, wakes, and debris. In the latter, static structures such as buildings or stationary vessels pose a challenge. The current literature in maritime background subtraction almost exclusively deals with the case of open seas.

The challenge of dynamic background is largely alleviated if long-wave infrared sensor, such as forward looking infrared (FLIR), is used because it maps the temperature of water, which is relatively uniform despite the dynamicity of water. This suppression of dynamicity occurs to a smaller extent in near-infrared and mid-wave infrared wavelengths [44]. Thus, static background subtraction techniques have better performance in long-wave infrared regime than in the visible spectrum. In the following, we describe the various background subtraction techniques based on background models and their corresponding learning strategies.

*1) Image Statistics:* Methods in this category use statistical information in a single infrared image. One of the earlier methods is by Bhanu and Holben [15], which modeled an image in terms of gray scale intensity and edge magnitude with the aim of segmenting the image into foreground and background using a relaxation function that gives a low value for background and a high value for the foreground. Similar idea was employed in [68], [69], where, instead of the relaxation function used in [29], confidence maps [68] and chi-squared measure of similarity [69] were used to segment the image into background and foreground regions.

Smith and Teal [37] compared the histogram of gray level intensities in a pixel's vicinity with a histogram of intensities of a reference background. If the histograms were similar, then the pixel was assigned to background. The reference background was obtained from the image itself by computing standard deviation over a $3 \times 3$ window at each pixel and assigning the pixels with less standard deviation as the background. The method fails if the reference background is computed incorrectly or if the background has wakes and debris such that their histograms may not correlate with the reference background histogram.

Van den Broek et. al [32] used horizon to determine the sea and sky regions. It was assumed that intensities in the sea region may vary perpendicular but not parallel to the horizon. Thus, mean and standard deviations of the intensities along thin strips parallel to the horizon were computed and polynomial functions fit upon the mean and standard deviations of the strips. Similar polynomial fitting was performed for the sky as well. These polynomial models for sea and sky were then used to compute a background map and subtract it from the image. This approach removed only low spatial frequency component from the image. Although a more robust approach proposed in [70] was used by [31] for visible range images, natural high spatial frequency components such as due to waves, sun, and clouds were still retained. Gal [71] used a co-occurrence matrix approach to learn sea and sky patterns which were subtracted from the original image. Fefilatyev et. al [16] used Gaussian low pass filtering in a narrow strip below horizon, followed by a color gradient filter to obtain regions of high color variations and finally applied a threshold computed using Otsu's method [72] to obtain the background.

Chen at. al [35] considered suppressing repeated spatial patterns by suppressing peaks in the Fourier transform of the image. Spatio-spectral residue and phase map of Fourier transform of eigenvectors representing 80% of input image were used in [20] for background suppression. Multi-scale approaches, combined with low-pass filter extracting low spatial frequencies [73], [74], which are representations of background, have also been found useful. For example, top-hat convolution filter, which is a low-pass filter, used in a multi-scale approach was shown to be effective in wake suppression [27]. Multi-scale spatio-spectral residue was used in [64]. Wang and Zhang [41] used multilevel filter and recursive Otsu approach [72] to detect and segment very small and either dark or bright targets from images with complex background.

*2) Gaussian Mixture Model (GMM):* Although wakes and foam appear distinct in the visible range images, they are not entirely suppressed even in infrared images. Thus, the histograms of both visible and infrared images are invariably multimodal. Gaussian mixture models (GMM) [75], [76] are suitable for representing multimodal backgrounds [12], [24], [60], [65], [77]. If a pixel belongs to the background, the probability $P(I)$ of observing an intensity $I$ at that pixel is given as:

$$P(I) = \sum_i w_i G(\mu_i, \sigma_i) \tag{4}$$

where $G(\mu_i, \sigma_i)$ represents $i^{\text{th}}$ Gaussian distribution with mean $\mu_i$ and standard deviation $\sigma_i$, and $w_i$ is the weight of the $i^{\text{th}}$ Gaussian distribution. Fefilatyev et. al [60] represented sky and sea regions as two Gaussian distributions (each being trivariate due to the red, green, and blue color channels) fitted with maximum possible separation between their means. In [24], if the distance of the test pixel's intensity from the mean of the closest Gaussian distribution was within 2.5 times its standard deviation, then the pixel was classified as background.

*3) Bayes Classifier:* Socek et. al [21] used Bayes classifier approach of [78] for background estimation and suppression. Given the feature vector at a test pixel $\bar{v}(p)$, if $P\big(p \in \mathbf{B}|\bar{v}(\mathbf{p})\big) > \mathbf{P}\big(p \in \mathbf{F}|\bar{v}(\mathbf{p})\big)$, where $\mathbf{B}$ and $\mathbf{F}$ indicate the background and the foreground, then the test pixel was classified as the background. The likelihoods $P\big(\bar{v}(p)|p \in \mathbf{B}\big)$

TABLE V
SUMMARY OF STATIC BACKGROUND SUBTRACTION APPROACHES USED IN OBJECT DETECTION

| Approach | Model | Learning | Advantages | Disadvantages |
|---|---|---|---|---|
| Single image statistics [20], [27], [29], [31], [32], [35], [37], [41], [64], [53] | Histogram correlation, polynomial functions fitting to strips parallel to horizon, spatial co-occurrence of intensities, spatial filtering | No temporal learning, only spatial patches/strips are used, initial background typically learnt as pixels using spatial standard deviations | Simple, no learning involved, no need of memory | Cannot deal with multi-modal approaches, does not use any form of temporal information |
| Gaussian mixture model (GMM) [24], [60], [65], [66] | Probability of intensity at a background pixel is a combination of Gaussian functions | Supervised learning using background labelled images or videos, adaptive learning can be used to update GMM | Adapts for multi-modal background and illumination changes | Requires supervised learning using a suitable dataset, adaptive learning may be complicated |
| Bayes classifier [21] | Compute the Bayes conditional probabilities of the pixel being background (or foreground) given an observed feature vector | Supervised learning through a suitable training dataset | Classification is simple | Learning is complicated and sensitive to the training dataset |
| Feature based background classifier [67] | Compute the feature attributes for every pixel and determine distance from pre-learnt class features | Supervised learning of class features using a suitable training dataset with both positive and negative samples | Robust due to multiple attributes class representation | Computation intensive learning, testing more complex than other methods above, multi-class may represent wakes, foam, clouds, etc. |

and $P\left(\bar{v}(p)|p \in \mathbf{F}\right)$ were determined through the histogram of the feature vector, learnt apriori. The supervised learning strategy enforced that a certain percentage of pixels should be classified as background using another background estimation method. It was found that the segmented frames contained too many noise-related and scattered pixels, which may be separately classified as outliers and removed, however at the expense of missing small objects that are few pixels wide.

*4) Feature Based Background Classifier:* In [67], both foreground and background objects were considered as belonging to different known classes, viz., clouds, islands, coastlines, oceanic waves, and ships. It used a total of seven types of features, viz., shape compactness, shape convexity, shape rectangularity or eccentricity, shape moment invariants, wavelet-based features, multiple Gaussian difference features, and local multiple patterns (discussed more in section V-A.7). It also considered what combinations of features were suitable for improving the detection accuracy of the different subclasses.

### C. Foreground Segmentation

In traditional maritime data processing, applying morphological operations such as identifying closed boundaries after background subtraction were considered sufficient for foreground segmentation [21], [23], [79] and the segmented contours were used as detected objects. Useful morphological operations for maritime human rescue problem have been adapted in [79] from [80], [81]. Several interesting edge based morphological segmentation techniques for detecting objects have been discussed in [29]. All are based on considering object segmentation as a two-class gray level problem in which objects belong to one set of gray levels and the background to the other. A further constraint is that the gradient inside an instance of each class is close to zero and gradients are high only along the edges. The underlying assumption in all morphological segmentation approaches is that the objects are not be occluded and are separate enough such that their boundaries may not merge.

### D. Comparison of Static Background Subtraction Techniques

A qualitative comparison is given in Table V. Here, we present quantitative comparison of a few static background subtraction techniques. The foreground is morphologically obtained after static background subtraction and enclosed in bounding boxes. They are compared against the bounding boxes of the objects annotated as ground truth. The performance is evaluated using intersection over union (IOU) ratio of the bounding boxes, defined as

$$\text{IOU}\left(O_i^{\text{GT}}, O_j^{\text{det}}\right) = \frac{\text{Area}\left(O_i^{\text{GT}} \cap O_j^{\text{det}}\right)}{\text{Area}\left(O_i^{\text{GT}} \cup O_j^{\text{det}}\right)} \tag{5}$$

where $O_i^{\text{GT}}$ and $O_j^{\text{det}}$ are the bounding boxes of $i$th ground truth (GT) object and the $j$th detected object and Area denotes the number of pixels. If more than one detected objects overlap with a GT object, the detected object with maximum overlap with the GT object is considered associated with the GT and dropped from further associations. The unassociated objects or associated objects with IOU less than 0.5 (based on [82]) are labelled as false positives (FPs). The remaining associated objects are true positives (TPs). The GT objects that are not associated to the any detected objects are labelled false negatives (FNs). $N_{\text{TP}}$ is the number of TPs in the video, analogously for $N_{\text{FP}}$ and $N_{\text{FN}}$. Precision and recall are computed as

$$\text{Precision} = N_{\text{TP}}/(N_{\text{TP}} + N_{\text{FP}}) \tag{6}$$

$$\text{Recall} = N_{\text{TP}}/(N_{\text{TP}} + N_{\text{FN}}) \tag{7}$$

Unfortunately, codes for static background subtraction in maritime research are not available. We implemented histogram comparison method (HistComp) of Smith and Teal [37] since sufficient implementation details were available. Further, we implemented a GMM for background subtraction in an image (StatGMM). We used the reference background computed in [37] for fitting one GMM each for the red and green color channels. Blue channel was not considered since the histogram of the blue channel's data is very narrow compared to the histograms of the other channels for maritime images. Kullback Leibler (KL) divergence [83] of these histograms from the GMMs are determined. In the KL map of each pixel, positive values indicate that the local histograms of the red and green color values are close to the static GMMs.

TABLE VI

QUANTITATIVE COMPARISON OF BACKGROUND SUBTRACTION APPROACHES FOR ON-SHORE VIDEOS OF SINGAPORE MARITIME DATASET. THE BEST VALUES IN EACH COLUMN ARE INDICATED IN BOLD

| | Precision ($\times 10^{-2}$) | | | | Recall ($\times 10^{-2}$) | | | | Time/ frame (ms) |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Q25 | Q50 | Q75 | Mean | Q25 | Q50 | Q75 | |
| Static background subtraction | | | | | | | | | |
| HistComp | 0.01 | 0.00 | 0.00 | 0.01 | 0.06 | 0.00 | 0.02 | 0.05 | >1000 |
| StatGMM | 0.01 | 0.00 | 0.00 | 0.01 | 0.08 | 0.00 | 0.00 | 0.03 | >1000 |
| Dynamic background subtraction | | | | | | | | | |
| TempMean | 0.07 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 43 |
| AdaMed | 0.25 | 0.06 | 0.13 | 0.26 | **14.72** | **6.96** | **11.00** | **19.87** | 61 |
| GMM | 0.56 | 0.05 | 0.29 | 0.67 | 9.71 | 1.91 | 7.96 | 13.77 | 54 |
| KDE | 0.59 | 0.08 | 0.48 | 0.89 | 8.93 | 1.87 | 9.52 | 13.31 | 83 |
| OptFlow | **11.64** | **1.65** | **7.21** | **14.50** | 13.35 | 0.91 | 6.85 | 17.99 | 360 |
| IMBS | 0.86 | 0.33 | 0.62 | 0.99 | 7.67 | 2.23 | 6.62 | 11.15 | 156 |
| CV methods for background subtraction | | | | | | | | | |
| LBP | 0.31 | 0.04 | 0.27 | 0.45 | 3.79 | 0.67 | 1.88 | 4.49 | 629 |
| LBSP | **8.00** | 0.04 | **3.36** | **9.29** | 6.08 | 0.03 | 2.73 | 8.92 | 589 |
| FuzzGMM | 0.01 | 0.00 | 0.00 | 0.01 | 0.06 | 0.00 | 0.02 | 0.05 | 63 |
| FAdaSOM | 0.55 | **0.18** | 0.32 | 0.85 | 11.01 | 3.07 | 9.67 | 13.89 | 133 |
| EigHMM | 0.26 | 0.05 | 0.14 | 0.30 | **14.55** | **4.68** | **12.63** | **18.48** | 245 |

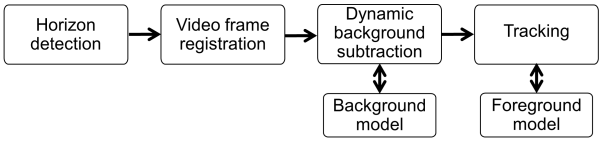


Fig. 5. General pipeline of maritime EO data processing for object tracking.

This implementation is intended to serve as the worst performance scenario of GMM for maritime on-shore videos.

The comparison results are presented in Table VI. The histogram comparison technique of [37] and static GMM perform almost similar, neither providing adequate precision and recall. We note that HistComp was tested on maritime intensity images acquired using low-resolution infrared cameras, and the high definition visible range color videos in Singapore-Maritime dataset may have caused the poor performance. Similarly, most static background subtraction techniques were tested on low-resolution intensity images. Thus, these methods may not be suitable for high-resolution maritime imaging.

## IV. OBJECT TRACKING

In much of the literature related to object tracking in maritime environment, the problem of object tracking is reduced to the problem of object detection in every frame. We differentiate between object detection algorithms and object tracking algorithms in that the latter use (i) temporal information across frames, e.g. optical flow, and (ii) employ dynamic background subtraction algorithms for more robust modeling of the background. A typical pipeline for maritime object tracking is shown in Fig. 5. Below, we discuss each of the modules in the pipeline.

### *A. Utility of Horizon Detection*

We discuss the use of horizon detection in object tracking. In object tracking, the main purpose of horizon detection is to allow for registration over consecutive frames and compensate

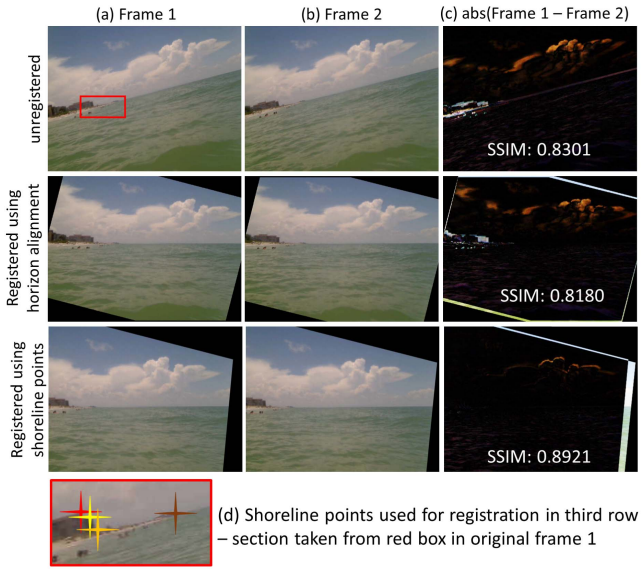


Fig. 6. The top row shows two consecutive frames and their difference. Second row: result of registration results horizon. Third row: registration results using just four fixed points on the shoreline. Fourth row: The four points used for registration in the third row. Saturation and brightness of the difference images in column (c) have been enhanced for better illustration. SSIM [84] for the image pairs is provided in column (c).

for the motion of camera or its mounting base (such as due to turbulence of water inducing motion in a boat). Horizon may also be used for determining special conditions for detecting objects close to horizon. For example, [85] used smaller video bricks close to horizon as compared to elsewhere. In some cases, horizon was used as an indicator of the distance between the camera and the vessel being tracked [30] or for motion segmentation of the vessel [86]. However, sensitivity of the distance computation to the error in horizon detection was noted as severely restrictive in [30], [86].

### *B. Registration*

In maritime scenario, consecutive frames may experience large angular or positional shift. The angular difference may be due to yaw, roll, and pitch of the vessel. The positional shift comes from the fact that the sensor itself is not necessarily mounted at the effective center of motion of the vehicle. If the horizon is present, the discrepancy in roll and pitch can be corrected by the change of angle and position of the horizon, respectively. However, the correction of the yaw cannot be achieved. To illustrate this point, consider two consecutive frames of a video shown in Fig. 6. Horizon based registration is partially effective (2nd row, 3rd image) but the mismatch along the horizontal direction indicates that the shift in yaw is not corrected. In order to correct for the yaw, we can use additional features from the scene that can help in registration as shown in the third row of Fig. 6. However, such features are not available in scenes of high seas. Fefilatyev et. al [16], [59] computed normalized cross-correlation in the horizontal direction along a narrow horizontal strip around the horizon in the images to be registered. The peaks of the normalized cross-correlation function indicated the amount of shift between the two frames. An example is given in Fig. 7.

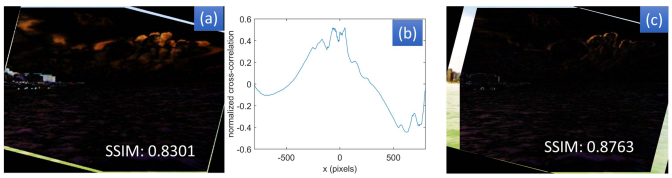


Fig. 7. Registration using cross-correlation of strip around the horizon. (a) The difference image obtained by registration using horizon only, reproduced from Fig. 6. (b) The cross-correlation function of Fefilatyev [59]. (c) The difference image after horizontal shift of 48 pixels, identified as the peak in (b). Saturation and brightness of the difference images (a,c) have been enhanced for better illustration. SSIM [84] for the image pairs is provided.

TABLE VII

QUANTITATIVE EVALUATION OF REGISTRATION APPROACHES ON ON-BOARD VIDEOS OF SINGAPORE MARITIME DATASET

| | Mean | Q25 | Q50 | Q75 | Q90 | Time/frame (ms) |
|---|---|---|---|---|---|---|
| Unregistered | 0.75 | 0.66 | 0.74 | 0.83 | 0.98 | not applicable |
| Using horizon | 0.81 | 0.75 | 0.81 | 0.88 | 0.94 | 68 |
| Correlation [59] | 0.81 | 0.75 | 0.81 | 0.89 | 0.94 | 465 |
| Feature matching | 0.75 | 0.66 | 0.73 | 0.83 | 1.00 | 279 |

We present a quantitative comparison of registration techniques discussed above on the on-board videos in Singapore Marine dataset. We used the ground truth of horizon for performing horizon based registration. For registration using correlation technique of [59], we found that strip of width 100 pixels centered at horizon gave the best result. Lastly, we used speeded up robust features (SURF) for registration using feature matching. We used all the features that could be matched between a pair of images to perform registration.

We use structural similarity index metric (SSIM) [84] to compute the similarity between two consecutive frames. It is specifically suitable for texture matching, and thus a good metric for maritime images with dynamic background. The mean SSIM for all the consecutive frame pairs is 0.75, as seen in Table VII. Even the Q25 value of SSIM is 0.66. Registration using horizon and using the cross-correlation technique [59] improves SSIM by 6-7%. However, the Q90 values show a decrease, indicating that some consecutive frames are less similar after registration using horizon only or cross-correlation technique of Fefilatyev et. al [59]. On the other hand, registration using feature matching hardly improved SSIM, with improvement appearing in only a few frames as noted from the Q90 value of SSIM. The results indicate that the cross-correlation technique [59], although simple, does not provide a significant advantage over registration using horizon for the current dataset although it was found to be effective for the videos taken from buoy mounted camera in [59]. We expect that this may be related to either the variety of structures seen along the horizon or the angle of the camera with the sea surface. On the other hand, registration using SURF features is not effective because of the lack of reliable stationary features in on-board maritime videos.

### C. Dynamic Background Subtraction

As discussed in section III-B, long-wave infrared sensors suppress dynamicity of water, which is amiss in videos acquired from visible range sensors. If static background methods are used for such videos, the dynamicity of water causes incorrect detections. Methods that explicitly model the background as being dynamic are more effective in this case. Here, we discuss the dynamic background subtraction approaches used for maritime EO videos processing.

*1) Relatively Stationary Pixels:* If a pixel corresponds to sky or water which is relatively stationary in the past few frames, the temporal distributions of intensities at a pixel over these frames are expected to be unimodal. The mean or median of the distribution is used to determine if the pixel belongs to the background or foreground [18], [37], [87]–[89]. A simple threshold approach was used for learning the background across the frames in [17]. At each pixel, $L_p$ norm of intensities over a small temporal window $I_{\text{thresh}}(x,t) = \|\big(I(x,t') - \tilde{I}(x,t')\big); \forall 0 \leq t - t' \leq T\|_p$ was computed. Here, $x$ and $t$ represent the pixel and the current frame, $T$ is the size of temporal window, $I$ and $\tilde{I}$ represent the actual and fitted intensities respectively, and $\| \ \|_p$ represents the $L_p$ norm. The fitted intensity $\tilde{I}$ was obtained by fitting a polynomial over the measured intensities $I$ in the temporal window of size $T$. If $\big(I(x,t') - \tilde{I}(x,t')\big) < I_{\text{thresh}}(x,t)$, the pixels were assigned to the background.

*2) Spatio-Temporal Filtering Approaches:* Wavelet transformation was used in [22], [39] for suppressing the background. In [40], wavelet transform and support vector machine on low frequency wavelets were used to detect objects, followed by correlation over 5 frames, and adaptive segmentation. Low frequency wavelets were assumed to contain less information of clutter and then the uncluttered background would not correlate over the frames, thus both clutter and background could be taken care of.

*3) Gaussian Mixture Models:* Gaussian mixture models have already been introduced in the context of stationary background in section III-B.2. In Bloisi and Iocchi [12] fitted a trivariate GMM for RGB values at each pixel over last few frames. The pixel was labeled as foreground if its RGB values differed from the GMM by a threshold $t$ that depended upon the illumination conditions. Gupta et. al [77] also learnt the Gaussian model over past few frames only, although using time-weighted intensity values. The time weighting allowed for the GMM to adapt to the changing conditions to a small extent. A test pixel was classified as background if the significance score, proportional to the square of distance between the intensity of the test pixel and the mean of the Gaussian model, was small.

*4) Kernel Density Estimation (KDE):* Mittal et. al [19] considered the ocean as dynamic background which was subtracted to detect people on shore. It used kernel density estimation (KDE) for background subtraction, in which the background model is represented as:

$$P(I_t) = \frac{1}{n}\sum_{i=1}^{n} K(I_t, \kappa_i) \tag{8}$$

where $P(I_t)$ is the probability of the intensity at a time $t$ at a given pixel is $I_t$, $K(I,\kappa)$ is the kernel function for the intensity $I$ with kernel parameters specified by $\kappa$, and $n$ is the number of kernels. Typically, single parameter kernels

TABLE VIII
SUMMARY OF DYNAMIC BACKGROUND SUBTRACTION APPROACHES FOR OBJECT TRACKING

| Approach | Model | Learning | Advantages | Disadvantages |
|---|---|---|---|---|
| Relatively stationary pixels [17], [18], [37], [88]−[90] | Temporal filter extracts statistical representative of background | Sliding temporal window, change from reference, spatial smoothing | Simple, computation efficient, online learning | Cannot deal with highly dynamic backgrounds, e.g., wakes |
| Spatio-temporal filtering approaches [22], [39], [40] | Background is modelled as low spatial frequency component, albeit with temporal variation | Sliding temporal window and fixed spatial window (blobs or neighborhood) are used for filter parameters' update | Simple, online learning, computation efficient, robust to small dynamics and illumination variation | Cannot deal with highly dynamic backgrounds, e.g., wakes |
| GMM [12], [77] | Intensities at a pixel as mixture of Gaussian distributions | Fitting GMMs on histograms of intensities over past few frames | Online, less memory intensive, simple, adaptive | Cannot accommodate complex intensity distributions and sudden illumination changes |
| Kernel density estimation[19] | Background modelled as sum of kernels of adaptive spreads | Learnt through fitting over last few frames | Asymmetric kernels may be used, online learning, fast and adaptive, can deal with small illumination variations, wakes, and foam better than GMM | Kernels should be good representative or else several kernels may be needed, adaptive nature makes it sensitive to variations |
| Optical flow [91] | Segment initial background, compare with background predicted by motion map | Learnt by spatial gradients' patterns over time as velocities | Suitable for wakes, big waves, and clouds | Not suitable for random wave motion, computation intensive |
| Multi-step approaches [13], [21] | Combination of more than one technique | | More robust and versatile, often made adaptive and capable of dealing with wakes | Complicated, computation intensive, slow due to frequent feedback —feed-forward steps |

are used. KDE is different from GMM in two respects. First, GMM uses Gaussian kernel whereas KDE allows the kernel to be asymmetric or have suitable statistical properties. Second, unlike GMM, KDE need not use supervised learning. Typically only a few frames are used for computing the probability distribution and the assumed kernels are fit upon it.

*5) Optical Flow:* Optical flow methods learn the patterns of motion from the videos. The flow vectors computed by comparing adjacent frames are used to warp each frame with respect to a reference frame such that stationary components can be identified as background. Ablavsky et. al [90] used optical flow technique for maritime images, specifically to model wakes as background. From a given image frame, pre-learnt motion maps were used to predict next image frame and correlate it with the actual new frame. The pixels with high correlation were labelled as background. We note that optical flow is used as one component in their multi-module interconnected framework for background subtraction. The other components are Bayesian probabilistic background estimation, motion map filter, and coherence analyzer.

*6) Multi-Step Approaches:* These approaches combine more than one technique to achieve better background subtraction.

Independent Multimodal Background Subtraction (IMBS) was proposed by Bloisi et. al [13]. It 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 modelling 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 in filtering out false detections due to shadows [91], reflections [92], and boat wakes. It models wakes as outliers of the foreground, forming a second background model specifically for such outliers.

Socek et. al [21] used a four-step process for background extraction: change detection, change classification, foreground segmentation, and background model learning and maintenance. Change detection was done by subtracting the incoming image from a reference (pre-learnt) stationary image. The detected change image was then analyzed to find if the changes correlate to the prediction of Bayesian background model [78] or are they likely to be foreground. The regions classified as foreground were then used with color-based segmentation approach to further strengthen the foreground estimation and thus contribute to more robust background detection. The background thus determined was used to update the reference stationary image and the Bayesian background model.

*7) Comparison of Techniques for Dynamic Background Subtraction:* A qualitative comparison of the dynamic background segmentation methods is given in Table VIII. In this section, we compare the performance of dynamic background subtraction techniques for the on-shore videos in Singapore Marine dataset. We have used the on-shore videos only to ensure that the dynamics correspond to the scene only and not to the sensor. The metrics and methodology of comparison are the same as discussed in section III-D. Since the source codes or executables of the methods for maritime dynamic background subtraction are not available, with the exception of IMBS [13], we have used temporal mean (TempMean) background model implementation of [93] as an example of techniques that use the concept of relatively stationary pixels, adaptive median filtering (AdaMed) [94] as an example of spatio-temporal filtering approaches, Gaussian mixture model (GMM) of [95], kernel density estimation (KDE) of [96], Lucas-Kanade approach [97] for optical flow (OptFlow) based background subtraction, and IMBS as an example of multistep approaches for background subtraction. We used the computer

vision toolbox of Matlab for optical flow segmentation using Lucas-Kanade [97] approach. The video was scaled down to 0.5 times its actual size in pixels before computing the optical flow. Bounding boxes with dimensions less than 10 pixels in the scaled down frames were filtered away to suppress the motion due to water. The remaining bounding boxes were used after scaling up to the original dimensions. For IMBS [13], we have used the source code of the authors. For the rest, we have used the codes in the background subtraction library [93].

The comparison results are shown in Table VI. With the exception of temporal mean approach, the other dynamic background subtraction approaches invariably perform better than the static background subtraction approaches. Further, the noticeably better precision of optical flow based approach is attributed to the down-scaling of the frames which suppressed detection of extremely small spurious characteristic of motion of water and the filtering away of small foreground segmentations. In terms of recall, the adaptive median approach of [94] performs the best, despite being simple. Nevertheless, none of the methods provide practically useful precision and recall.

### D. Tracking

In a typical object tracking pipeline, objects are extracted by background subtraction and segmented before they are tracked. However, in some cases, tracking is done even without segmenting the background, as discussed in section IV-D.6.

*1) Basic Tracking Techniques:* Hu et. al [18] formulated the problems of tracking as computation of an adaptive bounding box, where the bounding box in current frame is an adaptation of the bounding box in the previous frame within specified ranges of adaptivity to compensate for the background mismatch between the current and previous frames. Temporal high pass filter (analogous to fast moving objects) of segmented shapes was used in [23]. Robert-Inacio et. al [44] tested the locations of the objects in consecutive frames for expected speed range. Objects were tracked as long as such speed of the object persists. Westall et. al [79] used dynamic programming for tracking of the objects in the videos.

*2) Feature Based Tracking:* Methods in which prominent features of the objects are used for tracking are more suitable for dealing with occlusion. Bloisi et. al [14] used Haar features for detecting and tracking objects. It is notable that [14] used visible range color images as input and their features detection strategy for color images may not be directly useful for IR images. Other key point detectors, such as Harris corner detector, were found to be more effective [98].

*3) Shape Tracking With Level-Sets:* Casting segmented shapes (shape contours) as level-sets [11], [65] and then evolving the level-sets over frames has also been found useful for tracking. Notably, level-set techniques generally require the number of foreground objects and their initial contours to be specified [11], [65]. On one hand, knowing the number of objects require pre-segmentation of the objects, even if crude estimates are used. On the other hand, specifying initial contours imply that occluded objects may be difficult to deal with in such techniques. The level-set based approaches can benefit from some shape prior [65] which may be known through the general geometry knowledge of the expected objects (sea vessels of various kinds for maritime problem).

*4) Bayesian Predictive Network:* In the Bayesian approaches for tracking, the objects or features in a frame to be tracked are considered as a state vector of the Bayesian network and the training of a predictive model for transition between states is done to obtain a predictive model [79]. Often, once a Bayesian network is trained, it can be used to predict the state of the next frame given the state(s) in the previous frame(s). Thus Bayesian networks and hidden Markov models are suitable for tracking of objects that have a relatively smooth or predictable motion. They are suitable for dealing with multiple objects as well as occlusion as the objects' locations and features in a frame can be assigned a state variable each while the occlusion can be dealt with due to the predictive nature of the networks. However, it is difficult to make them adaptive and thus agile to learn complex motion characteristics such as some objects becoming stationary for some time.

*5) Kalman Filters:* Tracking large objects such as ships in port environment was performed using a mixture Kalman filter approach in [99]. Kalman filters were used in [100], [101] for learning the motion of the foreground objects. In the original form, Kalman filters cannot deal with multiple hypotheses, or multiple object motion tracking simultaneously [102]. Thus, a multi-hypotheses Kalman filter for tracking was proposed in [103] and its optimal implementation was presented in [104]. It was found useful in maritime problems [12], [16], [26], [60].

It was reported in [12] that the multi-hypotheses Kalman filter approach provides a good balance between computation load and tracking robustness. We note that Bloisi and Iocchi [12] tested this on high resolution video stream and its validity on lower quality videos is not assured. Wei et. al [26] used manually pre-assigned initial tracks and simple Kalman filters to track the objects instead of multi-hypotheses Kalman filters.

*6) Motion Segmentation Using Optical Flow Approach:* Direct foreground tracking without pre-segmenting the foreground can be done by 'motion segmentation'. The spatial information is incorporated implicitly as the features or pixels with same motion characteristics are likely to belong to the same foreground object. Although several motion segmentation approaches are used in the computer vision, as discussed later in section V-B, methods for maritime EO problem have used optical flow based motion segmentation only [12], [19]. The underlying assumptions in optical flow based techniques are that the objects are rigid and the motion is smooth.

Bloisi et. al [12] first computed connected segments in the foreground, which are not necessarily the foreground objects. Then, they computed sparse motion maps for each blob. The wakes and shadows do not have a consistent motion map and thus are suppressed in the optical flow approach. In order to deal with multiple foreground objects in one blob, a modified k-means clustering approach was applied on the optical flow map. First the motion map was over-clustered into many small clusters using k-means clustering. Then, the clusters were iteratively merged till further merging reduced the cluster

TABLE IX
SUMMARY OF TRACKING APPROACHES USED IN MARITIME OBJECT TRACKING

| Approach | Model | Learning | Advantages | Disadvantages |
|---|---|---|---|---|
| Basic tracking techniques [18], [23], [44], [79] | Adaptive bounding box, Temporal high pass filter | Online using a small temporal window and memory of previous tracking | Simple, computation efficient, adaptive bounding box can deal with wakes | Naive, non-predictive |
| Feature based tracking [14], [99] | Track features of segmented objects | Match features across frames | More robust than shape tracking, may allow some deformation (aspect change) | Features may not be consistently present, selection of appropriate features is important, computation intensive |
| Shape tracking with level-sets [11], [65] | Segmented shape contours are cast into level sets | Level sets are evolved through the frame | Apriori knowledge not required, but beneficial | Cannot deal with occlusion, sensitive to shape segmentation, computation intensive |
| Bayesian predictive network [79] | Features of objects case as state vectors of network representing motion model | Learning techniques such as expectation maximization are used to progressively update the network | Predictive nature, learns motion adaptively | Requires features as state variables, very computation intensive, cannot deal with complex motion patterns |
| Kalman filters [12], [16], [26], [60], [100] | Motion model is represented as state vector at a time t, current foreground segmentation's feature is case as input vector (to update the state vector) and Gaussian random motion perturbation is cast as noise (to be filtered) | State vector is updated for least square error between the actual measurement (input) and the predicted measurement (using previous state vector) | Almost real time, can filter away random perturbations, can adapt to multiple objects, can deal with complex motion variations occurring slowly | Does not work for complex random small motions, needs special framework for multiple object tracking |
| Optical flow [12], [19] | Computes motion maps or flow vectors to determine consistent motion patterns | Cluster using modified k-means clustering on flow maps or compute flow equation at each pixel | Wakes and waves are automatically suppressed due to inconsistent motion patterns | Cannot deal with boat maneuvers and is computation intensive |

separation instead of increasing the cluster separation. Notably, optical flow method fails in boat maneuvers because the optical flow may detect different directions for the different parts of the boat. Further, two boats having very similar motion characteristics and present in one blob cannot be separated by optical flow as well as k-means clustering.

Mittal et. al [19] computed the optical flow velocity vector $f$ by solving the flow constraint equation $\nabla g \cdot f + g_t = 0$, where $g$ is the measured value of a color channel, $\nabla g$ is the gradient of $g$, and $g_t$ is the temporal derivative of $g$. This equation was solved at each pixel such that the error in the estimated flow vector $f$ is minimised while satisfying the constraint that the velocity $f$ is locally constant.

*7) Comparison of Techniques for Tracking:* A qualitative comparison is given in Table IX. Here, we present a performance comparison of techniques of tracking on on-shore videos of Singapore Maritime dataset. We have used only on-shore videos so that the performance of tracking is not biased due to camera's motion. Performance metrics for tracking [105], namely precision, recall, multiple object tracking accuracy (MOTA), multiple object tracking precision (MOTP), and false alarm rate (FAR) are used, which we describe below.

An $i$th ground truth (GT) track is represented by its GT bounding box $O_{i,t}^{\mathrm{GT}}$ in frame $t$. Analogously, $O_{j,t}^{\mathrm{det}}$ denotes the bounding box of the $j$th track at time $t$ detected by a tracking technique (simply referred to as a tracker). Then, for $i$th ground truth track and $j$th detected track, $\mathrm{IOU}(i, j, t)$ denotes the value of IOU computed using eq. (5) for the pair $(O_{i,t}^{\mathrm{GT}}, O_{j,t}^{\mathrm{det}})$ at frame $t$. Matched pairs of ground truth tracks and detected tracks are determined using Hungarian method [106] with $1 - r(i, j, t)$ as input, where $r(i, j, t)$ is defined as

$$r(i, j, t) = \begin{cases} \mathrm{IOU}(i, j, t) & \text{if } \mathrm{IOU}(i, j, t) > 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Hereafter, $(i, j)$ denotes a matched pair of $i$th ground truth track and its corresponding $j$th detected track. In a frame $t$, all unmatched ground truth tracks contribute to one false negative (FN) each and all unmatched detected tracks contribute to one false positive (FP) each. For the matched tracks, if $\mathrm{IOU}(i, j, t)$ is more than 0.5 (based on [105]), the detection is said to be true positive (TP) for that frame. Otherwise it contributes to a mismatch (MM). Thus, $N_{\mathrm{TP},t}$, $N_{\mathrm{MM},t}$ , $N_{\mathrm{FP},t}$, and $N_{\mathrm{FN},t}$ are the numbers of TPs, mismatches, FPs, and FNs in a frame $t$. Further, total number of matched pairs of ground truth and detected tracks in a frame $t$ irrespective of the values of IOU is given as $N_{\mathrm{M},t} = N_{\mathrm{TP},t} + N_{\mathrm{MM},t}$ and the total number of frames in the video is denoted as $T$. Precision, recall, FAR, MOTA, and MOTP are then defined as

$$\text{Precision} = \frac{\sum_t N_{\mathrm{M},t}}{\sum_t \left(N_{\mathrm{M},t} + N_{\mathrm{FP},t}\right)} \tag{10}$$

$$\text{Recall} = \frac{\sum_t N_{\mathrm{M},t}}{\sum_t \left(N_{\mathrm{M},t} + N_{\mathrm{FN},t}\right)} \tag{11}$$

$$\text{FAR} = \sum_t N_{\mathrm{FP},t} / T \tag{12}$$

$$\text{MOTA} = 1 - \frac{\sum_t \left(N_{\mathrm{FP},t} + N_{\mathrm{FN},t} + N_{\mathrm{MM},t}\right)}{\sum_t \left(N_{\mathrm{M},t} + N_{\mathrm{FN},t}\right)} \tag{13}$$

$$\text{MOTP} = \frac{\sum_{(i,j),t} r\left((i, j), t\right)}{\sum_t N_{\mathrm{M},t}} \tag{14}$$

Precision and recall have their usual range of [0,1] with best values being 1. The unit of FAR is number of false positives per frame and its value may be any non-negative real number, with 0 being the best value. MOTA may take negative values but has a maximum and best value of 1. MOTP lies in the range [0,1], the best value being 1.

TABLE X

QUANTITATIVE EVALUATION OF DIFFERENT TRACKING TECHNIQUES FOR ON-SHORE VIDEOS OF SINGAPORE MARITIME DATASET. THE BEST VALUES FOR EACH METRIC ARE HIGHLIGHTED USING BOLD FONT

| Techniques related to maritime | | | | | | |
|---|---|---|---|---|---|---|
| Metric | | MST | KLT | DAOT | MOT | LKDoG |
| Precision | Mean | 0.57 | **0.73** | 0.65 | 0.01 | 0.00 |
| | Q25 | 0.38 | **0.60** | 0.53 | 0.00 | 0.00 |
| | Q50 | 0.52 | **0.70** | 0.65 | 0.01 | 0.00 |
| | Q75 | 0.83 | **0.86** | 0.77 | 0.01 | 0.00 |
| Recall | Mean | 0.53 | **0.77** | 0.68 | 0.04 | 0.02 |
| | Q25 | 0.31 | **0.68** | 0.57 | 0.0.00 | 0.00 |
| | Q50 | 0.50 | **0.76** | 0.68 | 0.03 | 0.01 |
| | Q75 | 0.77 | **0.88** | 0.82 | 0.06 | 0.02 |
| MOTA | Mean | 0.15 | **0.47** | 0.29 | -6.45 | -11.79 |
| | Q25 | -0.20 | **0.20** | 0.07 | -8.36 | -10.02 |
| | Q50 | 0.05 | **0.45** | 0.32 | -5.43 | -8.00 |
| | Q75 | 0.59 | **0.72** | 0.59 | -4.07 | -5.67 |
| MOTP | Mean | 0.74 | **0.80** | 0.68 | 0.61 | 0.45 |
| | Q25 | 0.71 | **0.76** | 0.66 | 0.59 | 0.51 |
| | Q50 | 0.74 | **0.81** | 0.68 | 0.60 | 0.55 |
| | Q75 | 0.77 | **0.82** | 0.69 | 0.62 | 0.59 |
| FAR | Mean | 3.46 | **2.70** | 3.56 | 46.85 | 81.10 |
| | Q25 | 0.97 | **0.91** | 1.16 | 42.21 | 46.45 |
| | Q50 | 2.66 | **2.43** | 2.83 | 48.73 | 58.36 |
| | Q75 | 5.79 | **4.40** | 5.20 | 52.43 | 81.17 |
| Time/frame (s) | | 1.66 | 0.51 | 0.73 | **0.26** | 1.14 |
| Other computer vision techniques | | | | | | |
| Metric | | AdaBoost | MIL | TLD | MedFlow | KCF |
| Precision | Mean | 0.82 | 0.62 | 0.26 | 0.76 | **0.86** |
| | Q25 | 0.68 | 0.48 | 0.16 | 0.69 | **0.77** |
| | Q50 | 0.86 | 0.71 | 0.26 | 0.75 | **0.90** |
| | Q75 | 0.91 | 0.78 | 0.35 | 0.86 | **0.96** |
| Recall | Mean | 0.83 | 0.63 | 0.27 | 0.77 | **0.87** |
| | Q25 | 0.75 | 0.48 | 0.16 | 0.72 | **0.79** |
| | Q50 | 0.85 | 0.70 | 0.27 | 0.79 | **0.90** |
| | Q75 | 0.92 | 0.79 | 0.36 | 0.84 | **0.95** |
| MOTA | Mean | 0.64 | 0.23 | -0.50 | 0.52 | **0.72** |
| | Q25 | 0.46 | -0.03 | -0.71 | 0.39 | **0.55** |
| | Q50 | 0.66 | 0.42 | -0.49 | 0.52 | **0.80** |
| | Q75 | 0.83 | 0.54 | -0.30 | 0.68 | **0.89** |
| MOTP | Mean | 0.79 | 0.75 | 0.63 | 0.79 | **0.80** |
| | Q25 | 0.77 | 0.69 | 0.60 | 0.76 | **0.78** |
| | Q50 | **0.80** | 0.75 | 0.63 | 0.79 | **0.80** |
| | Q75 | 0.82 | 0.80 | 0.65 | **0.83** | **0.83** |
| FAR | Mean | 1.73 | 3.52 | 6.56 | 2.24 | **1.37** |
| | Q25 | 0.70 | 1.51 | 4.07 | 1.17 | **0.19** |
| | Q50 | 1.13 | 2.45 | 6.21 | 1.57 | **0.91** |
| | Q75 | 2.37 | 4.99 | 8.53 | 3.35 | **2.09** |
| Time/frame (s) | | 12.03 | 3.31 | 42.73 | **0.42** | 0.47 |

We consider one technique per row of Table IX, with the exception of level-set based tracking, for which we could not find an implementation. We use mean shift tracking[3] (MST) [110] as an example of basic tracking techniques, Kanade-Lucas-Tomasi (KLT) feature tracker[4] [97], [111] for feature based tracking, distracter-aware online tracking (DAOT,[5] [112]) for tracking using Bayesian predictive network, motion-based multiple object tracking[4] (MOT) using Kalman filter [113] for Kalman filter based tracking, and optical flow based on Lukas-Kanade difference of Gaussian method[4] (LKDoG, [114]). No background subtraction technique has been applied in order to compare the performance of tracking only.

[3]https://www.mathworks.com/matlabcentral/fileexchange/35520-mean-shift-video-tracking

[4]Matlab provided function

[5]Matlab code provided by the authors of [112])

TABLE XI

SUMMARY OF LITERATURE ON OBJECT DETECTION AND TRACKING IN MARITIME SCENARIO

| Articles | Year | EO sensor | | Scene | | Object detection | | | Object tracking | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Infrared | Visible | On shore | Open sea | Horizon detection | Static background | Foreground segmentation | Horizon detection | Registration | Dynamic background | Foreground tracking |
| Bhanu [29] | 1990 | • | • | | | | • | | | | | |
| Sumimoto [23] | 1994 | | • | | | | | • | | | | |
| Strickland [22] | 1997 | | • | | | | | | | | • | |
| Smith [37] | 1999 | • | | | | | • | | | | | |
| Broek [32] | 2000 | • | | | | • | • | • | | | | |
| Voles [85] | 2000 | | | | | | | | • | | • | • |
| Caspi [108] | 2002 | • | • | • | | | | | | • | | • |
| Ablavsky [91] | 2003 | | | | | | | | | | • | • |
| Mittal [19] | 2004 | | • | | | | | | | | • | • |
| Socek [21] | 2005 | | • | | | | • | • | | | • | |
| Fefilatyev [15] | 2006 | | • | • | • | • | | | | | | |
| Wang [40] | 2006 | • | | | | | | • | | | • | |
| Robert-Inacio [44] | 2007 | • | • | • | | | • | | | | | • |
| Schwering [36] | 2007 | • | • | • | | | | | | | | |
| Bouma [31] | 2008 | • | | • | • | • | • | • | | | | |
| Broek [30] | 2008 | • | • | | | | • | | | | | |
| Zheng [109] | 2008 | • | • | | | | | | | | | |
| Bloisi [12] | 2009 | | • | | | | | | | | • | • |
| Gupta [77] | 2009 | | | | | | | | • | | • | • |
| Haarst [17] | 2009 | | • | | | | • | | | | | |
| Wei [26] | 2009 | | • | | | • | • | • | | | | |
| Fefilatyev [16] | 2010 | | • | • | • | | • | | • | • | | • |
| Zhu [67] | 2010 | | | | | | • | | | | • | |
| Bloisi [14] | 2011 | | • | • | | • | | | | | | |
| Hu [18] | 2011 | | • | • | | | | | | | • | • |
| Szpak [11] | 2011 | | • | | | | • | • | | | | |
| Wang [41] | 2011 | • | | | | • | • | | | | | |
| Broek [99] | 2012 | • | • | • | | | | • | | | | |
| Ren [20] | 2012 | | • | | | | | | | | • | |
| Zhang [66] | 2012 | | • | • | | | • | | | | | |
| Frost [65] | 2013 | | | | | | • | | | | | • |
| Gershikov [28] | 2013 | • | • | | | • | | | | | | |
| Tang [38] | 2013 | • | | | | • | • | | | | | |
| Bloisi [13] | 2014 | | • | | | | | | | | • | |
| Broek [33] | 2014 | • | | | | • | • | • | | | | |
| Broek [34] | 2014 | • | | | | • | • | • | | | | |
| Chen [35] | 2014 | • | | | | | • | • | | | | |
| Tu [39] | 2014 | • | | | | | | | | | • | |
| Wang [24] | 2014 | | • | | | • | • | | | • | | |
| Zhou [27] | 2014 | | • | | | | • | | | | | |
| Babaee [110] | 2015 | | • | | | | | | | • | | |
| Wang [25] | 2015 | | • | | | | | | | | | • |

MST, KLT, and DAOT are single object trackers and require initial guess (we used the first bounding box of each GT track). MOT and LKDoG do not require initial guess and track multiple object simultaneously. The results are presented in Table X. MST, KLT, and DAOT clearly benefit from the initial guess because the number of tracks remains close to the actual number of ground truth tracks. On the other hand, MOT and LKDoG suffer due to large number of false positives as a consequence of water dynamics. Among MST, KLT, and DAOT, MST performed the best in terms of all the metrics.

## V. COMPUTER VISION APPROACHES BEYOND MARITIME

A literature summary of **maritime** EO data processing for object detection and tracking is provided in Table XI.

Object detection and tracking has been studied for several decades in computer vision as well. However, due to the specific set of challenges presented by the maritime environment, not much attention has been paid in developing algorithms specific to this domain. Nevertheless, given the large number of algorithms already developed for object detection and tracking over the past years, it is only natural to seek out the algorithms that may be suitable in the maritime scenario. In this section, we identify some such algorithms which have reported at least one example with dynamic water background.

### A. Object Detection

In the context of current maritime EO data processing, the foreground obtained after background subtraction is segmented and the segmented regions are identified as the objects of interest. Thus, object detection is mainly performed through background subtraction. Different approaches have been taken for modelling and segmentation of background. There are several useful surveys on the topic of background suppression in video sequences [115]. Below, we discuss the techniques that are potentially effective in maritime videos.

*1) Relatively Stationary Pixels:* In the context of the section IV-C.1, we discuss other relevant works from non-maritime applications. Weighted average of intensities at a pixel across time was considered in [116], [117]. Median filter was employed for background suppression in [87], [88], [91], [118], [119]. First order low pass filtering was used in [118]. Toyama et. al [120] used pixel-wise temporal filter (Wiener filter). All the temporal filters essentially use temporal variation at pixels as indicator of foreground and background.

*2) Spatio-Temporal Filtering:* Ridder et. al [121] proposed to use Kalman filter for background estimation. This approach was found to be robust to illumination changes and incorporated pixel-wise automatic threshold (thus was less sensitive to control parameters). Zhong and Sclaroff [100] also used Kalman filter for representing dynamic textures.

*3) Gaussian Mixture Models:* While initial forms of Gaussian mixture models have already found use in maritime background subtraction [12], [24], [60], [65], [77], GMM has also been increasingly combined with other techniques in the computer vision community to improve the performance of object detection specifically in challenging dynamic environments. For example, the local variation persistence method [122] uses GMM for separating static background as the Gaussian component with large standard deviation and removing it, followed by numerical computation of negative differential entropy of the remaining Gaussian components which allowed for separating locally persistent variations as dynamic background. Varadarajan et. al [123] propose to use a square region based GMM, which inherently considers local spatial variations in addition to temporal variations in order to obtain a better background model for challenging dynamic backgrounds including water bodies.

*4) Kernel Density Estimation:* As mentioned in section IV-C.4, suitable kernels can be chosen for the KDE model of background. Chen and Meer [124] proposed to use Epanechnikov kernels [125], which is optimal in the least square error sense. Kato et. al [126] used a Gaussian distribution for intensity variation at a pixel, a Gaussian distribution for wavelet coefficient variation at a pixel, and their combination as a single 2-dimensional Gaussian kernel. An adaptive scheme for KDE model update was proposed in [95], where the volumes (spreads) of the kernels were made adaptive by changing the number of frames considered for the dynamic update.

*5) Optical Flow:* Ross [127] presented an interesting concept of texture-and-motion duality in optical flow in order to extract background. It used the single image segmentation approach of [128] to get an initial estimate of the background. Optical flow of the segmented regions was computed using an energy minimization approach [129]. Li and Xu [130] perform optical flow computation directions at the edges of super-pixelated regions to enhance the computation speed while allowing the identification of super-pixelated regions belonging to the dynamic background owing to the non-uniform flow vectors at their edges.

*6) Range Model:* A simple and popular approach for dynamic background extraction was considered in [131] which used a range of intensity values for a given pixel, quantified by minimum and maximum intensity values at a background pixel and maximum intensity difference between two consecutive frames, denoted as $m(x), n(x), d(x)$, respectively. For finding the parameters of this model, the pixel's intensity values in a reasonably long time sequence $I(x,t)$ were used. First, the instances $t'$ at which pixel can be considered as stationary were found as

$$|I(x,t') - \lambda(x)| < 2\sigma(x) \tag{15}$$

where $\lambda(x)$ and $\sigma(x)$ are the mean and standard deviation of $I(x,t), \forall t$. Then, $m(x) = \min(I(x,t'), n(x) = \max(I(x,t'), d(x) = \max(|I(x,t') - I(x,t'-1)|)$ were computed. The model parameters may be updated as often as needed. Haritaoglu [131] also suggested a technique for identifying that a moving object in earlier frames has become a stationary background in later frames.

Kim et. al [132] used a codebook of possible range values for addressing multi-class background. The code of a class was given by the range parameters discussed above. This was further augmented by average color data, frequency of occurrence of the code, and last access of the code.

*7) Dynamic Textures:* Local binary pattern [133] (LBP), either at a single pixel, or a small region around the given pixel [134], finds a binary number representing the boolean intensity changes in the neighborhood of the chosen pixel. It may be made shift and rotation invariant, as discussed in [133]. The LBP feature vector of a block of pixels in a frame is the histogram of the binary numbers obtained at all the pixels in the block [170]. Over time, one LBP feature vector is obtained for each frame and the net background feature vector is a weighted combination of feature vectors of the last $K$ (often heuristically chosen) number of frames. A distance measure for decision making and a model update scheme is discussed in [134]. This approach was found useful in applications involving underwater videos [135], [136].

Local binary similarity patterns (LBSP) [171] are a variation of LBP and include spatio-temporal binary similarity metric. A modification of LBP to deal with flat regions in an image is the local ternary patterns (LTP), presented and discussed in [134], [137]–[139]. Furthermore, [140], [141] proposed a mixture of dynamic textures, analogous to GMM, in order to allow for modeling of multiple dynamic textures. Mixture of dynamic textures showed good ability to deal with ocean's dynamic texture with synthetic translucent objects and flames.

*8) Hidden Markov Model of Dynamic Background:* Hidden Markov models (HMM) have two specific advantages as compared to other modelling approaches [126]. The first is its ability to incorporate temporal continuity. A pixel may be classified as belonging to background, foreground, or shadow in a particular frame. Nevertheless, it is likely that the pixel will have the same classification for at least a few continuous frames. This is so either because the object at the pixel is stationary or because the moving object occupies the pixels for some number of frames till the object crosses the pixel completely. HMM is inherently able to cover both these possibilities. Second, HMM does not require a specifically chosen training data. A scenario specific ordinary image sequence is sufficient for it to learn the hidden states that allow demarkation between background, foreground, and shadows. Further, it was noted in [115] that HMM approach is very effective in dealing with sudden illumination changes and providing a corrective temporary estimate of the background in such scenario.

Thus, despite being computationally expensive and difficult for dynamic modification of topology [115], [127], HMM has attracted a lot of attention for background suppression [21], [142]–[145], [161]. One of the most recent works in this context is [146], which showed some examples of boats in sea as well. It has many interesting and useful features, which include using dynamic textures [172] for simultaneous foreground-background modelling, augmenting the dynamic textures by introducing spatially smooth segmentation through HMM [140] and a specially designed expectation maximization approach with variational constraint.

We briefly discuss the update methods used for learning and updating the HMMs. Sheikh and Shah [147] used Markov random field with maximum-a-posteriori estimation to obtain spatial context in a simultaneous foreground-background modeling approach. Expectation maximization approaches for training HMM have been discussed in [142], [173], [174]. Stenger et. al [145] designed a dynamic update scheme for HMM which allows for adaptive topology modification of the HMM. Ostendorf and Singer [175] suggested that dynamic adaptation of HMM can be made fast by a state splitting approach. Brand and Kettnaker [176] suggested that an arbitrarily large number of states may be initially chosen and then entropy based training of HMM may be used to identify the less probable states and iteratively remove them. Wang et. al [177] used an offline Baum Welch algorithm [178] to learn HMM but employed an online algorithm for background detection and updating the HMM. Rittscher at. al [143] proposed a scheme for making HMM computationally less expensive and almost real-time. Brand and Kettnaker [176] and Ostendorf and Singer [175] discussed the optimal choice of number of states of HMM. Further, some amount of speed up of the HMM update may be achieved using subspace based approaches [100], [144].

*9) Saliency Based Approaches for Segmenting the Background:* Wixson [148] used a saliency measure defined on the cumulative optical flow directions of the moving objects (foreground). It incorporates net flow directions by computing maximum flow directions and finding observations consistent with the maximum flow direction. Such saliency measure based on maximum flow direction may be suitable for single object tracking but may need significant modification for incorporating multiple object tracking. On the other hand, Itti et. al [149] used a surprise based saliency map to segregate the non-surprising elements as the background (low saliency). It determined a surprising element as an element which has a large contrast compared to the surrounding pixels. The contrast should be consistently present at various length scales. This contrast is referred to as the center-surround difference. Although it can deal with the wavy nature of water to some extent, it is not effective in suppressing the wakes since they introduce a high contrast with respect to their surroundings.

Gao et. al [150] modified the saliency approach of [149] by retaining the center-surround and multi-scaling. It used discrimination (referred to as discriminant in [150]) between intermediate features in the center-surround instead of using the direct contrast feature [149] directly. It was tested on videos of floating bottle and surfer and showed better background identification than [149]. We note that [150] used local ternary patterns as the features of the background model.

Mahadevan and Vasconcelos [151] combined the discriminant saliency approach of [150] and mixture of dynamic textures [141] for determining spatially normal (high probability distributions) and temporally normal (high probability events) features of the videos with the moving crowd in urban scenarios. Furthering this concept, Wang et. al [177] proposed a saliency metric called spatiotemporal condition information (SCI). This metric computes the conditional information value (logarithm of conditional probability) of a pixel given the background and the spatiotemporal neighborhood of the pixel. Larger value of the conditional information indicates higher likelihood of the pixel being foreground.

Fang et. al [179] computed two saliency maps, one characterizing spatial saliency through proximity and continuity of a visually salient object region and the other characterizing temporal saliency which accounts for dynamic background variation and persistence of local contrast. These maps are merged by using an adaptive entropy-based uncertainty weighing approach to form the final spatiotemporal saliency map.

Recently, Liu et. al [180] used motion saliency map of [181] to determine the control parameter of the robust principle component analysis [157], which was then used for background subtraction and foreground extraction. In this definition of motion saliency, the sum of the background motion map $M$ such as due to water dynamics in maritime scenario and the stable background map $B$ is said to be a low rank representation of the video $V$. The background motion map

TABLE XII
BACKGROUND SUBTRACTION APPROACHES IN COMPUTER VISION HAVING POTENTIAL IN MARITIME SCENARIO (NOT COVERED IN TABLES V AND VIII)

| Approach | Description | Advantages | Disadvantages |
|---|---|---|---|
| Range model [132], [133] | Multiple class model, each with a range of intensities | Simple, pre-learnt ranges, may be made adaptive | Not discriminative |
| Local binary and ternary patterns [134]−[142] | Background model representing dynamic textures | Multiple patterns may be learnt for different dynamic textures such as water, wake, and waves, quite robust, can be learnt and adapted online | Computation intensive |
| Hidden Markov model [21],[127]−[148] | Local intensities (or other features) as state vectors in HMM | Uses temporal continuity of classification, does not require any pre-learning | Complex, computation intensive |
| Saliency based approaches [149]−[152] | Approaches for classifying pixels as background based on the used background model | More sophisticated than a simple threshold or range; can incorporate certain properties for classification, such as discriminative property and surprise | Complex, may be computation intensive |
| Fuzzy classifiers [153]−[156] | Fuzzy techniques for classification of pixel as background | Needs appropriate fuzzy classifier functions to be pre-learnt | Complex, needs supervised learning of classifier functions |
| Subspace based approaches [25], [66], [69], [101], [138], [145], [151], [157]−[170] | Learning the background model, compactly representing and fast updating of the model, finding the overlap of pixel features with the model | Fast, compact, amenable to fast linear programming | Assume linear separability of data, degrade with large dynamics in background |

and the stable background are solved by minimizing the sum of nuclear norm of $B$ and $L1$ norm of $M$.

*10) Fuzzy Classification of Background Pixels:* Fuzzy logic was used to compute an adaptive threshold for classifying the background pixels in complex background in IR images [152]–[154]. Although most of the experiments are in urban and semi-urban land scenarios, the techniques may inspire further interesting work in maritime IR images. A combination of fuzzy neural network and self organizing map [182] was used in [183]. It is shown to be robust to illumination changes and shadows, a property beneficial for maritime videos. In [184], the usual GMM background model was modified to be fuzzy mixture of Gaussian functions.

*11) Subspace Based Approaches in Background Modelling and Subtraction:* Spatio-temporal block of images, the collection of all the features of background model and the feature attributes of all the pixels may be represented as matrices. Then matrix decompositions can be used for manipulating the data, learning the model, compactly representing and fast updating the model, as well as finding the overlap of pixel features with the model in a powerful manner.

Thus, subspaces based approaches have been found useful [185] for compact representation of features at pixels before learning is performed. These include eigenbackground approach [66], [161], [186], [187], principal component analysis (PCA) [156], [160], [161], robust PCA [157], [188], independent component analysis [158], [189], and discriminant center surround [150], [159]. Subspace based learning was used for night-time videos as well in [190]. It is anticipated that these approaches may be improved by suitable combination of eigenvectors of the data and designing a robust update scheme for the background.

Autoregressive updates of background models is done by identifying the subspace of consistently recurring backgrounds [100], [120], [137], [162]. Methods in [69], [163], [164] that use correlation or covariance approaches may also be considered as subset of approaches that employ background subspace analyses. Some methods also use sparsity priors and implement background detection problem as sparse image reconstruction problem [144], [165]–[168]. Alternatively, compressive sampling based approaches can be used to reduce dimensionality of the data before using other background detection techniques for reducing the computational cost [25], [169].

*12) Comparison of CV Based Background Substraction Techniques:* A comparative summary of the background subtraction methods used in computer vision problems, but not covered in Tables V and VIII, is given in Table XII. Performance of five CV based techniques is compared in Table VI. These techniques are LBP [134], LBSP [171], FuzzGMM [184], FAdaSOM [182], and Eigen [161]. LBP and LBSP are dynamic texture approaches, FuzzyGGM is a fuzzy approach, AdaSOM is a neurofuzzy approach, and EigHMM uses a combination of HMM and eigenbackground (subspace based approach). The openCV implementations at the background subtraction library [93] are used for these methods. EigHMM performs the best in terms of recall and compares well with the recall values of AdaMed, showing good potential for maritime images. All methods perform poor in terms of precision, indicating large false positives. However, optimal choice of control parameters may render LBSP useful. Thus, we think that a combination of LBSP and eigen background may be helpful for maritime images.

### *B. Object Tracking*

Here, we discuss the object tracking techniques developed for non-maritime situations, but hold promise for maritime scenarios. Sections V-B.1 to V-B.3 discuss tracking of segmented objects while the methods in sections V-B.4 to V-B.6 do not need prior segmentation of objects.

*1) Foreground Models:* Many methods represent the foreground objects by mixture models such as GMM, local ternary patterns (LTP), and KDE, similar to the dynamic background models. The components of the mixture components are used as the features [140], [141], [146]. This permits cushion for deformability, swivel, and small randomness in motion,

which are useful for modelling moving sea vessels. Alternatively, the silhouette of the vessel may be tracked [201], [209]. Greenberg et. al [210] used morphological region growing and segmentation pruning after binarization for detecting objects with small false alarm rate.

A reverse approach was adopted in [180], [208], inspired by [211]. Zhou et. al [208] approximated the video as a low rank matrix and all the moving objects as systematic outliers to the low rank matrix belonging to an outlier support. Similar idea was used by Zhong and Sclaroff [100] and Liu et. al [180], where foreground objects were considered as outliers to the background model corrupting the estimate of the background.

*2) Temporal Persistence and Dynamic Programming for Tracking:* A sophisticated version of dynamic programming approach is used in [191], where dynamic programming updates the model parameters of the GMM representing the pre-segmented foreground object. A similar approach called temporal persistence was proposed in [147], which assumed that a mobile foreground object would remain in spatial vicinity in consecutive frames and maintain similar color or intensity values. We note that the approach of [147] falls in motion segmentation category where pre-segmentation of the foreground is not assumed and the persistently mobile Gaussian mixtures over a few frames are concluded as foreground.

*3) Machine Learning for Tracking Foreground Objects:* Machine learning techniques such as boosting based unsupervised or semi-supervised boosting techniques are often used for tracking or motion learning [192]–[196]. Often, initialization through manual segmentation (such as in [192], [194]) is needed. Boosting approaches are quite useful since they can often deal with occlusion intrinsically by considering each object's motion independently [197] and learning them over a subset or all of the frames. When the subset of frames is used at a time, often online learning can be used [193], [194], [196]. Another method is the use of principal component analysis, where a low-dimensional subset of principal components is updated as new frames arrive [198], [199]. Such approach allows for changes in views or shapes of the object with time. This flexibility is often absent in boosting approaches which match and boost the entire shape. However, boosting can be used with techniques such as multiple instance learning [196] to allow for deformable object tracking.

*4) Optical Flow Based Motion Segmentation:* Optical flow methods have been used for motion segmentation as well [19], [90], [148], [200], [201]. They incorporate spatial information by implying that the features or pixels with same motion characteristics are likely to belong to the same foreground object. Cues such as normalized color features [19] may be used to augment foreground object detection. Some methods [86], [201], [202] used partitioning of dense optical flow technique [203] to deal with large motion variations. Videos are decomposed into different motion layers with distinct motion characteristics such that each layer has smooth motion characteristics and sharp motions appear only at the edges of motion layers. Level-set techniques are often found useful for computing the dense layers and their boundaries [201].

*5) Feature Tracking and Clustering for Foreground Tracking:* Another class of methods trace the features of the objects over the frames [200], [204]–[206]. For example, sparse feature points are identified and tracked through out the video and then spectral [206] or subspace [185] clustering is applied to identify the clusters of features with same motion characteristics. Object segmentation is done by analysing the quality of clusters and post-processing [207]. We note that performing object segmentation after motion segmentation is different from performing tracking of features after object segmentation as discussed in section IV-D.2. Motion information is used for object segmentation in the former while object segmentation is used for extracting motion information in the latter.

Underlying assumptions in feature based motion segmentation are that the objects are rigid and the noise in data is limited to allow sparse tracking. While the first assumption is valid in maritime images for most scenarios, the second assumption may or may not be valid. It is notable that feature tracking methods are less computation intensive than optical flow approaches. Additionally, they can deal with random motion and large motion variations. Further, the features need not be pre-learnt. An arbitrarily large number of sparse features may be identified initially and only features with trackable motion may be retained. Other features may be classified as outliers and suppressed.

*6) Markov Random Field for Foreground Tracking:* Zhou et. al [208] modelled the motion of each individual outlier (which represents the foreground in the low rank background model) as a contiguous Markov random field. While the approach of [208] is computationally elegant and performs background suppression, foreground segmentation, and motion segmentation simultaneously, Mumtaz et. al [146] reported that [208] is not effective in suppressing wakes and shadows corresponding to the moving object. Thus, for making a method like [208] more effective for maritime object problem, two approaches may be considered. The first approach is to augment the background model of [208] with other background models such as those using local ternary patterns and visual saliency. The second approach is to use another model for background estimation, determine the corresponding support of the background (equivalent to the low rank matrix of [208]) and then use the approach of [208] to determine the outlier support and further motion modelling. Further, the estimation of the motion model may be considered as a predictive step and other suitable predictive models may be chosen if desired.

*7) Comparison of CV Based Tracking Techniques:* A comparative summary of tracking methods not covered in Table IX is given in Table XIII. We compare performances of five computer vision techniques in Table X. These techniques are AdaBoost (an online machine learning approach [193]), MIL (multi-instance learning, an online machine learning approach with support for multiple instance [196]), TLD (tracking-learning-detection, a semi-supervised machine learning approach [212]), MedFlow (an optical flow technique, [213]), and KCF (a color based feature tracking

TABLE XIII
OBJECT TRACKING APPROACHES IN COMPUTER VISION HAVING POTENTIAL IN MARITIME SCENARIO (NOT COVERED IN TABLE IX)

| Approach | Description | Advantages | Disadvantages |
|---|---|---|---|
| Temporal persistence [148], [192] | Dynamic programming, progressively update motion by finding features/objects from previous segmentation | Fast update | Simple, non-predictive, not robust to occlusion |
| Machine learning of motion [193]−[200] | Learning techniques for learning motion patterns from segmented features/objects; depending upon the technique, may need pre-learning of patterns and matching them in frames, or off-line processing of a small subset or complete set of image frames for determining the motion characteristics | Robust to occlusion, provide complete motion characteristics | Non-predictive, complex, not real-time |
| Optical flow [19], [86], [91], [149], [201]−[204] | Motion maps learnt for segmented objects or just features in image frames | Can deal with occlusion, can use multiple features simultaneously | Computation intensive, many dense motion layers for complex motion |
| Feature tracking and clustering [201], [205]−[208] | Features with same motion patterns are expected to belong to the same object | No explicit segmentation of object needed, very robust to occlusion, may be very discriminative | Not real-time, computation intensive |
| Markov random field [209] | Connected components network, where the features represent the state variables and all state variables may influence each other | Can deal with complex interdependent motion of multiple objects, robust to occlusions | Computation intensive, slow update |

approach, [214]). The implementation in the openCV tracker library is used for these methods. In general, AdaBoost, Medflow, and KCF perform better than the other methods in the whole table. KCF performs the best in all metrics and is fast as well.

## VI. CONCLUDING REMARKS

In this survey, contemporary works in maritime EO data processing have been discussed. For horizon detection, most of the work has been done by researchers working on maritime problems. In maritime EO data processing, object detection is done through segmentation of the foreground obtained after background subtraction. Thus, background subtraction is an important part of maritime EO data processing. Background subtraction may be performed on one image at a time assuming static background or may incorporate temporal information by modeling background as dynamic. In general, dynamic background approaches have better ability to deal with wakes, clouds, and foams. While a variety of methods are used in both categories of background subtraction, maritime EO data processing may benefit from other state-of-the-art background modelling techniques from the computer vision community as well. Different object tracking methods used in maritime EO processing are also discussed. Further, motion segmentation methods that do not segment the foreground to obtain objects but first learnt the motion patterns in the foreground and then cluster the patterns to identify objects are also discussed. Notably, while the gap between object tracking in maritime environment and computer vision community is relatively small, the gap in motion segmentation techniques is large. Nevertheless, we feel that motion segmentation may not be needed for on-board maritime processing since the scenes typically contain only a few objects of interest.

The study is supported with quantitative evaluation of performance of several representative maritime and computer vision techniques on Singapore Maritime Dataset. This dataset has been created with the aim of providing challenging maritime EO videos for future research. The evaluation indicates that computer vision techniques can aid maritime vision with suitable advancement.

## APPENDIX

### A. Postprocessing of Maritime EO Object Tracking Results

Here, we discuss some useful post-processing of maritime detection or tracking results. Vessel's positions and speeds are more useful in physical units [12], [42]. Bloisi et. al [12] used high mounted stationary cameras such that the height of the vessels is irrelevant and all the water surface may be considered flat and comprising of only lateral coordinates in $xy$ plane. Further, the center of each pixel observed in the camera was mapped to a physical point through careful and extensive pre-acquisition calibration, which is possible owing to the fixed nature of the cameras. A tracking boat with differential global positioning system device (GPS) was used for calibration as well as position and velocity tracking test such that GPS accuracy of few cm and few cm/sec is achievable for position and velocity respectively. This indicates the difficulty in mapping tracking results to actual physical units and highlights the importance of augmenting tracking results with information from radar sensors. Nevertheless, for a given camera's fixed position and orientation, estimates of pixel to distance relationships may be obtained with limited accuracy and may help in providing comparative or fuzzy information about the speed and location of vessels, such as vessels far away, vessels approaching or receding, closer vessels, and fast moving vessels. A simplified physical distance mapping was proposed in [42] and is given as:

$$d \approx \phi R - \sqrt{(\phi R)^2 - 2hR} \tag{16}$$

where $R$ is the radius of earth, $h$ is the height of camera, $\phi$ is the angle between the ship and horizon, and all of them are pre-known. Here, it is assumed that the object is at horizon. For a different point in the space, the angle between the point and the camera would be different. We note that this approximation is valid for points at far distances only.

It might be of interest to classify the vessels for their shape [67], size, speed, and visibility [11]. Such information may serve as indicator of the type of boat or vessel. Sometimes, crude or fuzzy classification such as very small (foe example, swimmer and debris), small (for example, jet ski, sail

boat, and speed boat), medium (for example, fast boat, fishing boat, and steamers), and large vessels (for example, cruise ship and cargo ship) may suffice. Or, identification of the exact type and model of the vessel may be considered crucial in military or rescue scenarios and surveillance [42], [98]. There are two major approaches to classification and identification, which we discuss below.

The first is the approach of shape library, where segmentations of ground truth may be stored as references in the library and segmented shapes may be compared with the stored shapes to classify the segmented shapes [65]. The shape library must be sufficiently generative to be robust enough for reliable detection and sufficiently discriminative to be specific to the vessel type [215]. For each vessel, shapes at different orientations [30], [98] and spatial resolutions [8] must be stored. Such approaches may require, in addition to the shape library, refined techniques for shape fitting, dominant point detections [216], and shape curvature analysis.

The other is a feature based approach in which a vessel is represented by a set of features discriminatively representing the vessel [98]. This approach requires selection of suitable features for shape classification [215]. SIFT features [33], [34], [98], Haar features [14], Fisher vectors [33], [34], and statistical moments and their variations [30], [33], [34], [67], [77], [98] have been found useful for maritime vessels. It is argued in [98] that simple, though less discriminative, features such as localized moments from the electro-optic images may be sufficient when the object is at large distances and appears only a few pixels wide.

### B. Multisensor Approaches

In this section, we review algorithms in which EO sensor data is processed in conjunction with other types of sensors, such as radar, sonar, gyroscope, and motion sensors. Although the focus of this survey is on EO sensor data processing alone, we believe it is useful to study the effectiveness of employing multiple sensors from other modalities to operate in tandem with EO sensors for tasks such as object detection and tracking. In particular, motion, gyro, and weather sensors can augment the EO data processing pipeline as shown in Fig. 8. Also, radar and sonar can be used to filter outliers in EO data processing and vice versa [14].

In marine environment, [109] combined electro-optical and sonar data stereoscopically to perform 3D reconstruction of floating objects. However, the work used a submerged electro-optical camera. Zheng et. al [108] fused images from visible electro-optical sensor and IR sensor using discrete wavelet transform in an iterative fusion scheme to generate fused and pseudocolored images which have more information than individual sensors alone. Van den Broek et. al [98] presented a system architecture for multi-sensor data association. Specifically, radar or other spectral detectors were used to locate the ships and then zoomed-in images from visible or IR cameras were used to classify the ships based on pre-learnt discriminative feature libraries of the ships. Robert-Inacio et. al [44] combined data from a high definition color image and a 3-camera IR imaging system for video surveillance at a shore,

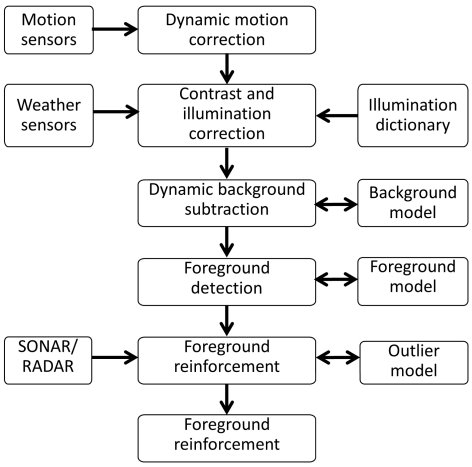


Fig. 8. Flowchart of a multi-sensor processing system.

with the particular intention of detecting terrorist threats. It showed that IR data has higher suppression of wake and thus enables better background detection, which was followed by high definition color data processing for threat analysis, as discussed in section IV-D.1. Caspi and Irani [107] demonstrated that video sequences in a port scenario from a visible range and an IR sensor could be aligned by identifying video tracking features in each sequence and then finding point-to-point correspondences between the two sets of features. It may work well for images with similar spatial resolutions, but fail if the resolution scales of the two sensors are quite different.

Zhang et. al [66] proposed to use rainfall radar, an in-situ multi-probe system equipped with turbidity, dissolved oxygen, temperature, conductivity and depth probes, and a visible sensor for a port-monitoring scenario. It reported that ships entering the port often coincide with spikes in data from the turbidity sensor, which are absent for small vessels. Schwering et. al [36] discussed extensively the design requirements and challenges of an integrated electro-optic system which uses several multiband, hyperspectral, IR, visible range, as well as radar sensors to provide a reliable shore system with tracking, monitoring, and surveillance capabilities. Furthermore, an architecture of using EO sensor for maritime vehicle traffic system in populated areas is presented in [43] and a practical account of maritime multi-sensor experiments is reported in [217].

### C. Commercial Maritime Systems

Here, we discuss two commercial maritime systems, namely, a Vessel Identification and Positioning System (VIPS) and a patent on anti-collision warning system. Both use multisensor approach and the EO sensor is used as a part of larger scheme.

*1) Vessel Identification and Positioning System of Stratech Group Limited:* VIPS is an integrated on-shore sensor system developed by Stratech Group Limited[6] for locating maritime vessels, estimating their heights and widths, and tracking them. The system uses electronic data from automatic identification

[6]http://www.thestratechgroup.com/iv_vips.asp

system (AIS), radar, and EO sensors. The AIS and radar video provide information of potential locations of the vessels. Pre-calibrated electro-optical sensor system then zooms into the vicinity of coordinates provided by the AIS and the radar. It does so by segmenting small image regions around the coordinates. The segmented region is tracked as well as processed to derive height and width information [218].

*2) Anti-Collision Warning System for Marine Vehicle:* A patent [219] approved in 2010 proposed a vessel-mounted anti-collision warning system which uses the EO sensors coupled to the compass as the main data source, which is augmented by AIS and radar for foreground reinforcement (see for example, the multi-sensor flowchart in Fig. 8). The method first detects the horizon and looks for an object close to horizon. Once an object is located, it chooses a small image region around it and performs back ground subtraction using single image statistics (specifically, average intensity thresholding). Using the segmented shape, pre-calibrated EO-sensor grid, and the compass information, the azimuth of the object with reference to the vessel and its approximate height are computed. Also, an after-glow pattern (change in intensity of the segmented shape with time) is computed. The temporal characteristics of the azimuth, the size, and the after-glow are compared with the reference visible objects' database and dangerous objects' database to determine whether an anti-collision warning should be generated. The comparison with the reference databases is done every 30 seconds and the history of an object is maintained for 20 minutes. The azimuth and size information can be checked against the AIS and radar system, or radar and AIS azimuthal tracks can be used instead of EO generated tracks.

## ACKNOWLEDGEMENT

The Singapore Maritime Dataset is available at https://sites.google.com/site/dilipprasad/home/singapore-maritime-dataset.

## REFERENCES

[1] T. Porathe, J. Prison, and Y. Man, "Situation awareness in remote control centres for unmanned ships," in *Human Factors in Ship Design & Operation*. London, U.K.: 2014, p. 93.

[2] A. Elfes, "Sonar-based real-world mapping and navigation," *IEEE J. Robot. Autom.*, vol. 3, no. 3, pp. 249–265, Jun. 1987.

[3] R. E. Hansen, "Synthetic aperture sonar technology review," *Marine Technol. Soc. J.*, vol. 47, no. 5, pp. 117–127, 2013.

[4] J. K. Horne, "Acoustic approaches to remote species identification: A review," *Fisheries Oceanogr.*, vol. 9, no. 4, pp. 356–371, Dec. 2000.

[5] M. P. Hayes and P. T. Gough, "Synthetic aperture sonar: A review of current status," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 207–224, Jul. 2009.

[6] K. D. Ward, C. J. Baker, and S. Watts, "Maritime surveillance radar. I. Radar scattering from the ocean surface," *IEE Proc. F, Radar Signal Process.*, vol. 137, no. 2, pp. 51–62, Apr. 1990.

[7] S. Watts, C. J. Baker, and K. D. Ward, "Maritime surveillance radar. Part 2: Detection performance prediction in sea clutter," *IEE Proc. F, Radar Signal Process.*, vol. 137, no. 2, pp. 63–72, 1990.

[8] R. Vicen-Bueno, R. Carrasco-Alvarez, M. P. Jarabo-Amores, J. C. Nieto-Borge, and M. Rosa-Zurera, "Ship detection by different data selection templates and multilayer perceptrons from incoherent maritime radar data," *IET Radar, Sonar Navigat.*, vol. 5, no. 2, pp. 144–154, Feb. 2011.

[9] G. Pasquariello, G. Satalino, V. la Forgia, and F. Spilotros, "Automatic target recognition for naval traffic control using neural networks," *Image Vis. Comput.*, vol. 16, no. 2, pp. 67–73, Feb. 1998.

[10] A. M. Ponsford, L. Sevgi, and H. C. Chan, "An integrated maritime surveillance system based on high-frequency surface-wave radars. 2. Operational status and system performance," *IEEE Antennas Propag. Mag.*, vol. 43, no. 5, pp. 52–63, Oct. 2001.

[11] Z. L. Szpak and J. R. Tapamo, "Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6669–6680, Jun. 2011.

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

[13] D. D. Bloisi, A. Pennisi, and L. Iocchi, "Background modeling in the maritime domain," *Mach. Vis. Appl.*, vol. 25, no. 5, pp. 1257–1269, Jul. 2014.

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

[15] S. Fefilatyev, V. Smarodzinava, L. O. Hall, and D. B. Goldgof, "Horizon detection using machine learning techniques," in *Proc. Int. Conf. Mach. Learn. Appl.*, Dec. 2006, pp. 17–21.

[16] S. Fefilatyev, D. Goldgof, and C. Lembke, "Tracking ships from fast moving camera through image registration," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3500–3503.

[17] T. Y. van Valkenburg-van Haarst and K. A. Scholte, "Polynomial background estimation using visible light video streams for robust automatic detection in a maritime environment," *Proc. SPIE*, vol. 7482, pp. 748209-1–748209-8, Sep. 2009.

[18] W.-C. Hu, C.-Y. Yang, and D.-Y. Huang, "Robust real-time ship detection and tracking for visual surveillance of cage aquaculture," *J. Vis. Commun. Image Represent.*, vol. 22, no. 6, pp. 543–556, Aug. 2011.

[19] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun./Jul. 2004, pp. 302–309.

[20] L. Ren, C. Shi, and R. Xin, "Target detection of maritime search and rescue: Saliency accumulation method," in *Proc. Int. Conf. Fuzzy Syst. Knowl. Discovery*, May 2012, pp. 1972–1976.

[21] D. Socek, D. Culibrk, O. Marques, H. Kalva, and B. Furht, "A hybrid color-based foreground object detection method for automated marine surveillance," in *Proc. 7th Int. Conf. Adv. Concepts Intell. Vis. Syst.*, 2005, pp. 340–347.

[22] R. Strickland and H. I. Hahn, "Wavelet transform methods for object detection and recovery," *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 724–735, May 1997.

[23] T. Sumimoto *et al.*, "Machine vision for detection of the rescue target in the marine casualty," in *Proc. Int. Conf. Ind. Electron., Control Instrum.*, vol. 2. Sep. 1994, pp. 723–726.

[24] Y. Wang *et al.*, "Aquatic debris monitoring using smartphone-based robotic sensors," in *Proc. Int. Symp. Inf. Process. Sensor Netw.*, 2014, pp. 13–24.

[25] Y. Wang, D. Wang, Q. Lu, D. Luo, and W. Fang, "Aquatic debris detection using embedded camera sensors," *Sensors*, vol. 15, no. 2, pp. 3116–3137, 2015.

[26] H. Wei, H. Nguyen, P. Ramu, C. Raju, X. Liu, and J. Yadegar, "Automated intelligent video surveillance system for ships," *Proc. SPIE*, vol. 7306, pp. 73061N-1–73061N-12, May 2009.

[27] J. Zhou, H. Lv, and F. Zhou, "Infrared small target enhancement by using sequential top-hat filters," *Proc. SPIE*, vol. 9301, pp. 93011L-1–93011L-5, Nov. 2014.

[28] E. Gershikov, T. Libe, and S. Kosolapov, "Horizon line detection in marine images: Which method to choose?" *Int. J. Adv. Intell. Syst.*, vol. 6, nos. 1–2, pp. 1–10, 2013.

[29] B. Bhanu and R. D. Holben, "Model-based segmentation of FLIR images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 1, pp. 2–11, Jan. 1990.

[30] S. P. van den Broek, H. Bouma, and M. A. C. Degache, "Discriminating small extended targets at sea from clutter and other classes of boats in infrared and visual light imagery," *Proc. SPIE*, vol. 6969, pp. 69690B-1–69690B-12, Apr. 2008.

[31] H. Bouma, D.-J. J. de Lange, S. P. van den Broek, R. A. Kemp, and P. B. W. Schwering, "Automatic detection of small surface targets with electro-optical sensors in a harbor environment," *Proc. SPIE*, vol. 7114, pp. 711402-1–711402-8, Oct. 2008.

[32] S. P. van den Broek, E. J. Bakker, D.-J. de Lange, and A. Theil, "Detection and classification of infrared decoys and small targets in a sea background," *Proc. SPIE*, vol. 4029, pp. 70–80, Jul. 2000.

[33] S. P. van den Broek, H. Bouma, R. J. M. den Hollander, H. E. T. Veerman, K. W. Benoist, and P. B. W. Schwering, "Ship recognition for improved persistent tracking with descriptor localization and compact representations," *Proc. SPIE*, vol. 9249, pp. 92490N-1–92490N-11, Oct. 2014.

[34] S. P. van den Broek, H. Bouma, H. E. T. Veerman, K. W. Benoist, R. J. M. den Hollander, and P. B. W. Schwering, "Recognition of ships for long-term tracking," *Proc. SPIE*, vol. 9091, pp. 909107-1–909107-12, Jun. 2014.

[35] H. Chen, H. Zhang, J. Li, D. Yuan, and M. Sun, "Real-time automatic small infrared target detection using local spectral filtering in the frequency," *Proc. SPIE*, vol. 9273, pp. 92730E-1–92730E-7, Nov. 2014.

[36] P. B. W. Schwering, S. P. van den Broek, and M. van Iersel, "EO system concepts in the littoral," *Proc. SPIE*, vol. 6542, pp. 654230-1–654230-12, May 2007.

[37] A. Smith and M. Teal, "Identification and tracking of maritime objects in near-infrared image sequences for collision avoidance," in *Proc. Int. Conf. Image Process. Appl.*, Jul. 1999, pp. 250–254.

[38] D. Tang, G. Sun, D.-H. Wang, Z.-D. Niu, and Z.-P. Chen, "Research on infrared ship detection method in sea-sky background," *Proc. SPIE*, vol. 8907, pp. 89072H-1–89072H-10, Sep. 2013.

[39] X. Tu and J. Chen, "Infrared image segmentation by combining fractal geometry with wavelet transformation," *Sensors Transducers*, vol. 182, no. 11, pp. 230–236, Nov. 2014.

[40] Z. Wang, J. Tian, J. Liu, and S. Zheng, "Small infrared target fusion detection based on support vector machines in the wavelet domain," *Opt. Eng.*, vol. 45, no. 7, pp. 076401-1–076401-9, Jul. 2006.

[41] X. Wang and T. Zhang, "Clutter-adaptive infrared small target detection in infrared maritime scenarios," *Opt. Eng.*, vol. 50, no. 6, pp. 067001-1–067001-12, Jun. 2011.

[42] P. J. Withagen, K. Schutte, A. M. Vossepoel, and M. G. Breuers, "Automatic classification of ships from infrared (FLIR) images," *Proc. SPIE*, vol. 3720, pp. 180–187, Jul. 1999.

[43] D. D. Bloisi, L. Iocchi, D. Nardi, and M. Fiorini, "Integrated visual information for maritime surveillance," in *Clean Mobility and Intelligent Transport Systems*. Edison, NJ, USA: IET, 2015.

[44] F. R. Inacio and A. Raybaud, "Multispectral target detection and tracking for seaport video surveillance," *Proc. Image Vis. Comput.*, New Zealand, 2007, pp. 169–174.

[45] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabally, 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.

[46] C. C. Chen, "Attenuation of electromagnetic radiation by haze, fog, clouds, and rain," Defense Technical Information Center, U.S. Department of Defense, Tech. Rep., 1975. [Online]. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a011642.pdf

[47] J. Park, J. Kim, and N.-S. Son, "Passive target tracking of marine traffic ships using onboard monocular camera for unmanned surface vessel," *Electron. Lett.*, vol. 51, no. 13, pp. 987–989, Jun. 2015.

[48] H. Wang and Z. Wei, "Stereovision based obstacle detection system for unmanned surface vehicle," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 2013, pp. 917–921.

[49] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[50] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, NOIDA, India: Pearson Education, 2009.

[51] G.-Q. Bao, S.-S. Xiong, and Z.-Y. Zhou, "Vision-based horizon extraction for micro air vehicle flight control," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 3, pp. 1067–1072, Jun. 2005.

[52] C. Demonceaux, P. Vasseur, and C. Pégard, "Omnidirectional vision on UAV for attitude computation," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2006, pp. 2842–2847.

[53] S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M. Waszak, "Towards flight autonomy: Vision-based horizon detection for micro air vehicles," in *Proc. Florida Conf. Recent Adv. Robot.*, 2002, pp. 617–640.

[54] S. Todorovic and M. C. Nechyba, "A vision system for horizon tracking and object recognition for micro air vehicles," in *Proc. Florida Conf. Recent Adv. Robot.*, 2004, pp. 1–8.

[55] Y. Sheng, X. Yang, D. McReynolds, Z. Zhang, L. Gagnon, and L. Sevigny, "Real-world multisensor image alignment using edge focusing and Hausdorff distances," *Proc. SPIE*, vol. 3719, pp. 173–184, Mar. 1999.

[56] S. Todorovic and M. C. Nechyba, "A vision system for intelligent mission profiles of micro air vehicles," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1713–1725, Nov. 2004.

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

[58] B. M. H. Romeny, *Front-End Vision and Multi-Scale Image Analysis: Multi-Scale Computer Vision Theory and Applications, Written in Mathematica*, vol. 27. Dordrecht, The Netherlands: Springer, 2003.

[59] S. Fefilatyev, "Algorithms for visual maritime surveillance with rapidly moving camera," Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. South Florida, Tampa, FL, USA, 2012.

[60] 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 Eng.*, vol. 54, pp. 1–12, Nov. 2012.

[61] D. K. Prasad, D. Rajan, C. K. Prasath, L. Rachmawati, E. Rajabaly, and C. Quek, "MSCM-LiFe: Multi-scale cross modal linear feature for horizon detection in maritime images," in *Proc. IEEE TENCON*, Singapore, Nov. 22–25, 2016.

[62] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "MuSCoWERT: Multi-scale consistence of weighted edge Radon transform for horizon detection in maritime images," *J. Opt. Soc. Amer. A*, vol. 33, no. 12, pp. 2491–2500, 2016.

[63] 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.

[64] Z. Yao, "Small target detection under the sea using multi-scale spectral residual and maximum symmetric surround," in *Proc. Int. Conf. Fuzzy Syst. Knowl. Discovery*, Jul. 2013, pp. 241–245.

[65] D. Frost and J.-R. Tapamo, "Detection and tracking of moving objects in a maritime environment using level set with shape priors," *EURASIP J. Image Video Process.*, vol. 2013, no. 1, pp. 1–16, 2013.

[66] D. Zhang, E. O'Connor, K. McGuinness, N. E. O'Connor, F. Regan, and A. Smeaton, "A visual sensing platform for creating a smarter multi-modal marine monitoring network," in *Proc. ACM Int. Workshop Multimedia Anal. Ecol. Data*, 2012, pp. 53–56.

[67] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010.

[68] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," in *Proc. Int. Conf. Pattern Recognit.*, vol. 4. Sep. 2000, pp. 627–630.

[69] M. Mason and Z. Duric, "Using histograms to detect and track objects in color video," in *Proc. Appl. Imag. Pattern Recognit. Workshop*, Oct. 2001, pp. 154–159.

[70] P. Meer, C. V. Stewart, and D. E. Tyler, "Robust computer vision: An interdisciplinary challenge," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 1–7, 2000.

[71] O. Gal, "Automatic obstacle detection for USV's navigation using vision sensors," in *Robotic Sailing*, A. Schlaefer and O. Blaurock, Eds. Berlin, Germany: Springer, 2011, pp. 127–140.

[72] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[73] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[74] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal Saliency detection using phase spectrum of quaternion Fourier transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

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

[76] 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.

[77] K. M. Gupta, D. W. Aha, R. Hartley, and P. G. Moore, "Adaptive maritime video surveillance," *Proc. SPIE*, vol. 7346, pp. 734609-1–734609-14, Apr. 2009.

[78] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in *Proc. ACM Int. Conf. Multimedia*, 2003, pp. 2–10.

[79] P. Westall, J. J. Ford, P. O'Shea, and S. Hrabar, "Evaluation of maritime vision techniques for aerial search of humans in maritime

environments," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl.*, Dec. 2008, pp. 176–183.

[80] D. Casasent and A. Ye, "Detection filters and algorithm fusion for ATR," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 114–125, Jan. 1997.

[81] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," *Proc. SPIE*, vol. 3809, pp. 74–83, Oct. 1999.

[82] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. *The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results*, accessed on: Nov. 30, 2016. [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html

[83] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[84] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[85] P. Voles, A. A. W. Smith, and M. K. Teal, "Nautical scene segmentation using variable size image windows and feature space reclustering," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 324–335.

[86] T. Brox, A. Bruhn, and J. Weickert, "Variational motion segmentation with level sets," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 471–483.

[87] N. Sang and T. Zhang, "Segmentation of FLIR images by target enhancement and image model," *Proc. SPIE*, vol. 3545, pp. 274–277, Sep. 1998.

[88] J. Barnett, "Statistical analysis of median subtraction filtering with application to point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1050, pp. 10–18, Jun. 1989.

[89] P. Voles, M. Teal, and J. Sanderson, "Target identification in a complex maritime scene," in *Proc. IEE Colloq. Motion Anal. Tracking*, May 1999, pp. 15/1–15/4.

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

[91] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.

[92] A. L. Rankin, L. H. Matthies, and A. Huertas, "Daytime water detection by fusing multiple cues for autonomous off-road navigation," in *Proc. Transformat. Sci. Technol. Current Future Force, (With CD-ROM)*, 2004, p. 177.

[93] A. Sobral, "BGSLibrary: An OpenCV C++ background subtraction library," in *Proc. 9th Workshop Vis. Comput. (WVC)*, 2013, pp. 1–16. [Online]. Available: https://github.com/andrewssobral/bgslibrary

[94] 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.

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

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

[97] B. D. Lucas *et al.*, "An iterative image registration technique with an application to stereo vision," in *Proc. IJCAI*, vol. 2. 1981, pp. 674–679.

[98] S. P. van den Broek, P. B. W. Schwering, K. D. Liem, and R. Schleijpen, "Persistent maritime surveillance using multi-sensor feature association and classification," *Proc. SPIE*, vol. 8392, p. 83920O-1–83920O-11, May 2012.

[99] D. Angelova and L. Mihaylova, "Extended object tracking using Monte Carlo methods," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 825–832, Feb. 2008.

[100] J. Zhong and S. Sclaroff, "Segmenting foreground objects from a dynamic textured background via a robust Kalman filter," in *Proc. Int. Conf. Comput. Vis.*, 2003, pp. 44–50.

[101] D. Koller, J. Weber, and J. Malik, *Robust Multiple Car Tracking With Occlusion Reasoning*. Berlin, Germany: Springer, 1994.

[102] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Proc. Eur. Conf. Comput. Vis.*, 1996, pp. 343–356.

[103] D. B. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.

[104] I. J. Cox and S. L. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 2, pp. 138–150, Feb. 1996.

[105] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–10, 2008.

[106] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistics Quart.*, vol. 2, nos. 1–2, pp. 83–97, Mar. 1955.

[107] Y. Caspi and M. Irani, "Spatio-temporal alignment of sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1409–1424, Nov. 2002.

[108] Y. Zheng, K. Agyepong, and O. Kuljaca, "Multisensory data exploitation using advanced image fusion and adaptive colorization," *Proc. SPIE*, vol. 6968, p. 69681U, Apr. 2008.

[109] M. Babaee and S. Negahdaripour, "3-D object modeling from 2-D occluding contour correspondences by opti-acoustic stereo imaging," *Comput. Vis. Image Understand.*, vol. 132, pp. 56–74, Mar. 2015.

[110] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," in *Proc. IEEE Workshop Appl. Comput. Vis., (WACV)*, 1998, pp. 214–219.

[111] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features*. Pittsburgh, PA, USA: Carnegie Mellon Univ., 1991.

[112] H. Possegger, T. Mauthner, and H. Bischof, "In defense of color-based model-free tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2113–2120.

[113] X. Li, K. Wang, W. Wang, and Y. Li, "A multiple object tracking method using Kalman filter," in *Proc. IEEE Int. Conf. Inf. Autom.*, Jun. 2010, pp. 1862–1866.

[114] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, 1994.

[115] 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 Comput. Sci.*, vol. 1, no. 1, pp. 32–54, 2008.

[116] A. Cavallaro and T. Ebrahimi, "Video object extraction based on adaptive background and statistical change detection," *Proc. SPIE*, vol. 4310, pp. 465–475, Dec. 2000.

[117] D. Koller *et al.*, "Towards robust automatic traffic scene analysis in real-time," in *Proc. Int. Conf. Pattern Recognit.*, vol. 1. 1994, pp. 126–131.

[118] A. El Maadi and X. Maldague, "Outdoor infrared video surveillance: A novel dynamic technique for the subtraction of a changing background of IR images," *Infr. Phys. Technol.*, vol. 49, no. 3, pp. 261–265, 2007.

[119] B. Shoushtarian and H. E. Bez, "A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking," *Pattern Recognit. Lett.*, vol. 26, no. 1, pp. 5–26, 2005.

[120] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. 7th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Sep. 1999, pp. 255–261.

[121] C. Ridder, O. Munkelt, and H. Kirchner, "Adaptive background estimation and foreground detection using Kalman-filtering," in *Proc. Int. Conf. Recent Adv. Mechatronics*, 1995, pp. 193–199.

[122] D.-S. Pham, O. Arandjelović, and S. Venkatesh, "Detection of dynamic background due to swaying movements from motion features," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 332–344, Jan. 2015.

[123] S. Varadarajan, P. Miller, and H. Zhou, "Region-based mixture of Gaussians modelling for foreground detection in dynamic scenes," *Pattern Recognit.*, vol. 48, no. 11, pp. 3488–3503, 2015.

[124] H. Chen and P. Meer, "Robust computer vision through kernel density estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 236–250.

[125] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory Probab. Appl.*, vol. 14, no. 1, pp. 153–158, 1969.

[126] J. Kato, T. Watanabe, S. Joga, J. Rittscher, and A. Blake, "An HMM-based segmentation method for traffic monitoring movies," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1291–1296, Sep. 2002.

[127] M. G. Ross, "Exploiting texture-motion duality in optical flow and image segmentation," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2000.

[128] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[129] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, 1981.

[130] X. Li and C. Xu, "Moving object detection in dynamic scenes based on optical flow and superpixels," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 84–89.

[131] I. Haritaoglu, D. Harwood, and L. S. Davis, "$W_4$: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 809–830, Aug. 2000.

[132] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *Proc. Int. Conf. Image Process.*, vol. 5. Oct. 2004, pp. 3061–3064.

[133] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[134] 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.

[135] J. Shen, T. Fan, M. Tang, Q. Zhang, Z. Sun, and F. Huang, "A biological hierarchical model based underwater moving object detection," *Comput. Math. Methods Med.*, vol. 2014, p. 8, 2014.

[136] B. Zhong *et al.*, "Background subtraction driven seeds selection for moving objects segmentation and matting," *Neurocomputing*, vol. 103, pp. 132–142, Mar. 2013.

[137] L. Lin, Y. Xu, X. Liang, and J. Lai, "Complex background subtraction by pursuing dynamic spatio-temporal models," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3191–3202, Jul. 2014.

[138] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[139] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikäinen, and S. Z. Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1301–1306.

[140] A. B. Chan and N. Vasconcelos, "Layered dynamic textures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1862–1879, Oct. 2009.

[141] A. B. Chan and N. Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 909–926, May 2008.

[142] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," in *Proc. 13th Conf. Uncertainty Artif. Intell.*, 1997, pp. 175–181.

[143] J. Rittscher, J. Kato, S. Joga, and A. Blake, "A probabilistic background model for tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2000, pp. 336–350.

[144] V. Cevher, M. F. Duarte, C. Hegde, and R. Baraniuk, "Sparse signal recovery using Markov random fields," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 257–264.

[145] B. Stenger, V. Ramesh, N. Paragios, F. Coetzee, and J. M. Buhmann, "Topology free hidden Markov models: Application to background modeling," in *Proc. Int. Conf. Comput. Vis.*, vol. 1. 2001, pp. 294–301.

[146] 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.*, 2014, pp. 368–375.

[147] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.

[148] L. Wixson, "Detecting salient motion by accumulating directionally-consistent flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 774–780, Aug. 2000.

[149] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[150] 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, 2008, Art. no. 13.

[151] V. Mahadevan and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171–177, Jan. 2010.

[152] C. Xia, H. Huang, T. Wang, and Z. Lin, "Segmentation of infrared image using fuzzy thresholding via local region analysis," in *Proc. Int. Congr. Image Signal Process.*, Oct. 2012, pp. 706–710.

[153] J. Xia, J. Sun, F. He, and H. Li, "Segmentation of FLIR images based on background suppression," in *Proc. Int. Symp. Intell. Inf. Technol. Appl.*, vol. 3. Dec. 2008, pp. 311–314.

[154] X. Yang, T. Zhang, and Y. Lu, "Method for building recognition from FLIR images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 26, no. 5, pp. 28–33, May 2011.

[155] J.-Y. Chang and J.-L. Chen, "Applying fuzzy logic in the modified single-layer perceptron image segmentation network," *J. Chin. Inst. Eng.*, vol. 23, no. 2, pp. 197–210, 2000.

[156] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background subtraction based on cooccurrence of image variations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2003, pp. 65–72.

[157] F. De la Torre and M. J. Black, "Robust principal component analysis for computer vision," in *Proc. Int. Conf. Comput. Vis.*, vol. 1. Jul. 2001, pp. 362–369.

[158] D.-M. Tsai and S.-C. Lai, "Independent component analysis-based background subtraction for indoor surveillance," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 158–167, Jan. 2009.

[159] W. Kim, C. Jung, and C. Kim, "Spatiotemporal saliency detection and its applications in static and dynamic scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 446–456, Apr. 2011.

[160] P. W. Power and J. A. Schoonees, "Understanding background mixture models for foreground segmentation," in *Proc. Int. Conf. Image Vis. Comput.*, Auckland, New Zealand, 2002, pp. 10–11.

[161] 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.

[162] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *Proc. Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1305–1312.

[163] T. Matsuyama, T. Ohya, and H. Habe, "Background subtraction for non-stationary scenes," in *Proc. Asian Conf. Comput. Vis.*, 2000, pp. 662–667.

[164] C. Guyon, T. Bouwmans, and E.-H. Zahzah, "Foreground detection via robust low rank matrix decomposition including spatio-temporal constraint," in *Proc. Asian Conf. Comput. Vis. Workshops*, 2013, pp. 315–320.

[165] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 155–168.

[166] M. Dikmen and T. S. Huang, "Robust estimation of foreground in surveillance videos by sparse error estimation," in *Proc. Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[167] J. Huang, X. Huang, and D. Metaxas, "Learning with dynamic group sparsity," in *Proc. Int. Conf. Comput. Vis.*, Sep. 2009, pp. 64–71.

[168] J. Mairal, R. Jenatton, F. R. Bach, and G. R. Obozinski, "Network flow algorithms for structured sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1558–1566.

[169] Y. Shen, W. Hu, J. Liu, M. Yang, B. Wei, and C. T. Chou, "Efficient background subtraction for real-time tracking in embedded camera networks," in *Proc. ACM Conf. Embedded Netw. Sensor Syst.*, 2012, pp. 295–308.

[170] M. Heikkilä, M. Pietikäinen, and J. Heikkilä, "A texture-based method for detecting moving objects," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2004, pp. 1–10.

[171] 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.

[172] G. Doretto, D. Cremers, P. Favaro, and S. Soatto, "Dynamic texture segmentation," in *Proc. Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1236–1242.

[173] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*. The Netherlands: Springer, 1998, pp. 355–368.

[174] S. J. Nowlan, *Soft Competitive Adaptation: Neural Network Learning Algorithms Based on Fitting Statistical Mixtures*. Pittsburgh, PA, USA: Carnegie Mellon Univ., 1991.

[175] M. Ostendorf and H. Singer, "HMM topology design using maximum likelihood successive state splitting," *Comput. Speech Lang.*, vol. 11, no. 1, pp. 17–41, Jan. 1997.

[176] M. Brand and V. Kettnaker, "Discovery and segmentation of activities in video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 844–851, Aug. 2000.

[177] Y. Wang, K.-F. Loe, and J.-K. Wu, "A dynamic conditional random field model for foreground and shadow segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 279–289, Feb. 2006.

[178] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[179] Y. Fang, Z. Wang, W. Lin, and Z. Fang, "Video saliency incorporating spatiotemporal cues and uncertainty weighting," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3910–3921, Sep. 2014.
[180] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2502–2514, Aug. 2015.
[181] Y. Xue, X. Guo, and X. Cao, "Motion saliency detection using low-rank and sparse decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2012, pp. 1485–1488.
[182] 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.
[183] 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.
[184] F. E. Baf, T. Bouwmans, and B. Vachon, "Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2009, pp. 60–65.
[185] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.
[186] H. Sajid and S.-C. S. Cheung, "Background subtraction under sudden illumination change," in *Proc. IEEE Int. Workshop Multimedia Signal Process.*, Sep. 2014, pp. 1–6.
[187] H. Sajid and S.-C. S. Cheung, "Background subtraction for static & moving camera," in *Proc. IEEE Int. Conf.*, Sep. 2015, pp. 4530–4534.
[188] X. Cao, L. Yang, and X. Guo, "Total variation regularized RPCA for irregularly moving object detection under dynamic background," *IEEE Trans. Cybern.*, vol. 46, no. 4, pp. 1014–1027, Apr. 2015.
[189] Z. Tu, A. Zheng, E. Yang, B. Luo, and A. Hussain, "A biologically inspired vision-based approach for detecting multiple moving objects in complex outdoor scenes," *Cognit. Comput.*, vol. 7, no. 5, pp. 539–551, 2015.
[190] Y. Zhao, H. Gong, L. Lin, and Y. Jia, "Spatio-temporal patches for night background modeling by subspace learning," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
[191] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee, "Using adaptive tracking to classify and monitor activities in a site," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1998, pp. 22–29.
[192] V. Nair and J. J. Clark, "An unsupervised, online learning framework for moving object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2004, pp. 317–324.
[193] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *Proc. Brit. Mach. Vis. Conf.*, vol. 1, no. 5, p. 6, 2006.
[194] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 234–247.
[195] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2006, pp. 798–805.
[196] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1619–1632, Aug. 2011.
[197] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, and D. G. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 28–39.
[198] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 125–141, 2008.
[199] M. J. Black and A. D. Jepson, "EigenTracking: Robust matching and tracking of articulated objects using a view-based representation," *Int. J. Comput. Vis.*, vol. 26, no. 1, pp. 63–84, 1998.
[200] R. Vidal and Y. Ma, "A unified algebraic approach to 2-D and 3-D motion segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 1–15.
[201] D. Cremers and S. Soatto, "Motion competition: A variational approach to piecewise parametric motion segmentation," *Int. J. Comput. Vis.*, vol. 62, no. 3, pp. 249–265, 2005.
[202] T. Amiaz and N. Kiryati, "Piecewise-smooth dense optical flow via level sets," *Int. J. Comput. Vis.*, vol. 68, no. 2, pp. 111–124, 2006.
[203] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Comput. Vis. Image Understand.*, vol. 63, no. 1, pp. 75–104, 1996.
[204] R. Tron and R. Vidal, "A benchmark for the comparison of 3-D motion segmentation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[205] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *Proc. Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1219–1225.
[206] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 282–295.
[207] P. Ochs and T. Brox, "Object segmentation in video: A hierarchical variational approach for turning point trajectories into dense regions," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 1583–1590.
[208] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, Mar. 2013.
[209] M. Bertalmio, G. Sapiro, and G. Randall, "Morphing active contours," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 7, pp. 733–737, Jul. 2000.
[210] S. Greenberg, R. Yehezkel, Y. Gurevich, and H. Guterman, "NLEBS: Automatic target detection using a unique nonlinear-enhancement-based system in IR images," *Opt. Eng.*, vol. 39, no. 5, pp. 1369–1376, 2000.
[211] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, May 2011.
[212] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.
[213] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2756–2759.
[214] M. Danelljan, F. S. Khan, M. Felsberg, and J. van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1090–1097.
[215] D. K. Prasad, "Survey of the problem of object detection in real images," *Int. J. Image Process.*, vol. 6, no. 6, pp. 441–446, 2012.
[216] D. K. Prasad, M. K. H. Leung, C. Quek, and S.-Y. Cho, "A novel framework for making dominant point detection methods non-parametric," *Image Vis. Comput.*, vol. 30, no. 11, pp. 843–859, 2012.
[217] L. Elkins, D. Sellers, and W. R. Monach, "The autonomous maritime navigation (AMN) project: Field tests, autonomous and cooperative behaviors, data fusion, sensors, and vehicles," *J. Field Robot.*, vol. 27, no. 6, pp. 790–818, 2010.
[218] K. M. D. Chew, "Method and system for surveillance of vessels," U.S. Patent 7 889 232, Jun. 14, 2011.
[219] P. Waquet, "Anti-collision warning system for marine vehicle and anti-collision analysis method," U.S. Patent 7 679 530, Mar. 16, 2010.

**Dilip K. Prasad** received the B.Tech. degree in computer science and engineering from IIT Dhanbad, 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 Research Fellow with the Rolls-Royce@NTU Corporate Laboratory, Singapore. He has authored over 55 internationally peer-reviewed research articles. His research interests include image processing, pattern recognition, 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 Bombay, Mumbai, India. He is an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His 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 is a Staff Technologist with Rolls-Royce Pvt. Ltd., Singapore 797575. She has seven years of research experience in technology related to maritime.

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

**Chai Quek** received the B.Sc. and Ph.D. degrees from Heriot-Watt University, Edinburgh, U.K. He is with the School of Computer Engineering, Nanyang Technological University, Singapore. He has published over 250 international conference and journal papers. His research interests include neurocognitive informatics, biomedical engineering, and computational finance. He is a member of the IEEE Technical Committee on Computational Finance and Economics. He has been invited as a Program Committee Member and Reviewer for several conferences and journals, including IEEE TRANSACTIONS ON NEURAL NETWORKS and IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION.