

## A survey on video detection and tracking of maritime vessels

**RODRIGO DA SILVA MOREIRA<sup>1</sup>, NELSON FRANCISCO FAVILLA EBECKEN<sup>2</sup>,  
ALEXANDRE SOARES ALVES<sup>3</sup>, FRÉDÉRIC LIVERNET<sup>4</sup>, ALINE CAMPILLO-NAVETTI<sup>4</sup>**

<sup>1</sup> Civil Engineering Program

COPPE, Federal University of Rio de Janeiro

Centro de Tecnologia, Bloco B, Cidade Universitária, Rio de Janeiro, Rio de Janeiro

BRAZIL

rocypa@yahoo.com.br

<sup>2</sup> Civil Engineering Program

COPPE, Federal University of Rio de Janeiro

Centro de Tecnologia, Bloco B, Cidade Universitária, Rio de Janeiro, Rio de Janeiro

BRAZIL

Senior member of the IEEE and ACM. nelson@ntt.ufrj.br

<sup>3</sup> Electronic Engineering Department

Naval Research Institute – IpqM

Rua Ipiru, n. 2, Cacúia, Ilha do Governador, Rio de Janeiro, Rio de Janeiro

BRAZIL

asoares@ipqm.mar.mil.br

<sup>4</sup> Techniques Navales

Délégation Générale de l'Armement

Avenue de la Tour Royal, BP 40915, Toulon

FRANCE

frederic.livernet@intradef.gouv.fr, aline.campillo@intradef.gouv.fr

*Abstract:* - Maritime surveillance systems can be employed to increase the security of ports, airports, merchant and war ships against pirates, terrorists or any hostile vessel attacks, to avoid collisions, to control maritime traffic at ports and channels and for coastal and oil platforms defense. Cameras are one of the main sensors of these systems. They are cheap and complement other types of sensors. There are few papers about video maritime surveillance systems present in literature compared with other kinds of video surveillance systems. This survey was motivated by the importance of the subject, to motivate new researches and because there are no surveys about video detection and tracking of marine vehicles or they are not widespread. The paper presents the state of the art algorithms.

*Key-words:* - Maritime surveillance systems, tracking, detection, features, image processing, radars

### 1 Introduction

Video surveillance systems in dynamic environments is one of the most active research topics in computer vision [1], receiving much attention in the last decade [2]. Maritime surveillance can be defined as the effective recognition of all maritime activities that impact the security, the economy or the environment [3]. About

80% of all world trade is carried by sea transport. With the growing use of maritime transport, an increase of pirate attacks, activities such as traffic of prohibited substances, illegal immigration and fishing, terrorist attacks at port areas and collisions between marine vehicles primarily at channels and near the ports and coasts is occurring. It is estimated that the losses due to piracy may reach US\$ 16

billion per year [4]. The attack against civilian and military marine vehicles is one way to hurt the economy and security of a country [5, 6]. The terrorist attack against the U.S. warship Cole DDG 67 occurred at port Aden, Yemen, caused the death of 17 people [5]. The French tanker Limburg also suffered terrorist attack at the Yemen coast. Pirate attacks are very common in Somalia, in the Strait of Malacca and Indonesia [7].

The manual operation of surveillance systems is not efficient due to fatigue, stress and the limited ability of human beings to perform certain tasks, the development of automated systems for maritime surveillance is essential to reduce the occurrence of unwanted events [2, 3, 5 -19].

The use of cameras in maritime surveillance systems has increased [19]. Cameras are essential to assist and supplement the radars and other sensors. They are cheap, flexible [6, 11, 17, 20] and can be installed on almost every platform type [2]. The magnetometer detects vehicles by the change in the magnetic field around the vehicle, but are limited to detect vehicles within walking distance [10]. Low and high frequency radars are expensive, hampered by clutter [10], have blind zones close to the transmitting antenna [6, 11] and detect with low efficiency the vehicles built with non-conductive materials [4, 6, 7].

Efforts have been made worldwide for the development of maritime surveillance systems. The European project AMASS - Autonomous Maritime Surveillance System - was created to develop a surveillance system with FLIR cameras installed on advanced platforms [17]. The AVITRACK system [21] and MAAW - Maritime Activity Analysis Workbench - [5] are surveillance systems based on cameras. The ARGOS system [1] has been active since 2007 and is used to monitor the maritime traffic at Gran waterway in Venice, Italy. The SELEX Sistemi Integrati system integrates the data obtained by cameras and by radars and are operating in Russia, Italy, Poland, China and Panama. Burkle et al. [13] proposed a surveillance system based on cameras installed on different platforms and land bases to increase the system coverage area.

New technologies have emerged allowing the data fusion extracted from different systems and sensors. The cameras are one of the main system components [2, 5]. The AMFIS system [13], the AIS

system- Automatic Identification System - [15], the ASV system- Automatic Sea Vision - [15], the VMS system- Vessel Monitoring System - [2] and the AIVS3 system - Automated Intelligent Video Surveillance system for Ships - [6] are examples of maritime surveillance systems that perform data fusion.

## 2 Components of a video surveillance system

A complete video surveillance system consists of five main components, the initial detector, the image processor, the classifier, the tracker and the behavior analyzer. Figure 1 shows a complete video surveillance system.

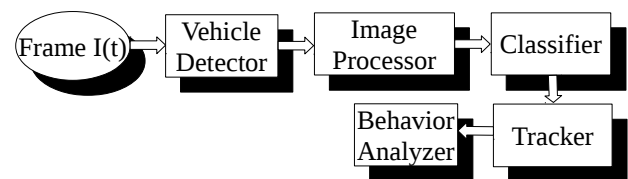


Fig 1. main components of a complete video surveillance system.

Some surveillance systems may not contain all these components. The initial detector is a motion detector that detects all pixels in motion [1, 2, 4, 5, 7, 9, 15, 16] or an object detector based on a classifier set [22]. The information obtained by the initial detector is handled by the image processor to eliminate noise, to segment the most relevant regions and to detect the connected components. These regions are evaluated and classified into objects that are or are not of interest by the classifier. The objects of interest are modeled and are therefore called objects being tracked OT. The tracker attempts to locate the OT in a region of interest ROI at each frame  $I(t)$  and determines the OT position  $P(OT(t))$ . The ROI is the frame region where the probability OT be found is higher. The vehicle trajectory and speed are sent to the behavior analyzer. It generates an alert to a control center if it classify the event as a suspicious activity [3, 5, 6, 16, 18]. The trajectory and speed analysis can also improve the efficiency of the detection and classification of marine vehicles [9].

Marine vehicles do not have particular characteristics that can be used for an efficient classification [12]. It is difficult to construct a representative database for vessel classification due to the variety of marine vehicles types [6, 15].

Although some surveillance systems perform the classification [5, 18, 19], these systems classify a limited number of marine vehicles types and the classification efficiency depends on the position and distance relative to the camera.

### 3 Difficulties

Conventional algorithms for detection and tracking vessels at video, when applied to a maritime environment without proper adjustments, do not produce efficient results, as the background is quite dynamic. The maritime scenario presents challenges that may hinder the initial detector and the tracker. The dynamic and unpredictable ocean appearance makes their mathematical modeling difficult [7, 9]. The images captured by the cameras may not be clear due to the presence of noise and clutter caused by the electronics equipments or by the adverse environmental condition, such as storms, haze and low luminosity [4, 14, 20]. The white foam on the water surface caused by the vehicle propeller or by the waves, the sunlight reflection, the change in lighting conditions, the constant change of each pixel value caused by waves, the presence of objects that float over the ocean, the great variability of certain maritime vehicles features such as size, maneuverability, appearance, geometric shape, the low contrast of the image captured by the cameras or between the marine vehicle and the background and the presence of birds, clouds, fog and aircraft that arises immediately above the horizon hinder the detector and tracker [1, 2, 4, 7, 9, 10, 11, 14, 16, 19]. Figure 2 shows an image with low contrast and clutter. Figure 3 demonstrates the error caused by white foam.

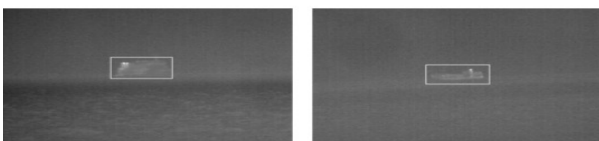


Fig 2. image with low contrast and clutter [18].



Fig 3. white foam generated by the ship [16].

It is common to use FLIR cameras - Forward Looking Infrared - because they are more insensitive to changes in lighting conditions, they do to capture the sunlight reflection over the sea

surface or over the vehicle and they decrease the influence of white foam [9, 10], but they limit the quantity of features that can be extracted [20] and have high energy consumption [2, 10].

Most of the surveillance systems use fixed cameras [15]. Systems based on cameras installed on buoys have to compensate their movement to lower the probability of tracking mistakes [2]. In these cases, the horizon line is used as a reference. Cameras installed on aircrafts or low port marine vessels can produce tracking fails caused by the vibratory camera movement, being necessary to use a smoothing filter [20].

### 4 Horizon line detection

The initial detector usually detects a maritime vessel around the position of the horizon line PHL. After estimating the PHL, the surveillance system detects the maritime vehicles that arises next and above the horizon line, limiting the search region and reducing the execution time of the initial detector [2, 10, 14]. The ROI region can be reduced to the ocean region, below the PHL [6, 15, 19].

Authors like Fefilatyeve et al. [14], Todorovic [23] and Ettinger et al. [24] estimate the PHL by minimizing the intra-class variance of the sky and the sea pixels values. To minimize the influence of the coast and marine vehicles present near the horizon, Fefilatyeve [10] proposed the Unsupervised Slice algorithm. The image is divided into  $N$  parts with  $N-1$  vertical lines evenly distributed. The line segments that minimize the intra-class variance of each part are calculated and combined to estimate the PHL. Fefilatyeve et al. [25] minimize the intra-class variance using features extracted from the pixel values. Cornall et al. [26] estimate the PHL by segmenting the pixels with a threshold. The centroids of the sky and sea pixels define a segment perpendicular to PHL. McGee et al. [27] and Fefilatyeve et al. [25] segment the sky and sea pixels with an SVM classifier. McGee et al. [27] estimate the PHL as the line that separates the sky and sea pixels with less error among all candidate lines. Fefilatyeve et al. [25] define a quantized pixel map  $\{-1,1\}$  according to the classification. The PHL is the line that minimizes the intra-class pixel variance on the map. The surveillance system proposed by Kruger et al. [17] have cameras with inertial units that determine the camera position in space, stabilize the image and reduce the total number of

possible candidate positions for the PHL.

Fefilyatyev et al. [2] discard the frames in which the PHL estimation is unreliable to increase the detector and the tracker robustness. The reliability reduction can occur when the sky or the sea comes out of the camera field of view and when water droplets are deposited on the lens. Considering the hypothesis that the sea and sky pixel values have Normal distributions, Fefilyatyev et al. [2] select a small set of candidate lines with a less robust algorithm based on the Hough transform applied to a gradient map and then select among the candidate lines the one that maximizes an function that indicates the variance between the two classes to accelerate the PHL estimation. Wei et al. [6] apply the Hough transform on a gradient map calculated over the application result of a smoothing filter to the first frame. If the line is not accurately detected, the search region becomes the entire image. Bloisi et al. [19] applies the Hough transform on a gradient image to determine a candidate PHL. The PHL estimation is validated if 90% of sampled pixels above and below the PHL have different values.

The approaches based on the Hough transform [2,6,19] and on optical flow [28] have higher computational complexity.

## 5 Initial vessel detection

An efficient initial detection of the maritime vehicle is important because the performance of all other surveillance system components depend on its performance. Marine environments are very dynamic and difficult to be processed, which can generate a lot of false detections FP and missing vessels FN [10].

The initial detection based only on frame differences can fail in cases where the vehicle is docked or moves toward the camera as little difference between the pixel values at consecutive frames is produced [19]. The ocean pixels values are constantly varying due to the waves, which generates many FP [10]. There are detection algorithms based on the frequency information [29] and on histograms [30], however, recent works use Gaussian functions to model the sea pixel values and detect vehicles by background subtraction [1, 2, 4, 5, 7, 9, 15, 16]. The optical flow analysis is not much used for the initial detection due to the higher computational complexity [7]. Some authors [9, 29]

and divide the image into N regions and extract features, such as entropy, energy, uniformity and contrast of each area. Maritime regions where vehicles are present have different characteristics from the other regions.

The constant movement of the water is one of the factors that cause failures in algorithms based on the background subtraction [4, 6]. The background subtraction statistically exploits the fact that each pixel value follows a normal distribution (equation (1)) or a mixture of normal distributions over time. The probability P of each pixel I(x,y) belong to the ocean or to the vehicle is related to the difference of its value and the mean of each distribution considering the distributions variances (equation (2)).

$$BM(x,y) = N_{(x,y)}(\mu, \sigma) \quad (1)$$

$$P(I(x,y)) \in BM(x,y) \frac{\sigma}{|I(x,y)-\mu|} \quad (2)$$

Many authors have reported that using a background model BM represented by a mixture of Gaussians is less efficient. Szpak and Tapamo [7] conducted statistical tests based on DIP - Departure from Unimodality - and concluded that the pixel values in most cases have a Normal distribution, however, Bloisi et al. [1] reported that a mixture of Normal distributions can represent the ocean better. The right conclusion is that the best representation depends on the application.

Pires et al. [15], Grupta et al. [5] and Robert-Inácio et al. [9] represent BM(x,y,t) by an adaptive Normal distribution. A maritime vehicle is detected whether a connected component area larger than a threshold L is located on the region corresponding to the water surface at the map of relevant pixels MRP. MRP is a map that contains only the pixels that have a low probability to belong to the BM. Pires et al. [15] and Robert-Inácio et al. [9] put in the MRP only the pixels whose value of the difference between I(x,y,t) and  $\mu(x,y,t)$  is greater than a threshold L2. Pires et al. [15] calculates the difference pixel-by-pixel and Robert-Inácio et al. [9] split the image with a regular grid and define I(x,y,t) as the average of the pixels values at each region. Grupta et al. [5] put in the MRP only the pixels whose squared value of the difference between I(x,y,t) and  $\mu(x,y,t)$  divided by  $\sigma(x,y,t)$  is greater than a threshold L3. Hu et al. [16] detect marine vehicles with background subtraction. The initial frames are used to define the BM.

$BM(x,y)$  is the average of the last six  $I(x,y)$  values inserted into a buffer.  $I(x,y,t)$  is inserted into the buffer only if the difference between  $\mu(x,y,t)$  and the average value of the pixel at  $(x,y)$  and its  $3 \times 3$  neighborhood is greater than a threshold  $L$  at  $K$  consecutive frames. Szpak and Tapamo [7] define  $BM(x,y)$  as a Normal distribution initially estimated with the first  $N$  frames and adjusted every frame considering higher weights to more recent frames. The probability of a pixel to belong to a marine vehicle is proportional to the deviation of its value and its neighbors to the interval  $[\mu(x,y,t)-3.\sigma(x,y,t); \mu(x,y,t)+3.\sigma(x,y,t)]$ . At every  $Z$  frames, an active contour starts at the image edges and evolves to the position where a new marine vehicle is. The  $BM$  proposed by Bloisi et al. [1] is a mixture of seven Normal distributions defined by clusterization of the RGB pixel values at  $(x,y)$  contained in the training images. It was chosen seven distributions to represent all possible sea appearances. The vehicle is detected when a connected component has a low probability to belong to the 7 distributions. Wei et al. [6] define  $BM(x,y) = ax + by + c$ . The real values  $a$ ,  $b$  and  $c$  are the ones that minimize a mean squared error function weighted by the pixel values that are below the horizon line. They are updated at each frame. The detection is performed with the search for connected components present at the residue image  $I(x,y,t)-BM(x,y,t)$  segmented by thresholding.

The initial detection validation by a classifier is present in the literature [18, 19], however, due to high variability of the appearance and the geometric shape of marine vehicles, this approach is not very explored. Bloisi et al. [19] proposed an initial detector based on an ensemble classifier trained offline with Haar wavelet features. The ensemble was designed to increase the robustness of initial detection in cases where a vessel is anchored and when sunlight reflections or white foam are present at the sea surface. Teutsch and Kruger [18] train a SVM classifier with the features invariant moments, some statistical measures such as mean and variance, texture analysis, co-occurrence matrices and the gradient analysis to classify vehicles in two steps. At the first step the detected candidates are classified into objects over the ocean or clutter. If it is classified as an object over the ocean, the object is classified as a marine vehicle or an irrelevant object in the second step. Sullivan and M. Shah [31] detect marine vehicles with the similarity value between the result of the FFT transform applied to vehicles

images recorded in a database and the result of the FFT transform applied to candidate regions at each frame. Feineigle et al. [8] detect marine vehicles by the Euclidean distance between SIFT feature points detected at each frame and SIFT feature points present in a image dataset.

Detection algorithms based on connected components localization must consider the vehicle proximity to the camera [4]. Using a camera focused at infinity and installed on a buoy, Fefilatyev [10] detects marine vehicles by exploiting the gradient information of the pixels above the PHL. Morphological operations of erosion and dilation followed by the connected components localization are used to detect a pixel set with high gradient present above the PHL. Figure 4 shows a marine vehicle detection by exploiting the gradient information.

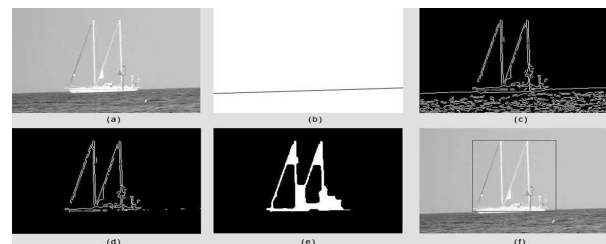


Fig 4. detection of marine vehicles by exploiting the gradient information of the pixels above the PHL [10].

Fefilatyev et al. [14] and Fefilatyev et al. [2] accelerated algorithm proposed by Fefilatyev [10]. They removed the need for morphological operations. The threshold values for the pixel segmentation are obtained by applying the Otsu segmentation method on a gradient map. Frost and Tapamo [4] detect marine vehicles by locating connected components based on segmentation by thresholding applied to a probability map estimated by a Gaussian kernel function. Only connected components with geometric shape similar to pre-defined models are considered marine vehicles.

The use of different and independent features is important to increase the robustness of the initial detector and the tracker to the variability of vehicle and environment appearances. Kruger and Orlov [17] and Teutsch and Kruger [18] combine the result of 3 detectors based on the extraction of distinctive features to determine if a vehicle is present near the PHL. Westall et al. [32] exploit the information provided by different color spaces. At each frame

point in 3 different resolutions are extracted a color histogram  $H$  and a histogram gradient orientations  $HoG$  through integral images to accelerate the extraction [11]. Connected points that have the histograms  $H$  and  $HoG$  different from its neighbor's histograms belong to a marine vehicle. Fefilatyeu [10] compared the efficiency of texture measurements like entropy, average, standard deviation, and moments up to the fourth order calculated with RGB value of each pixel and their neighborhood  $11 \times 11$  pixels normalized to the interval  $[0,1]$ . Applying segmentation by thresholding, the pixels that belong to the sky, to the sea and to marine vehicles are separated into distinct groups. Islam et al. [12] proposed a detector whose initial image  $Q_0$  is blurred by a linear filter to generate the image  $I$ . A Gaussian filter with  $\sigma=1$  and other one with  $\sigma=3$  are applied to  $Q_0$  and  $I$  to form the filtered images  $Q_1$ ,  $Q_3$ ,  $I_1$  and  $I_3$ . The differences  $Q_1-Q_3$  and  $I_1-I_3$  are applied to an anomaly detector to produce the  $A$  and  $B$  images.  $A(x,y)-B(x,y)$  is proportional to the probability of the pixel at  $(x,y)$  to belong to a marine vehicle.

### 5.1 Techniques Used To Lower The Quantity Of FP And FN Detections

The ocean is a dynamic environment that has waves, white foam and light reflections on the water surface, which can generate a considerable FP and FN amounts. Some authors [6, 9] report that the detection and tracking applied on IR images are more efficient because the water temperature is not influenced by these events.

Different methods are employed to decrease the FP and FN rates detections. Many authors [2, 7, 9- 11, 14, 15, 18, 19] only validate the initial detection of a marine vehicle if the tracking result is consistent and reliable at  $N$  consecutive frames. Fefilatyeu [10] and Fefilatyeu et al. [14] only consider an initial detection if the detection is reliable and the centroid and bounding box trajectories of the OT are consistent at 10 of the first 20 frames. Fefilatyeu et al. [2] added to these rules the need of the vehicle appearance to be almost constant at  $N$  consecutive frames and the need for an object to have a considerable size.

To decrease the FP rate caused by ocean waves, birds, aircrafts or objects of negligible size, Fefilatyeu [10] and Grupta et al. [5] consider marine vehicles the connected components located at the

MRP that are distanced from PHL at least  $N$  pixels apart. The contour size of a FP caused by foam, shadows, reflections and waves decreases at every frame to disappear when the active contour tracker based on level set functions proposed by Szpak and Tapamo [7] is applied. To reduce the FP rate, Bloisi et al. [19] use an ensemble classifier trained offline to validate the initial detection. To decrease the FP rate caused by white foam, Frost Frost and Tapamo [4] analyze if each pixel value remains different from the BM value for more than  $N$  consecutive frames. Foam pixels have lower persistence than the vehicle pixels. At the first step, Hu et al. [16] and Grupta et al. [5] eliminate the foam pixels removing small connected components located at the MRP. Hu et al. [16] apply an algorithm that eliminates shadows to reduce its influence. Pixels with high brightness and chromaticity distortion are white foam candidates. The candidate pixels that have brightness variation greater than a threshold are considered white foam.

To decrease the FN and FP rates, some authors apply morphological operations [5, 6, 32]. Grupta et al. [5] apply the erosion, dilation and smoothing operations before looking for connected components. Westall et al. [32] apply the opening, closing, erosion and dilation operations to eliminate noise and decrease the FP rate. Beyond these operations, Wei et al. [6] applied the operations of opening and closing to the residue image  $I(x,y,t) - BM(x,y,t)$  to eliminate clutter.

## 6 Maritime Vehicle Tracking

There are many object tracking methods in the literature. The mean-shift, successive clustering, active contour and template matching are the most used methods in marine environments. The use of Kalman filter [33] as an estimator produces good tracking applications results because the vehicle movement is not too complex [18].

### 6.1 Kalman Filter

The Kalman filter KF [33] is an optimal estimation method of the state of a stochastic, non-stationary, dynamic and linear process. Kalman [33] introduced the representation of linear dynamical systems by state equations. The process is governed by discrete and linear equations (equations (3) and (4)) [20].

$$x(t+1) = A \cdot x(t) + B \cdot u(t) + w(t) \quad (3)$$

$$z(t) = C.x(t) + D.u(t) + v(t) \quad (4)$$

Where  $x$  is the process state vector, which may contain variables related to the object translation, scale and orientation and its first and second order derivatives,  $u$  is the control vector,  $z$  is the measurement vector obtained by a tracking algorithm,  $A$  is the state transition matrix,  $B$  is the state control matrix,  $C$  is the observation matrix,  $D$  is the measurement control matrix,  $w$  is the noise associated with the state and  $v$  is the noise associated with the measure. By hypothesis,  $w$  and  $v$  noise vectors are independent and have Gaussian multivariate probability distribution functions of zero mean and diagonal covariance matrix  $Q$  and  $R$  respectively ( $w \sim N(0, Q)$  and  $v \sim N(0, R)$ ).

KF is a recursive algorithm that consists of two phases: time update and measurement update. The time update phase (equations (5) and (6)) estimates the state vector  $x(t|t-1)$  value and the error matrix  $P(t|t-1)$  value considering the observations obtained at  $I(t-1)$ .

$$x(t|t-1) = A.x(t-1|t-1) \quad (5)$$

$$P(t|t-1) = E((x(t) - x(t|t-1)).(x(t) - x(t|t-1))^T) \quad (6)$$

$$= A.P(t-1|t-1).A + Q$$

Where  $x(t)$  is the state at frame  $t$ ,  $x(t|t-1)$  and  $x(t-1|t-1)$  are a priori and a posteriori estimation of the state vector,  $P(t|t-1)$  and  $P(t-1|t-1)$  are the a priori and a posteriori estimation of the error matrix and  $E$  is the expected value.

The measurement update phase (equations (7), (8) and (9)) corrects the  $x(t|t-1)$  and  $P(t|t-1)$  values by incorporating the  $z(t)$  measurement obtained by the tracker at each frame.

$$K(t) = P(t|t-1).C^T.(C.P(t|t-1).C^T + R)^{-1} \quad (7)$$

$$x(t|t) = x(t|t-1) + K(t).(z(t) - C.x(t|t-1)) \quad (8)$$

$$P(t|t) = P(t|t-1) - K(t).C.P(t|t-1) \quad (9)$$

Where  $K(t)$  is the Kalman gain at frame  $t$ .

A very common application of KF is the prediction of each object position at frame  $t+1$  to define the ROI position [1, 2, 5, 6, 10, 14, 15, 18].

## 6.2 Successive Clustering

The clustering applied to successive frames is one of

the simplest tracking methods [9]. An image segmentation algorithm is applied at each frame image to generate a probability map. Then, a clustering algorithm forms the connected components in the map.  $P(OT(t))$  is usually considered the centroid position of the connected component that is nearest to the OT centroid position estimated by the KF.

The surveillance system ASV [15] determines the vehicle spatial position geometrically by considering the camera height, the vehicle pixel closest to the water surface and the PHL. The tracking is based on successive clustering.  $P(OT(t))$  is determined by associating the bounding box positions and centroid velocities estimated by the KF for each vehicle and the ones calculated for each connected component. Fefilyatyev [10], Fefilyatyev et al. [14] and Fefilyatyev et al. [2] track marine vehicles by applying the Kalman filter and successive clustering at each frame. When one OT is not detected within the ROI estimated by the KF, the OT is considered occluded and the OT model is not updated, but the KF continues estimating future states of the OT bounding box and centroid. Bloisi et al. [1] and Gupta et al. [5] group together the clusters being tracked that are close to each other and have similar movements in a single OT. Gupta et al. [5] segment the image by background subtraction and analyze only the proximity and movement of the centroids. Bloisi et al. [1] segment the image by analyzing the optical flow similarity of the pixels and cluster the neighbor segments with a K-means algorithm. The optical flow is a dense field vector of displacements that define the translation of the pixels at successive frames. The OT movement can be estimated by analyzing the optical flow of the OT pixels. The optical flow is a vector that indicates the displacement of the pixels between successive frames and its calculation is performed considering the hypothesis that the brightness of the pixels at successive frames do not vary abruptly [34] (equation (10)).

$$I(x, y, t) - I(x + \Delta x, y + \Delta y, t + \Delta t) = 0 \quad (10)$$

A high frame per second rate is required to secure this hypothesis. The equation that connects the optical flow vector  $V = (\partial x / \partial t, \partial y / \partial t)^T$  and the first order intensity derivative (equation (11)) is deduced by Taylor series expansion up to the first order term (equation (10)) [35].

$$(\nabla I(x,y,t))^T \cdot V + \frac{\partial I(x,y,t)}{\partial t} = 0 \quad (11)$$

### 6.3 Mean-Shift

The mean-shift algorithm is a nonparametric clustering technique based on gradient ascent applied to data in the feature space FS. Was initially proposed by Fukunaga and Hostetler [36], and then was adapted by Cheng [37] for image analysis, by Comaniciu and Meer [38] for image segmentation and by Bradski [39] and Comaniciu et al. [40] for object tracking.

The mean-shift algorithm considers the data as points in FS associated with a empirical probability density function, where regions of dense data present in FS correspond to local maximum or modes of the data distribution. A local gradient ascent algorithm is applied to the empirical probability density function to determine the data region corresponding to the mode. Given  $n$  points  $p_i$ ,  $i=1, \dots, n$  in  $R^d$ , the empirical probability density function EPDF( $p$ ) that has a radially symmetric kernel (equation (13)) centralized at  $p$  and has a bandwidth  $h$  is defined by equation (12) [37, 38].

$$EPDF(p) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{p-p_i}{h}\right) \quad (12)$$

$$K(a) = c_k \cdot k\left(\left\|\frac{a}{h}\right\|^2\right) \quad (13)$$

Where  $c_k$  is a normalization constant. The EPDF modes are localized at the points where the gradient of the EPDF is null ( $\nabla EPDF(p)=0$ ).

The OT model is represented by the function EPDFR [40]. Equivalently, a pixel region  $R$  is represented by the function EPDFC. Both functions are estimated by the histograms Hepdfr and Hepdfc (equations (14) and (15)). At each interaction step, the mean-shift vector (equation (17)) shifts  $R$  toward a region of maximum similarity between the histograms calculated by the Taylor series expansion of the Bhattacharyya coefficient (equation (16)). The final  $R$  position is the OT position.

$$Hepdfr_u(p) = C \sum_{i=1}^n k\left(\left\|\frac{p-p_i}{h}\right\|^2\right) \cdot d[b(p_i)-u] \quad (14)$$

$$Hepdfc_u(p) = C_h \sum_{i=1}^{nh} k\left(\left\|\frac{p-p_i}{h}\right\|^2\right) \cdot d[b(p_i)-u] \quad (15)$$

Where  $b$  returns the histogram bin for the pixel  $p_i$ ,  $u$

is a bin,  $d$  is the Kronecker delta function,  $n$  and  $nh$  are the pixel amount at  $p$  neighborhood defined by the kernel  $k$ ,  $C$  and  $C_h$  are normalization constants and  $p_i$  is a neighbor pixel of  $p$ .

$$C_{Bhattacharyya} = \sum_{u=1}^m \sqrt{Hepdfr(u) \cdot Hepdfc(u)} \quad (16)$$

$$m_h(p) = \frac{\sum_{i=1}^n p_i \cdot w_i \cdot g\left(\left\|\frac{p-p_i}{h}\right\|^2\right)}{\sum_{i=1}^n w_i \cdot g\left(\left\|\frac{p-p_i}{h}\right\|^2\right)} - p \quad (17)$$

Where  $g(a) = -k'(a)$  and  $w_i$  is calculated by equation (18).

$$w_i(p) = \sum_{i=1}^n \sqrt{\frac{r_u}{c_u \cdot p}} \cdot d[b(p_i) - u] \quad (18)$$

Where  $r_u$  is the of the bin  $u$  in the OT histogram and  $c_u$  is the value of the bin  $u$  in the R histogram.

Bibby and Reid [41] developed a tracker based on mean-shift algorithm, but his approach fails in cases of total occlusions.

Liu et al. [11] modifie the segmentation threshold by selecting online the most discriminative features with the algorithm proposed by Collins and Liu [42] and applies the mean-shift algorithm starting from the position estimated by KF to determine  $P(OT(t))$ . The feature pool has three color components, three differences between color components and the results of eight transformations applied to the Hue component. The function that measures the discrimination degree is based on the similarity between the histograms of OT and its neighboring pixels.

### 6.4 Template Matching

The template matching in the context of object tracking is defined as the location of a small pixel set called template within the ROI [43]. The OT model is the template to be found within the ROI. Templates are constructed with the pixels inside a simple geometric shape region. The position of the candidate region  $C$  that maximizes the similarity between the OT model  $M$  and all candidates reveal  $P(OT(t))$ . The Hamming distance (equation (19)) [44], the Euclidean distance [45], the Cross Correlation [46], NCC (equation (20)) - Normalized



Cross Correlation - [47], the SSD (equation (21)) - Sum of Squared Difference - [46] and SAD (equation (22)) - Sum of Absolute Difference - [46] are examples of similarity functions between templates. The weightless neural network WiSARD can generate different similarity functions and can be adapted to tracking [48]. The simplest similarity function is the sum of the differences between pixel values of two templates (equation (19)). The values  $dx$  and  $dy$  that minimizes the function determines  $P(OT(t))$ .

$$Dif(M, C_i) = \frac{\sum_{(x,y) \in M} (I(x+dx, y+dy) - M(x,y))}{N_x \cdot N_y} \quad (19)$$

$$NCC(x,y) = \frac{\sum_{(x,y) \in M} (I(x+dx, y+dy) \cdot M(x,y))}{\sqrt{\sum_{(x,y) \in M} M^2(x,y)}} \quad (20)$$

$$SSD(x,y) = \sum_{(x,y) \in M} (M(x,y) - I(x+dx, y+dy))^2 \quad (21)$$

$$L1(x,y) = \sum_{(x,y) \in M} |M(x,y) - I(x+dx, y+dy)| \quad (22)$$

The L1 norm distance (equation (22)) rises the robustness to noise because it generates a lower penalty than the quadratic SSD function penalty. To limit the effects caused by variations in the environment lighting conditions, the normalized SSD can be used in place of the SSD (equation (23)).

$$NSSD(x,y) = \sum_{(x,y) \in M} (A - B)^2 \quad (23)$$

$$A = \frac{M(x,y) - \mu(M(x,y))}{\sigma(M(x,y))}$$

$$B = \frac{I(x+dx, y+dy) - \mu(I(x+dx, y+dy))}{\sigma(I(x+dx, y+dy))}$$

Where  $\mu$  and  $\sigma$  are the average and the standard deviation.

The similarity does not necessarily have to be calculated with the pixel values. Any feature extracted from a pixel region can be used.

Fefilyatyev et al. [14] and Fefilyatyev et al. [2] stabilize the image obtained by the camera installed on a buoy by minimizing the NCC between two images IMG1 and IMG2. IMG1 is the difference between the OT template and the average of the template pixel values. IMG2 is the difference between the frame  $I(t)$  and the average value of the

pixels inside the ROI at  $I(t)$ . The tracking algorithms proposed by Fefilyatyev et al. [14] and Fefilyatyev et al. [2] define  $P(OT(t))$  by the NCC template matching algorithm used to stabilize the camera when the result of the segmentation by thresholding applied to an gradient image is unreliable. If the template matching is also unreliable,  $I(t)$  is discarded. The threshold value is calculated by Otsu method. Moreira and Ebecken [48] proposed a tracker based on the weightless neural network WiSARD. The OT model is stored at the network RAM nodes. Candidate regions of quantized pixels are put at the network input.  $P(OT(t))$  is defined as the position of the region that maximizes the network response. The tracker proposed by Hu et al. [16] defines  $P(OT(t))$  by a template matching algorithm that uses the MAD function (equation (24)) - Median of Absolute Differences - as the similarity function.

$$MAD = \frac{1}{W \cdot H} \sum_{i=0}^W \sum_{j=0}^H |OT(x,y,t) - I(x+i, y+j, t)| \quad (24)$$

Where  $W$  and  $H$  are the length and height of the OT bounding box.

## 6.5 Histogram Matching

The histogram matching is a technique frequently used for tracking objects because the histogram is invariant to rotation and scale transformations applied to the object and it is robust to partial occlusions [49]. The appearance model is defined extracting a histogram with the OT pixels.  $P(OT(t))$  is the frame position that provides the maximum similarity measure between the OT histogram HM and histograms extracted from candidate regions HC (equation (25)).

$$S(HM, HC) = \sum_{j=1}^n (HM(j) - H(j)) \quad (25)$$

Where  $n$  is the total bin quantity and  $H(j)$  is the value of the bin  $j$  of the histogram  $H$ .

Puzicha et al. [50] present other ways of calculating the similarity between histograms as the weighted bin to bin difference, the histogram intersection (equation (26)) and  $\chi^2$ . The log likelihood statistics and log likelihood ratio statistics functions of similarity between histograms have been simplified by Ojala et al. [51] (equation (27)) and (equation

(28)).

$$I(H_{OR}, H_{OC}) = \frac{\sum_{j=1}^n \min(H_{OR}(j) - H_{OC}(j))}{\sum_{j=1}^n (H_{OR}(j))} \quad (26)$$

$$L(H_{OR}, H_{OC}) = \sum_{j=1}^n H_{OC} \cdot \log(H_{OR}(j)) \quad (27)$$

$$L(H_{OR}, H_{OC}) = 2 \cdot \sum_{j=1}^n H_{OC} \cdot \log\left(\frac{H_{OC}(j)}{H_{OR}(j)}\right) \quad (28)$$

The linear approximation of the Bhattacharyya coefficient [40] is the most used similarity function for tracking objects, because it is easily calculated and because there are many authors who reported the success of their application [49].

The tracker proposed by Bloisi et al. [19] determine  $P(OT(t))$  with histograms matching based on the Bhattacharyya coefficient. The pixel values are in the HSV color space to minimize the influence of shadows and lighting variations caused by sunlight reflection over the sea surface. To decrease the quantity of tracking and detection failures, Bloisi et al. [19] proposed the radar and camera data fusion. Fusion occurs in a normalized plane where  $P(OT(t))$  is defined by the nearest neighbor rule. Westall et al. [32] detect the head of missing people at sea using information in RGB, YCbCr, YIQ and HSV color spaces considering by hypothesis that these color spaces are independent.

## 6.6 Active Contour

The active contour tracking method represents the vehicle contour by one or more curves. The curves move dynamically at every frame toward the position of the vehicle edges, which by hypothesis is the place where the discontinuity of the pixel values are higher. Trackers generally use the final contour position at the previous frame as the initial position at the current frame. The main advantage of the active contour is that it is relatively insensitive to lighting variations. Figure 5 shows the active contour evolution.

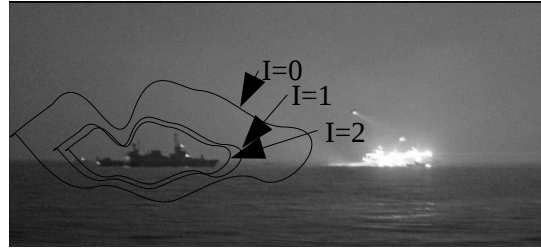


Fig 5. active contour evolution. I is the iteration step number.

Goldenberg et al. [52] describes the mathematical theory related to the parametric and non-parametric active contour methods. There are two ways to represent the object contour: the explicit representation, as is the case of snakes, or implicit representation, such as the level set function [4, 7]. Snakes have not been applied for tracking marine vehicles yet. For this reason, only the tracking based on level set functions will be presented in this paper.

A distance function that implicitly determines the curve  $C$  position is defined by equation (29) [4, 7].

$$C = \{(x,y) | \phi(x,y) = 0\} \quad (29)$$

$C$  is the set of image points whose level set function value is null. Many authors define the function as the Euclidean distance between the point  $(x,y)$  and  $C$  (equation (30)).

$$\phi(x,y) = \begin{cases} -d(x,y), & \text{if } (x,y) \text{ is inside } C \\ 0, & \text{if } (x,y) \text{ is over } C \\ d(x,y), & \text{if } (x,y) \text{ is outside } C \end{cases} \quad (30)$$

Where  $d(x,y)$  is the Euclidean distance between the pixel at  $(x,y)$  and the curve  $C$ .

The curve evolution is defined by the equation (31) [4]. The update of the level set function values at each point generates the implicit curve movement.

$$\frac{d\phi}{dt} = V |\nabla \phi| \quad (31)$$

Where  $V$  is a speed function that depends on the pixel values and is independent of the parametrization [52].  $V$  can be defined as a gradient function [4]. The update of the level set function depends on the  $V$  value.

The level set function that describes the OT contour at each frame moves by minimizing an energy

function. The energy function proposed by Frost and Tapamo [4] is minimized by a gradient descent method. It is composed by a sum of three functions: the color histogram, the FFT transform and statistical measures like entropy, contrast, homogeneity and energy. These functions indicate the difference between the pixel values of the OT model and the pixel values inside the active contour. Szpak and Tapamo [7] applies the active contour method directly on a probability map that estimates the probability of each pixel to be a background pixel. The Chan-Vese energy function was chosen. This function measures the sum of the probability variances of the pixels inside and outside the curve.

### 6.7 Occlusion Handling

Partial and total occlusions may occur. The occlusion can cause a tracking failure. Figure 6 shows an occlusion case. Teutsch and Kruger [18] proposed a tracker that combines 3 different trackers to increase the robustness to partial occlusions. When the response of one or two trackers is unreliable,  $P(OT(t))$  obtained by them receives a lower weight. T1 and T2 trackers are based on pixel regions and T3 is based on feature points extracted by the algorithm proposed by Shi and Tomasi [53]. T1 tracker performs segmentation by adaptive thresholding at each frame  $I(t)$  and defines  $P(OT(t))$  by the nearest neighbor rule applied to the centroids of connected pixel regions present at  $I(t)$  and  $I(t-1)$ . T2 performs the association between blobs extracted at  $I(t)$  and  $I(t-1)$ . T3 performs the association between feature points extracted from the ROI and the OT feature points and defines  $P(OT(t))$  as the average position of each associated feature point. Teutsch and Kruger [18] associate an independent KF for each OT and only update their models when the OT is not occluded. A total occlusion occurs when none of the trackers determines  $P(OT(t))$  with high confidence. In this case, the KF continues estimating  $P(OT(t))$ . If the OT is not detected at  $N$  consecutive frames with high confidence, the reference to the OT is erased.

Bacho et al. [20] model the OT by a grid of size  $31 \times 31$  pixels which has associated the parameters of affine transformations. The tracker is based on a particle filter and defines  $P(OT(t))$  as a weighted average of the positions estimated by each particle. When an occlusion occurs, the variance associated with the state transition matrix of each particle increases to scatter the particles over the space and

to detect the vehicle after the occlusion with greater efficiency.  $P(OT(t))$  is determined by the particle that is more similar to the OT template.

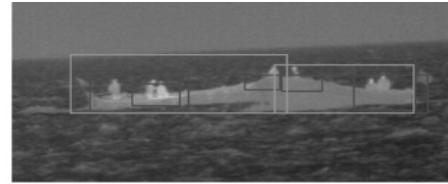


Fig 6. occlusion example [18].

## 7 Conclusion

This paper presented the state of the art methods of video detection and tracking of marine vehicles. The maritime environment is very challenging and dynamic. The algorithms of object detection and tracking, when applied to a maritime environment without proper adjustments, do not produce efficient results. Many errors of detection and tracking may occur due to noise, clutter, waves, dynamic and unpredictable ocean appearance, sunlight reflections, bad environmental conditions, low luminosity and image contrast, presence of objects that float over the ocean, white foam, the great variability of certain maritime vehicles features such as size, maneuverability, appearance, geometric shape and the presence of birds, clouds, fog and aircraft that arises immediately above the horizon.

Video maritime surveillance systems are very important. They can be used to increase the coastal and ship security against hostile vessel attacks, to avoid collisions, to control the maritime traffic at ports and channels and for oil platforms defense.

There are not many researches about video detection and tracking of marine vehicles. The algorithms seem not to perform well in some real situations when little vessels that have low contrast with the background arise in the camera field of view. The video maritime surveillance is still a not complete solved problem and need to be more explored.

## 8 References

- [1] D. Bloisi and L. Iocchi, Argos: A Video Surveillance System for Boat Traffic Monitoring in Venice, *International Journal of Pattern Recognition and Artificial Intelligence*, 2008, pp. 1-24.
- [2] S. Fefilyatyev, D. Goldgof, M. Shceve and C. Lembke, Detection and Tracking of Ships in Open

- Sea with Rapidly Moving Buoy-Mounted Camera System, *Ocean Engineering*, Vol.54, 2012, pp. 1-12.
- [3] S. Kasemi, S. Abghari, N. Lavesson, H. Johnson and P. Ryman, Expert Systems with Applications, *Expert Systems with Applications*, Vol.40, 2013, pp. 5719-5729.
- [4] D. Frost and J.-R. Tapamo, Detection and Tracking of Moving Objects in a Maritime Environment with Level-set with Shape Priors, *EURASIP Journal on Image and Video Processing*, Vol.1, No. 42, 2013, pp. 1-16.
- [5] K. M. Grupta, D. W. Aha, R. Hartley, and P. G. Moore, Adaptive Maritime Video Surveillance, *Proceedings of SPIE*, Vol.7346, No.09, 2009, pp. 1-12.
- [6] H. Wei, H. Nyguen, P. Ramu, C. Raju, X. Liu and J. Yadegar, Automated Intelligence Video Surveillance System for Ships, *Proceedings of SPIE* Vol.7306, No.1N, 2009, pp. 1-12.
- [7] Z. L. Szapak and J. R. Tapamo, Maritime Surveillance: Tracking Ships Inside a Dynamic Background Using a Fast Level-Set, *Expert System with Applications*, Vol.38, 2011, pp. 6669-6680.
- [8] P. A. Feineigle, D. D. Morris and F. D. Snyder, Ship Recognition Using Optical Imagery for Harbor, *Proceedings of Association for Unmanned Vehicle Systems International*, 2007.
- [9] F. Robert-Inácio, A. Raybaud and É. Clément, Multispectral Target Detection and Tracking for Seaport Video Surveillance, *Proceedings of Image and Vision Computing*, 2007, pp. 169-174.
- [10] S. Fefilatyevev, Detection of Marine Vehicles in Images and Videos of Open Sea, *Master thesis (Computational Science and Engineering) – South Florida University, United States, Florida, Tampa*, 2008.
- [11] H. Liu, O. Javed, G. Taylor, X. Cao and N. Haering, Omni-Directional Surveillance for Unmanned Water Vehicle, *8th International Workshop on Visual Surveillance*, 2008.
- [12] M. M. Islam, M. N. Islam, K. V. Asari, and M. A. Karim, Anomaly Based Vessel Detection in Visible and Infrared Images, *Proceedings of SPIE-IS & Electronic Imaging*, Vol.7251, No.0B, 2009, pp. 1-6.
- [13] A. Burkle and B. Essendorfer, Maritime Surveillance with Integrated Systems, *Proceedings of Waterside Security Conference*, 2010.
- [14] S. Fefilatyevev, D. Goldgof and C. Lembke, Tracking Ships from Fast Moving Camera Through Image Registration, *Proceedings of the 20th International Conference on Pattern Recognition*, 2010, pp. 3500-3503.
- [15] N. Pires, J. Guinet and E. Dusch, ASV: An Innovative Automatic System for Maritime Surveillance, *NAVIGATION*, Vol.58, No.232, 2010, pp. 47-66.
- [16] W.-C. Hu, C.-Y. Yang and D.-Y. Huang, Robust Real-Time Ship Detection and Tracking for Visual Surveillance of Cage Aquaculture, *Journal of Visual Communication & Image Representation*, Vol.22, 2011, pp. 543-556.
- [17] W. Kruger and Z. Orlov, Robust Layer-Based Boat Detection and Multi-Target-Tracking in Maritime Environments, *Proceedings of Waterside Security Conference*, 2010.
- [18] M. Teutsch and W. Kruger, Classification of Small Boats in Infrared Images for Maritime Surveillance, *Proceedings of Waterside Security Conference*, 2010.
- [19] D. Bloisi, L. Locchi, M. Fiorini and G. Grasio, Automatic Maritime Surveillance with Visual Target Detection, *CiteSeerX Scientific Literature Digital Library and Search Engine*, 2011.
- [20] A. K. Bacho, F. Roux and F. Nicolls, An Optical Tracker for The Maritime Environment, *Proceedings of SPIE, Signal Processing, Sensor Fusion, and Target Recognition XX*, Vol.8050, 2011.
- [21] F. Fusier, V. Valentin, F. Brémond, M. Thonnat, M. Borg, D. Thride and J. Ferryman, Video Understanding for Complex Activity Recognition, *Machine Vision and Applications*, Vol.18, 2007, pp. 167-188.
- [22] Z. Kalal, J. Matas and K. Mikolajczyk, P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints, *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 49-56.
- [23] S. Todorovic, Statistical Modeling and Segmentation of Sky-Ground Images, *Master thesis, Florida University*, 2002.
- [24] S. M. Ettinger, N. C. Nechyba, P. G. Ifju, and M. Waszac, Vision-guided Flight Stability and Control for Micro Air Vehicles, *Advanced Robotics*, Vol.17, No.7, 2003, pp. 617-640.
- [25] S. Fefilatyevev, V. Smarodzinava, L. O. Hall and D. B. Goldgof, Horizon Detection Using Machine Learning Techniques, *5th International Conference on Machine Learning and Applications*, 2006, pp. 17-21.
- [26] T. Cornall and G. Egan, Calculate Attitude from Horizon Vision, *11th Australian Aerospace Congress*, 2005.
- [27] T. G. Mcgee, R. SenGrupta and K. Hedrick, Obstacle Detection for Small Autonomous Aircraft Using Sky Segmentation, *Proceedings of the 2005*

*IEEE International Conference on Robotics and Automation*, 2005, pp. 4679-4684.

[28] D. Dusha, W. Poles and R. Walker, Fixed-Wing Attitude Estimation Using Computer Vision Based Horizon Detection, *Proceedings of Australian International Aerospace Congress*, 2007, pp. 1-19.

[29] J. Sanderson, M. Teal and T. Ellis, Characterization of a Complex Maritime Scene Using Fourier Space Analysis to Identify Small Craft, *7th International Conference on Image Processing and its Applications* Vol.2, 1999, pp. 803-807.

[30] A. A. Smith and M. Teal, Identification and Tracking of Maritime Objects in Near Infrared-Image Sequences for Collision Avoidance, *7th International Conference on Image Processing and its Applications*, Vol.1, 1999, pp. 250-254.

[31] M. D. R. Sullivan and M. Shah, Visual Surveillance in Maritime Port Facilities, *Proceedings of SPIE* Vol.6978, No.11, 2008, pp. 1-8.

[32] P. O. Westall, P. Shea, J. J. Ford and S. Hrabar, Improved Maritime Target Tracker Using Colour Fusion, *International Conference on High Performance Computing & Simulation*, 2009.

[33] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME – Journal of Basic Engineering*, Vol.82, 1960, pp. 35-45.

[34] J. Li, Y. Wang and Y. Wang, Visual Tracking and Learning Using Speeded Up Robust Features, *Pattern Recognition Letters*, Vol. 33, 2012, pp. 2094-2101.

[35] J. Barron, P. Fleet and S. Beauchemin, Performance of Optical Flow Techniques, *International Journal of Computer Vision*, Vol.12, No.1, 1994, pp. 42-77.

[36] K. Fukunaga, and L. Hostetler, The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition, *IEEE Transactions on Information Theory*, Vol.21, No.1, 1975, pp. 32-40.

[37] Y. Cheng, Mean Shift, Mode Seeking and Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.17, No.8, 1995, pp. 790-799.

[38] D. Comaniciu and P. Meer, Mean Shift: A Robust Approach Toward Feature Space Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002, pp. 603-619.

[39] G. R. Bradski, Computer Vision Face Tracking for Use in a Perceptual User Interface, *Intel Technology Journal Q2*, 1998.

[40] Comaniciu, D., Romesh, V. and Meer, P., Kernel-based Object Tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, No.5, 2003, pp. 564-577.

[41] C. Bibby and I. D. Reid, Visual Tracking at Sea, *In: ICPA*, 2005, pp. 1841-1846.

[42] R. T. Collins and Y. Liu, On-line Selection of Discriminative Tracking Features, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.10, 2005, pp. 1631-1643.

[43] H. Schweitzer, R. Deng and R. F. Anderson, A Dual-Bound Algorithm for Very Fast and Exact Template Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.33, No.3, 2011.

[44] E. Rublee, V. Rabaud, K. Konolige and G. Bradski, ORB:an Efficient Alternative to SIFT or SURF, *IEEE International Conference on Computer Vision*, 2010.

[45] D. Sinha, and G. Sanyal, Development of Human Tracking System for Video Surveillance, *Computer Science & Information Technology*, Vol.3, 2011, pp. 187-195.

[46] A. I. Kravchonok, Region Growing Detection of Moving Objects in Video Sequences Based on Optical Flow, *Pattern Recognition and Image Analysis*, Vol.22, No.1, 2012, pp. 224-255.

[47] S. X. Li, H.-C. Chang and C. F. Zhu, Adaptive Pyramid Mean Shift for Global Real-Time Visual Tracking, *Image and Vision Computing*, Vol.28, 2010, pp. 424-437.

[48] R. D. S. Moreira and N. F. F. Ebecken, Parallel WiSARD Object Tracker: a RAM-Based Tracking System, *Computer Science & Engineering: An International Journal*, Vol.4, No.1, 2014.

[49] J. Ning, L. Zhang, D. Zhang and W. Yu, Joint Registration and Active Contour Segmentation for Object Tracking, *IEEE Transactions on Circuits Systems for Video Technology*, Vol.23, No.9, 2013.

[50] J. Puzicha, Y. Rubner, C. Tomasi, and J. Buhmann, Empirical Evaluation on Dissimilarity Measures for Color and Texture, *Proceedings of the Seventh International Conference on Computer Vision*, 1999, pp. 1165-1173.

[51] T. Ojala, M. Pietikainen and T. Maenpaa, Multiresolution Grey-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, 2002, pp. 971-987.

[52] R. Goldenberg, R. Kimmel, E. Rivlin and M. Rudzsky, Fast Geodesic Active Contours, *IEEE Transactions on Image Processing*, Vol.10, No.10, 2001.

[53] J. Shi and C. Tomasi, Good Features to Track, *9th IEEE Conference on Computer Vision and Pattern Recognition*, 1994.