

INTEGRATED VISUAL INFORMATION FOR MARITIME SURVEILLANCE

Domenico D. Bloisi^a, Luca Iocchi^a, Daniele Nardi^a and Michele Fiorini^b

^aDepartment of Computer, Control, and Management Engineering, Sapienza University of Rome, via Ariosto 25, 00185 Rome, Italy.

^bSELEX-ES S.p.A. - A Finmeccanica Company, Rome, Italy.

Abstract

Surveillance systems for the maritime domain are becoming more and more important. However, available technology still provides only a limited support to this kind of applications. In this book chapter, we describe a framework for intelligent surveillance in the maritime domain, designed for fusing information from heterogeneous sources. Electro-optical cameras are used as main sensors, in order to enhance the functionalities of current Vessel Traffic Services (VTS) systems. Furthermore, the framework can be used for the detection of non-cooperative targets and can be deployed in populated areas, where radar-based systems cannot be used due to electromagnetic radiation emissions. A quantitative evaluation of the proposed approach has been carried out on a large dataset of images and videos, collected from different real sites. Such data are publicly available from the MarDCT - Maritime Detection, Classification and Tracking dataset.

Keywords: coastal surveillance, boat detection, maritime dataset, vessel traffic services, multi-sensor data fusion

1. Introduction

The protection of coastal areas is becoming a necessary requirement due to the increasing threats such as illegal smuggling, immigration, illegal fishing, oil spills, and, in some part of the world, piracy. Moreover, the control of vessel traffic is often correlated to environment protection issues, since vessels carrying dangerous goods (e.g., oil-tankers) can cause huge environmental disasters.

There exist various surveillance systems for the maritime domain, relying on different types of sensors, including Vessel Monitoring Systems (VMS), Automatic Identification System (AIS) [1], ship- and land-based radars, air- and space-born SAR systems, harbour-based visual surveillance, and Vessel Traffic Services (VTS) systems [2].

The above mentioned technologies integrate several information sources: Examples are systems combining AIS data with SAR-imagery [3], buoy-mounted sensors with land radars [4, 5], visual- with radar-based surveillance [6], and multiple ship-based sensors [7].

In particular, VTS systems combine radar and AIS data and are often equipped with long-range surveillance cameras. However, using radar and AIS data only is not sufficient to ensure a complete solution for the maritime surveillance problem, due to the following limitations:

- The AIS signal may be not available (AIS device not activated or malfunctioning);
- The recognition task for non-cooperative (non-AIS) targets can only be addressed visually by human operators;
- Radar-based systems are not suitable for vessel tracking monitoring in populated areas, due to high electro-magnetic radiation emissions.

Replacing the radar with cameras is a feasible solution for addressing the maritime surveillance problem, without the need to place a radar antenna in a populated area. An example is ARGOS system [8], designed to monitor the naval traffic in the center of the city of Venice, Italy by using cameras only.

The aim of this book chapter is to illustrate a framework for extending the functionalities of the currently used VTS systems, by adding a visual dimension to the usual radar-like view together with the possibility of automatically recognizing non-cooperative targets.

The framework is designed to:

1. Detect boats by means of a classifier-based approach, using EO images;
2. Find the horizon line in the observed scene;
3. Track multiple boats even in presence of occlusions;
4. Fuse information from existing VTS systems with visual data;
5. Provide feedback information to the VTS system in order to adjust the radar sensitivity;
6. Be deployable in populated areas.

The framework has been tested with real data coming from different currently working VTS systems, as well as with publicly available data coming from the VOC dataset [9]. Moreover, an operative scenario where traditional VTS systems can benefit from the proposed approach is presented.

The main contribution of this book chapter is to provide a data fusion scheme for combining in a unique view radar and visual data. The experimental evaluation of the performance of the modules included in the framework has been carried out by using data from the publicly available dataset of videos and images called MarDCT - Maritime Detection, Classification and Tracking [10], containing ground truth information.

The remainder of the book chapter is organized as follows. After analysing related work in Section 2, the proposed framework is detailed in Section 3, presenting also a possible operative scenario. A quantitative evaluation of the vessel detection module is provided in Section 4, while conclusions are given in Section 5.

2. Related Work

The following aspects made maritime automatic surveillance a very challenging task (see Fig. 1):

- Wide areas to be monitored;
- Highly dynamic background (due to waves and boat wakes);
- Weather issues (such as heavy rain or fog);
- Night-time monitoring.

In addition, the objects to be monitored can have very different size (ranging from few to hundreds of meters in length) and their number can be very high: 23 million tonnes of cargo and 55,000 passengers travel by ship every day worldwide [13]. Thus, in order to perform an accurate and effective monitoring of coastal areas, it is necessary to manage a huge amount of data coming from multiple (e.g., multiple cameras) and heterogeneous information sources (e.g., daylight and night cameras, radar and AIS).

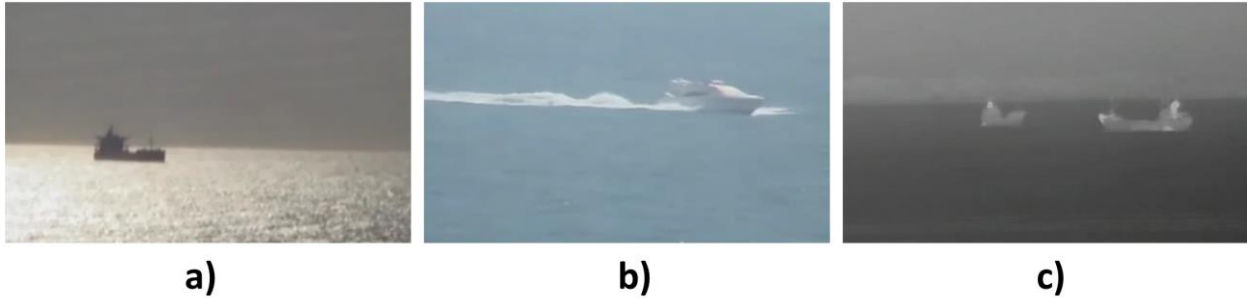


Figure 1: Challenges in maritime automatic surveillance. a) Reflections on the water surface. b) Boat wakes. c) Night-time monitoring.

Vision is becoming one essential component in recent maritime surveillance systems and this trend is supported by the deployment of a number of working systems [5]. Several approaches concerning information fusion solutions for the maritime domain are discussed in a recent survey [14] and are analysed in the following.

2.1. Systems Using Radars

Raytheon Marine Small Target Tracker [15] is a system deployed for surveillance of the Straits of Gibraltar and waterways near New York airports. The system fuses information from multiple radars, detecting and tracking small boats at a distance of 10 nautical miles. The use of microwave radars limits the surveillance within the horizon.

Accipiter Radar developed a radar surveillance network solution [16], which is currently deployed for surveillance of portions of Lake Erie and Lake Ontario between USA and Canada. The network is built using off-the-shelf radars, that can be placed on rooftops, water towers, mobile vehicles, aerostats, and towers. Since only radar sensors are considered, the output of the system can provide only a 2D view of the observed situation.

2.2. Satellite-based Systems

ASV [17] is an automatic optical system for maritime safety by using infrared (IR), GPS, and AIS data. To detect relevant objects, the sea area captured by the sensor is segmented and its statistical distribution is calculated. Any irregularities from this distribution are supposed to correspond to objects of interest. However, such an approach can produce false positives if there are wakes on the water.

COMMANDER [18] is a system developed by Thales Canada in which the Command, Control, and Communications (C3) nodes are connected through a satellite communication network to provide Canada-wide coverage. It is a system for the sharing of a Common Operating Picture (COP), including contact data, messages, and geo-referenced map overlays, between vessels, helicopters, and shore installations. However, the management of non-cooperative targets remains problematic.

2.3. Systems Using Sonars

Harbor Surveillance System [19] uses multiple sensors, including radar, sonar, and EO-IR devices to detect divers, swimmers and small boats. The system is developed by DSIT Solutions and uses an Autonomous Underwater Vehicle (AUV) that performs underwater surveys using forward-looking and side scan sonar systems. The declared detection ranges for the underwater surveillance unit is 500 meters.

Data coming from radar, EO-IR, and sonar are used in the system called HarborGuard2 [20] to provide over and underwater surveillance. The system, developed by L-3 Klein, is currently deployed by the US Navy for protection of base facilities; local governments for bridge, port/harbor and critical infrastructure security; and commercial companies for oil drilling rig and critical asset protection. However, the detection is based only on the radar sensors and the human operator is responsible to assigns a tracked target to the CCTV/Thermal Imaging camera controller.

2.4. Camera-based Systems

SeeCoast system [21] detects, classifies and tracks vessels by fusing EO and IR video data with radar and AIS data and provides decision support. It has been deployed in Coast Guard sites in Virginia, USA. The detection is carried out by estimating the motion of the background and segmenting it into components. However, motion-based vessel detection can experience difficulties when a boat is moving directly toward the camera or is anchored off the coast due to the small amount of inter-frame changes.

Maximum Average Correlation Height (MACH) filters are employed for vessel classification in [22]. Vessel detections are cross-referenced with ship pre-arrival notices in order to verify the access of vessels to the port. As reported by the authors, such an approach tends to misclassify small vessels (e.g., fishing boats).

An object detection system for finding ships in maritime video is detailed in [23]. The used approach is based on the Histogram of Oriented Gradients (HOG) [24]. Since the calculation of the detection features involves a significant amount of computational resources, real-time performance can be obtained only by means of hardware acceleration with programmable components such as FPGAs.

A method for visual surveillance in maritime domain with non-stationary camera installed on an untethered buoy is presented in [25]. After the detection of the horizon line, a color gradient filter is applied to obtain a grayscale image with intensities corresponding to the magnitude of color changes. Detection of the objects of interest is performed through thresholding of such grayscale image into a binary map. The algorithm is limited by the assumption that all marine targets are located above the horizon line.

2.5. Discussion

As argued in [14], none of the above presented systems addresses the requirements for performing in highly cluttered conditions, for tracking targets in ambiguous situations, and for reporting suspicious activity. Moreover, most of the cited systems use the radar as the main sensor and such a solution may not be applicable in populated areas. As an example, just consider the maritime traffic in the Venetian Laguna or in Venice down-town itself [8] and, more in general, in a touristic or residential environment where it is not acceptable to live near a radar antenna.

In this book chapter, we propose a camera-based framework in which cameras are used in combination with radar data in order to obtain a vessel traffic monitoring system with a high accuracy. In our approach, cameras are the main information sources, thus a system that adopts the presented framework can be effectively deployed both in a site where radar is already existing (augmenting its performance) and in a site where it is not possible or not convenient to install a radar.

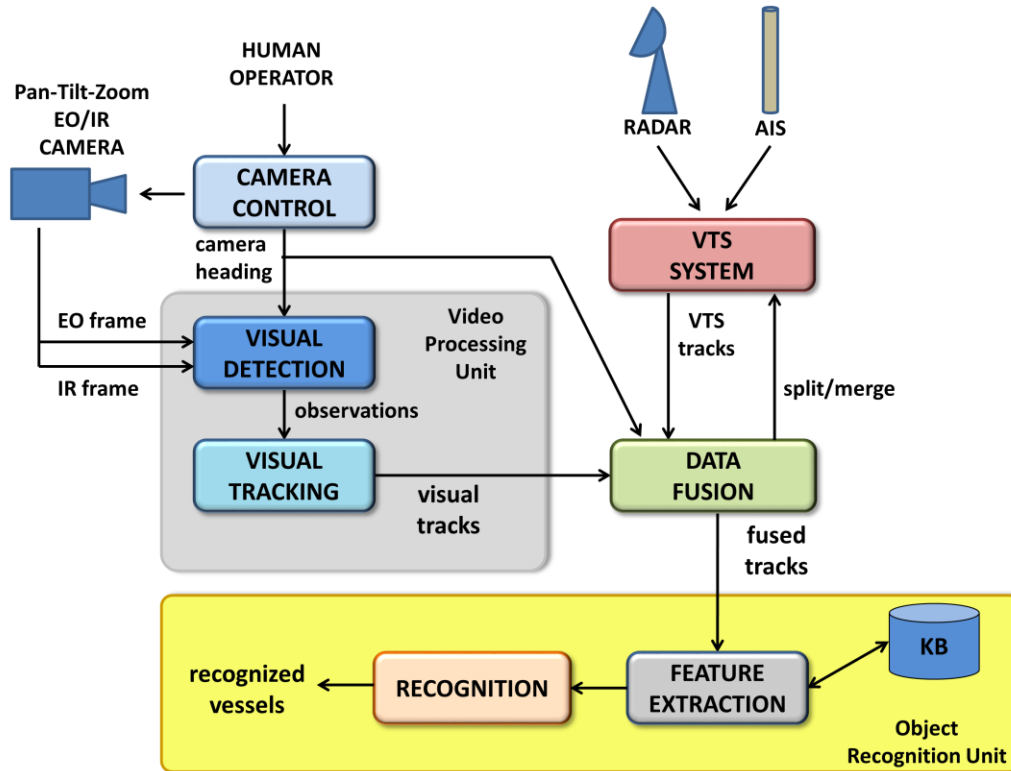


Figure 2: Functional architecture of the proposed framework. The data fusion module manages information coming from the Visual Processing Unit and the VTS system and sends validated tracks to the Object Recognition Unit.

On the other hand, the proposed approach has to deal with the challenges related to automatic video surveillance in the maritime domain [12]:

- Non-stationary cameras (e.g., Pan-Tilt-Zoom cameras);
- Different classes of targets (ranging from huge oil-tankers to small phishing boats);
- Reflections and boat wakes on the water surface;
- Boats anchored off the coast, that can appear motionless.

In the next section, the architecture of the framework is presented together with the approaches and methods that can be used to deal with the above listed issues.

3. Architecture of the Framework

The functional architecture of the proposed framework is shown in Fig. 2. An EO-IR visual device is the main sensor and it can be moved through a control module by a human operator. The control module provides the orientation and the field-of-view (FOV) of the visual sensor to the Video Processing Unit (VPU). The VPU is responsible to detect and track the vessels by using visual information. It is worth noting that the detection task is made complex by the camera movements, since it is not possible to create a model of the observed scene.

The data fusion module receives data from both the VPU and the VTS system. It aims at associating the visual tracks coming from the video analysis with the tracks generated from radar and AIS data. In this way, it is possible to provide the user a new visual dimension, in addition to the traditional geo-referenced, radar-like VTS view. Moreover, the data fusion module sends feedback information to the VTS system, that can be useful to improve the detection accuracy of the radar by adjusting its parameters.

The tracks generated by the data fusion module are sent to the object recognition unit, that classifies them according to specific visual features. In the remainder of this section, each module of the framework is described in details.

3.1. Visual Detection

The visual detection module is part of the VPU and it aims at processing the current video frame by searching for the targets of interest. Visual detection is a crucial task, since detection accuracy affects all the stages in the VPU pipeline and it must be as high as possible, while maintaining an acceptable computational load. The visual detection module is made of three main blocks: 1) the Haar-based boat classifier, 2) the horizon detection function, and 3) a filter for removing possible noise (see Fig. 3).

3.1.1. Haar-based classifier

Since the camera can be moved by the user, a foreground/background modelling approach to detect vessels (e.g., background subtraction) can result ineffective. A possible solution consists in adopting a classification-based method. Classification is used to learn a model from a set of labelled data instances (training phase) and then to classify a test instance into one of the pre-defined classes by using the learned model (testing phase) [26]. Classification-based detection can operate on still images, thus avoiding to create a model of the background.

The training phase learns the parameters of the classifier by exploiting the available training data and the testing phase classifies a test instance as being an object of interest or not, using the classifier.

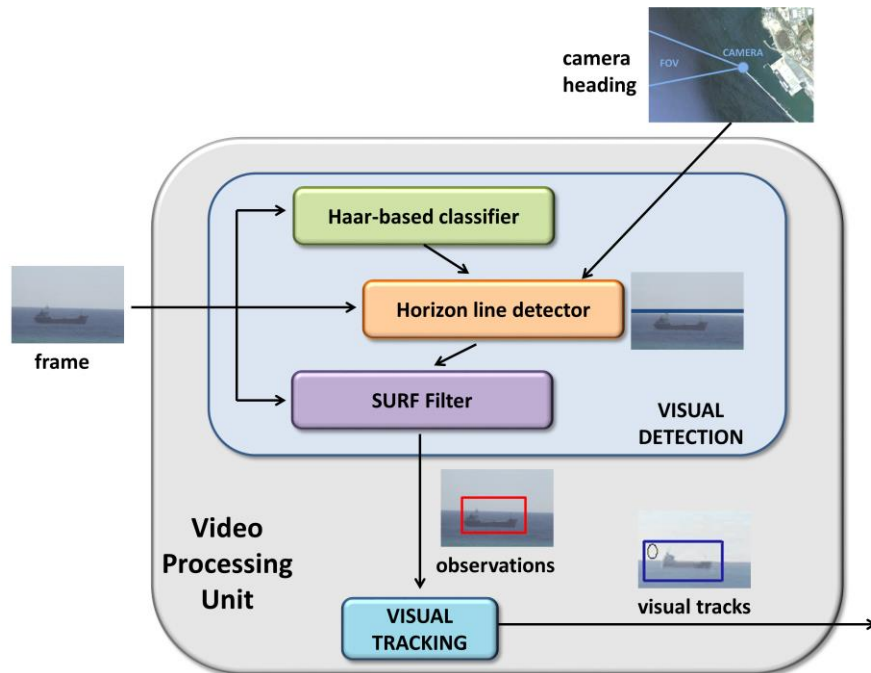


Figure 3: Video Processing Unit. Video frames coming from the camera are analysed in order to find boats and the horizon line. A SURF-based filter is used to filter out false positive detections.

Different methods can be used to create the classifier [22, 9]. In order to obtain real-time performance, we discarded computationally expensive testing methods (e.g., [24, 27]) and adopted an approach based on Haar-like features [28].

It is worth noting that such an approach has been originally designed for face detection, thus we suitably adapted it for boat detection. Moreover, the chosen approach gives the possibility of combining different weak classifiers in a cascade, in order to obtain a strong classifier.

A set of 4000 images not containing boats (negative images) and a set of 1500 images taken from the Internet, depicting various types of vessels with different view angle (positive images), have been used as input for the off-line training phase, obtaining a 24 level classifier (the training

stage was stopped when the false alarm rate reached 5×10^{-6}). The OpenCV [29] HaarTraining functions have been used to create the classifier with a 60×30 search window. The size of search window has been experimentally found. The used set of positive and negative images can be downloaded from the MarDCT - Maritime Detection, Classification and Tracking dataset [10].

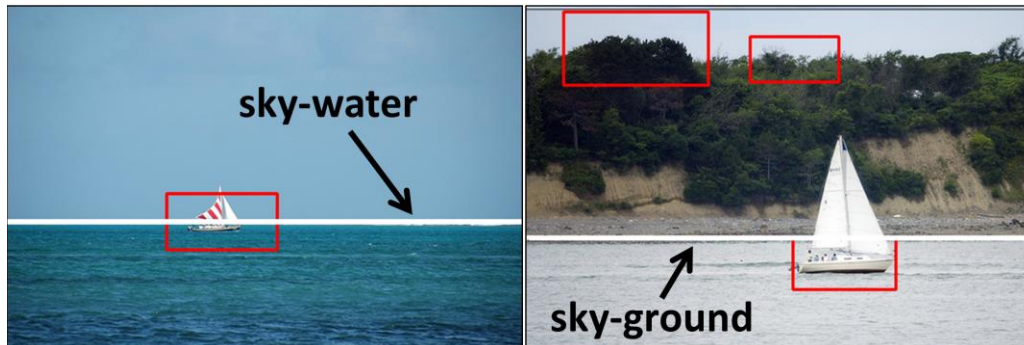


Figure 4: Sky-water and sky-ground line. The probability of finding false positives increases in presence of the coast, due to the rich texture of the scene (right side of this figure). The images are from the VOC dataset [9].

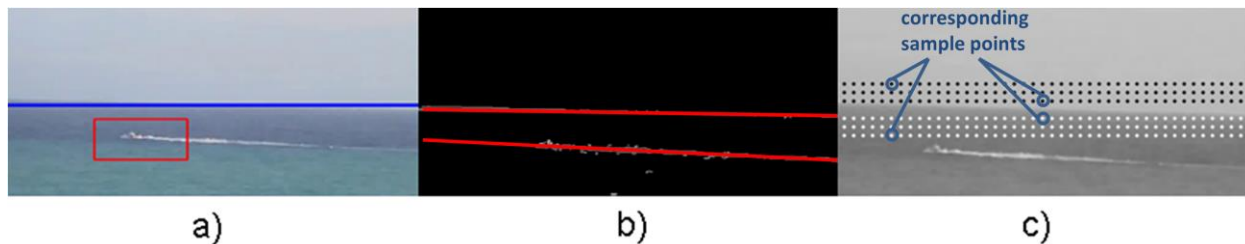


Figure 5: Horizon line detection. a) The horizon line correctly detected. b) The edge map with the Hough line in red. c) The set of sample points used for validating the line (two corresponding pairs are highlighted).

3.1.2. Horizon Line Detector

Along with boats, it can be useful to detect the limit of the sea surface also. Depending upon the heading of the camera, the module called Horizon Line Detector is responsible to find the horizon line, differentiating between sky-water and sky-ground line (Fig. 4). Indeed, in presence of the coast the probability of finding false positives increases due to the rich texture of the scene, thus it is convenient to filter out detections laying above the sky-ground line (see the right side of Fig. 4).

In order to detect the limit of the sea surface (Fig. 5a), the Hough transform is applied to the edge map of the frame, creating a list of candidate lines (red lines in Fig. 5b). Then, each candidate line is validated with respect to a set of sample points belonging to a rectangular region created around the line (Fig. 5c). The points above the candidate line are compared with the corresponding points under the line. If the percentage of corresponding pairs with different intensity values is over a predefined threshold, then the line is considered valid. We set the threshold at 95%, in order to filter out false positive detection that can be generated by long wakes. As an example, in Fig. 5b two candidate lines are individuated. The lower line is discarded because the sample points over and under the line have similar intensity values, while the upper line is considered as a valid horizon line.



Figure 6: False detections caused by wakes can be filtered out analysing the SURF key points in the bounding box. If a negligible number of key points is found, then the observation is rejected.

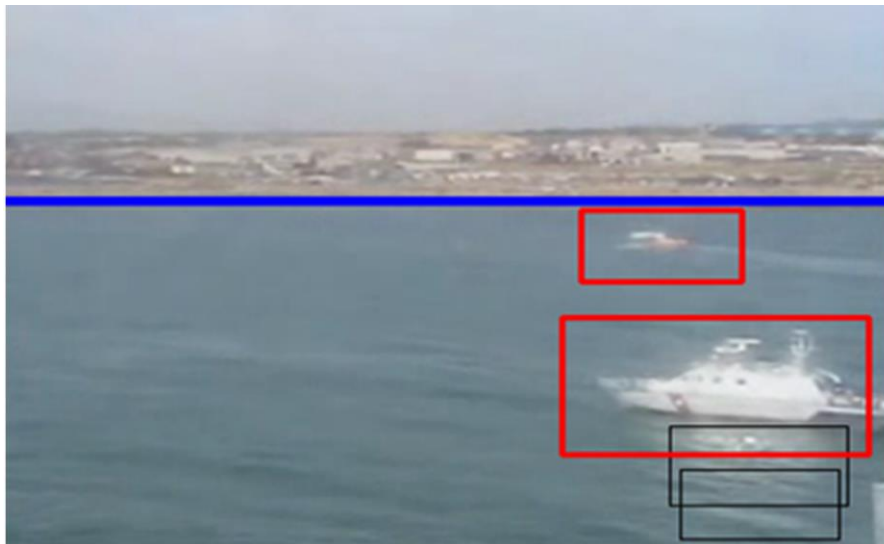


Figure 7: In case of reflections, SURF can be used for rejecting false positive detections, since we experimentally noted that the number of key points generated by reflections on the water surface is limited.

3.1.3. SURF filter

The acronym SURF (Speeded-Up Robust Features) indicates a scale- and rotation-invariant detector and descriptor [30]. SURF can be used to filter out possible false detections made by the classifier. Indeed, in the maritime domain, false detections can be caused by waves, boat wakes, and reflections on the water surface. When the classifier detects an object of interest, its bounding box is analysed in order to extract the SURF key points in it. If a negligible number of key points is present in the bounding box, then the observation is rejected.

In Fig. 6 the above presented algorithm can be used to filter out a false positive detection caused by wakes (black bounding box). In Fig. 7 the same algorithm allows to filter out two false positive detection caused by reflections on the water surface (the two black bounding box on the bottom of the image).

3.1.4. Observations

The output of the detection module is a list of observations, each of them being a bounding box representing a detected boat. Our classifier [11] is able to detect targets of different size, with blurring noise, in presence of boat wakes and reflections on the water surface (see Fig. 8).

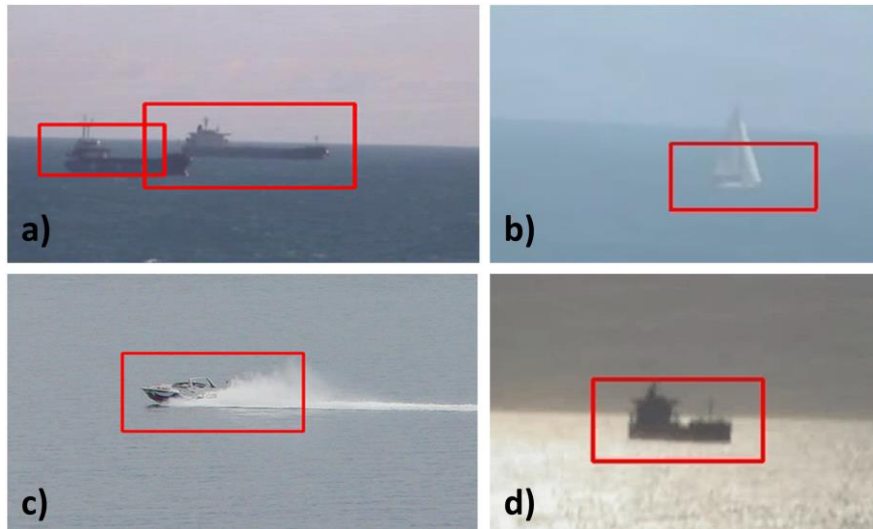


Figure 8: Examples of observations. The classifier is able to detect targets of different sizes (a), with blurring noise (b), and in presence of boat wakes (c) and reflections on the water surface (d). Picture c is from the VOC dataset [9].

3.3. Visual Tracking

The visual tracking module takes as input the image flow and the observations generated by the detection module. It returns a set of visual tracks, i.e., bounding boxes with an identification number, having as main role the one of temporal filtering the false positives.

The association between tracks and observations is made on the basis of a nearest-neighbour policy with the Bhattacharyya distance between the HSV value histograms of the track (updated over time) and those of the current observation as measure. Only tracks that present a sufficient number of associated observations are considered of interest (the association threshold has been set to 10): such a policy allows to filter out a high number of false positive detections.

3.3.1. Group Tracking

Sometimes occlusions can occur when boats are aligned with respect to the camera view or when boats are close to each other. Occlusion situations cause considerable difficulties for visual tracking. A possible solution is to consider partially overlapping tracks to form a group instead of tracking them separately (see Fig. 9). When two or more tracks have their bounding boxes moving closer to each other (Fig. 9a), the tracker starts considering them to form a group (Fig. 9b) and it stores the HSV histograms of the involved tracks - the histograms will be used as models for re-identify the tracks when the occlusion phase is over. To this end, when only one track is completely visible, the tracker considers the occluded track as frozen (Fig. 9c). When the occluded object becomes visible again, a new track is identified, but the tracker can use the saved histograms to re-assign the correct identification number to the already registered target (Fig. 9d).

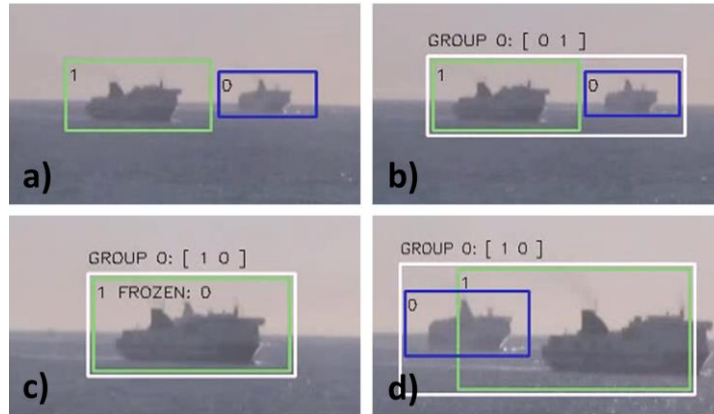


Figure 9: Group tracking.

3.4. VTS System

A Coastal Surveillance Site (CSS) is responsible for detecting and tracking vessels in a range up to 12 nautical miles (NM), the defined “Territorial Waters” [31]. A VTS system is a CSS designed to merge information coming from radars and AIS devices. The user is provided with a 2D geographical view, showing all the boats navigating near the site. For AIS equipped vessels, the VTS system can also show additional information such as the name of the boat and its route.

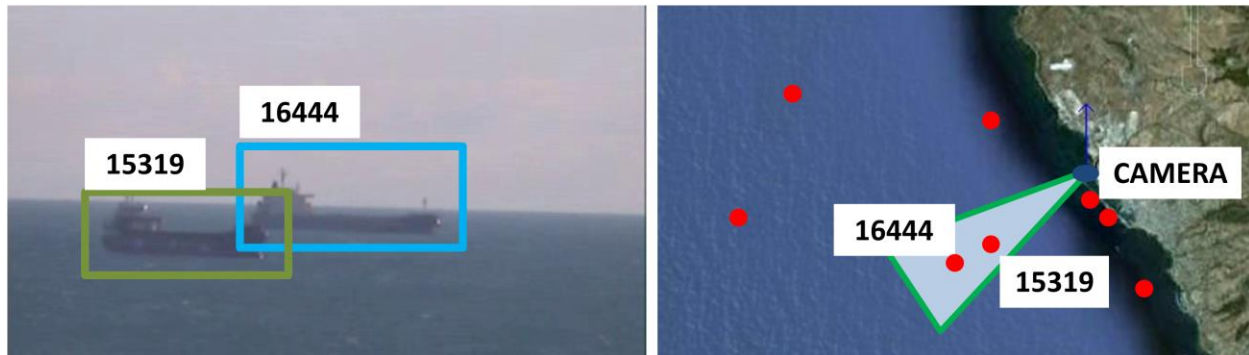


Figure 10: Combined view. The state-of-the-art VTS systems can be improved adding a visual dimension (left side of this figure) to the currently available radar-like view (right side of this figure).

3.5. Data Fusion

The data fusion module aims at creating a combined view for the monitored situation, which is not available in today's VTS systems (see Fig. 10).

The following steps are performed by the data fusion module:

1. Data Synchronization: VTS and video data are synchronized to perform a consistent data association. Time synchronization is required since VTS data have a refresh period of about 3 seconds, while the video stream is transmitted at 25 frames per second.
2. VTS Tracks Selection: VTS tracks in the FOV of the camera are selected (see right side of Fig. 10) by considering the current heading of the camera, that is received from the camera control module.
3. Rotation: VTS tracks are aligned with respect to the camera heading, since the 2D geographical view can be rotated with respect to the frames acquired by the camera.

4. Common Space Projection: Visual and VTS tracks are projected in a common, normalized 2D space (see Fig. 11). This step can be performed since all the boats are moving on a plane (the water surface).
5. Association: A probabilistic association between VTS and visual tracks is carried out in order to build the final integrated view (see Fig. 12).

Data fusion between video and VTS data is performed on a probabilistic basis. VTS and visual tracks are projected onto a two dimensional common reference space in order to perform the association. For visual data, the first dimension (x) in the common reference space is the distance (in pixels) of the bounding box from the left margin of the frame, while the second dimension (y) is the distance of the bounding box from the bottom of the frame (see the left side of Fig. 11).

For VTS data, the distance (in pixel) of the VTS track from the left side of the FOV and the distance from the camera position in the geographic view represent the x and y dimensions, respectively (see the right part of Fig. 11).

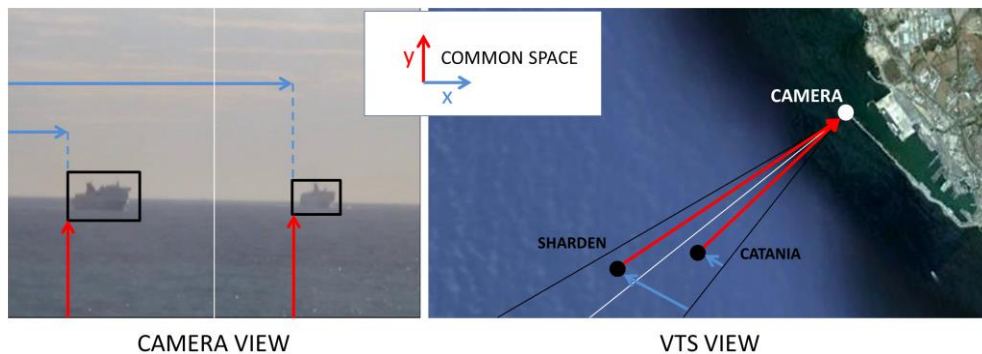


Figure 11: Visual and VTS data projection. The values for the x and y coordinates are extracted from the video frame (left part of the figure) and the geographical view (right part of this figure).

Since the video frame and the geographical view have different scales, the x and y values related to the visual track and the x and y coordinates related to the VTS track are normalized with respect to the width and the height of the common reference space window. The projected visual and VTS tracks are associated on the basis of a nearest-neighbour policy (see Fig. 12), by using a predefined threshold.

The proposed data fusion scheme is justified by the fact that the boats are moving on a planar surface, thus a boat that is closer to the camera than another one will appear in a lower position in the image [12].

The final output presented to the user is a global view of the observed situation, as reported in Fig. 13.

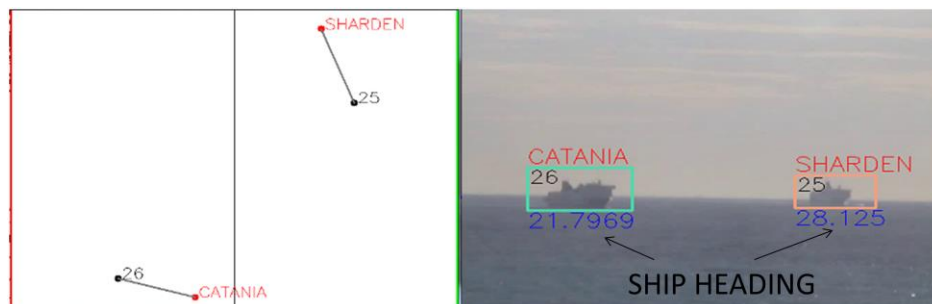


Figure 12: Probabilistic association. The common space is shown on the left, with VTS data marked in red and visual tracks in black. The observed scene is reported on the right, with the AIS labels associated with the images of the boats. The current headings of the ships are also calculated.

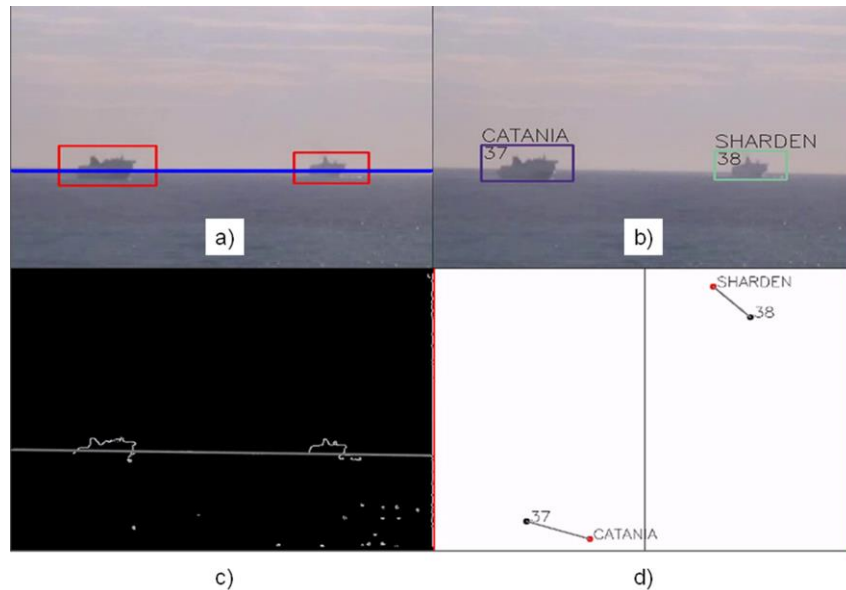


Figure 13: Global view of the observed situation. a) Visual detection. b) Visual tracking and VTS data fused in the same view. c) Edge map for sky-water line detection. d) Data association in the common space.

3.5.1. Feedback to VTS System

The data fusion module can help in correcting erroneous detection made by the radar. In presence of targets having a relevant size (e.g., oil-tankers), it is possible to have multiple VTS tracks for the same target (split case). On the other hand, in the case of a small boat navigating near a bigger one, it is also possible to have a single VTS track instead of two (merge case).

Fig. 14 shows an example of a split situation: the system detects two visual observations (Fig. 14a) and two visual tracks (Fig. 14b), but multiple VTS tracks (projected in Fig. 14c as red dots) also. The multiple radar detections can be clustered by exploiting the visual tracks (the black dots in Fig. 14c referring to the visual tracks in Fig. 14b) as centroids.

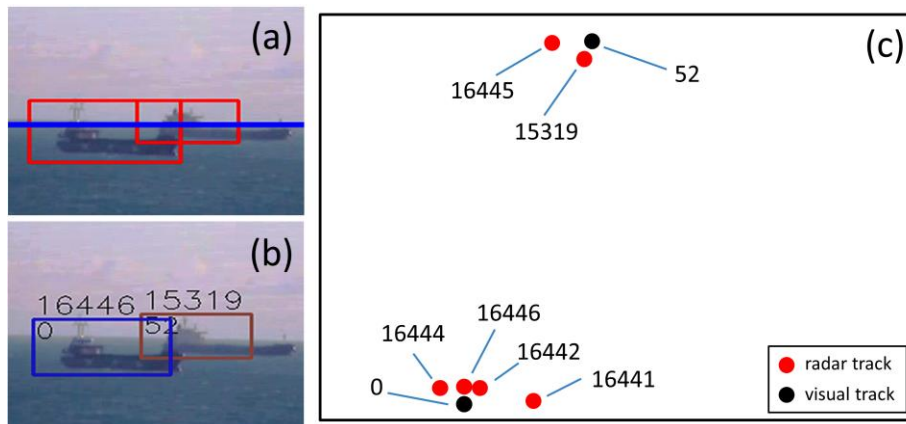


Figure 14: Feedback information can be sent to adjust the radar parameters in case of misdetection.

3.6. Object Recognition

The Object Recognition Unit is responsible for recognizing the boats in the FOV of the camera. In order to deal with such a complex problem, a novel algorithm, called “barcode matching”, is proposed. The method is designed to work in a well-defined operative scenario and to manage non-closed curves and low information images. The “barcode matching” requires a database containing a set of records of known vessels, each record containing the silhouette of a boat and a heading. Multiple silhouettes of the same boat can be stored, corresponding to different headings. All the stored silhouette images are normalized to the same size, in order to allow a comparison between them.

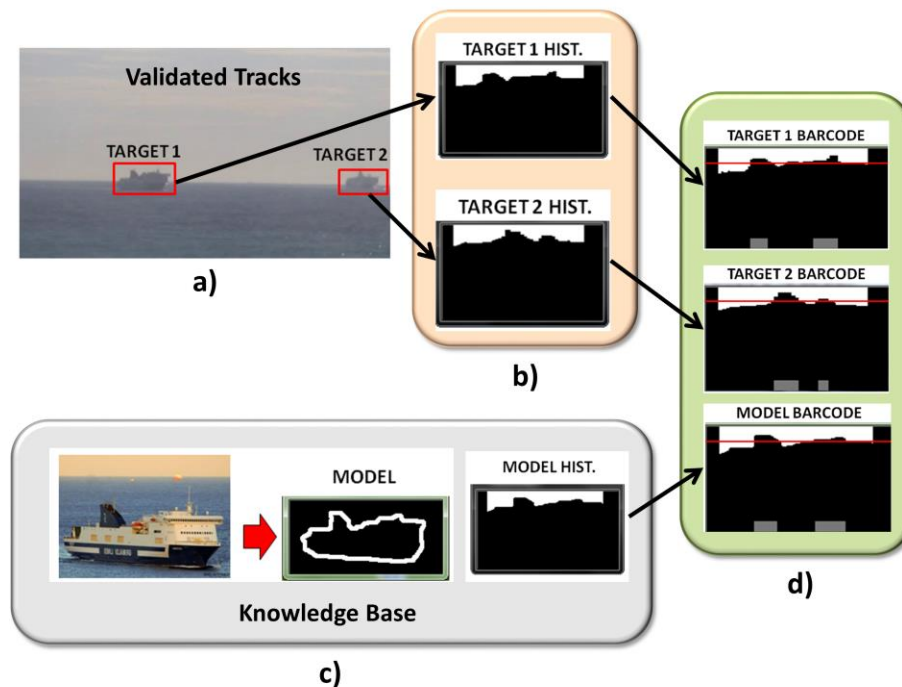


Figure 15: Target silhouettes extraction and comparison. a) Validated tracks are analysed in order to extract silhouettes. b) Targets' silhouettes are used to generate histograms. c) The knowledge base in the Object Recognition Unit is made of histograms ordered by heading values. d) Barcode matching.

The barcode matching method consists of the following steps (see Fig. 15).

1. The bounding box around the object to recognize is obtained by the list of valid tracks (Fig. 15a).
2. The object's silhouette inside the bounding box is extracted and the edges belonging to the horizon line (sky-water line), if present, are filtered out. A histogram is computed starting from the object's silhouette. The histogram represents the distance of the object's silhouette from the top of the image (Fig. 15b).
3. A model histogram is chosen in the database according to the heading of the object (Fig. 15c).
4. The chosen model histogram is compared with the object's histogram by calculating the Bhattacharyya distance between them. Then, the barcode for the object's histogram is computed and it is compared with the barcode from the model silhouette (Fig. 15d).

The more similar the two histograms are, the lower the Bhattacharyya distance value is. However, even if such a distance grows in proportion to the amount of differences in the two histograms, it is not possible to completely rely on it, in order to recognize an object. Indeed, it is very difficult to determine a threshold that can be used to recognize an object given a model's histogram. Thus, in order to improve the discriminative power of the method, an additional set of features has been used. By computing the mean value μ of the histogram, it is possible to identify the bins in the histogram, whose values are below μ . As an example, the silhouette produced by a chimney or by an antenna of a boat can be

identified by analyzing the number and the position of the bins that are above the mean μ . Indeed, the barcode for each histogram highlights particular features of the silhouette of the vessel, that can be used to distinguish between two possible matches (see Fig. 15).

It is worth noting that, even if the described recognition algorithm is strongly affected by the angle with which the camera captures the boat, in real scenarios the heading of the boats near the coast is rather stable and predictable, because there is always an approaching corridor to enter or to leave a port that the boats have to navigate in. Fig. 16 shows an example where two boats are navigating in opposite directions, inside a corridor that is established by nautical charts. The example demonstrates that, in operative scenarios, the angle between the boats and the camera is predictable, since all the boats are forced to follow almost the same path and the camera can be deployed accordingly, in order to obtain the best possible view.

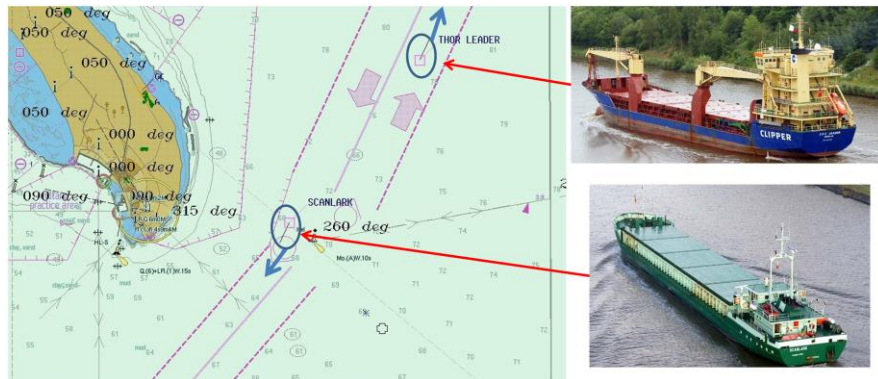


Figure 16: Operative scenario. Two boats are navigating in a corridor to enter/leave a port, maintaining a predictable heading. Images are from the website www.marinetraffic.com.

4. Experimental Results

The boat detection approach has been tested in order to evaluate both the detection accuracy and the required computational load. Two different data sets have been used:

- 1) MarDCT, containing video sequences coming from a real VTS site;
- 2) VOC, containing images coming from the publicly available VOC database [9].

4.1. MarDCT Data

In order to quantitatively evaluate the accuracy of the detection method, a data set of maritime video streams and images, called MarDCT - Maritime Detection Classification and Tracking, containing ground truth data that can be used for evaluating the performance of automatic video surveillance systems and of computer vision techniques [10], has been built. In particular, for each video in the data set the following details are provided:

- Sensor type;
- Camera type (static or moving);
- Location;
- Light conditions;
- Foreground masks to evaluate foreground segmentation;
- Annotations with the bounding box vertices to evaluate detection;
- Identification numbers to evaluate data association and tracking over time.

In addition, the videos have been recorded with varying observing angles and whether conditions.

MarDCT data set contains:

1. Visual data recorded in a VTS centre in Italy;
2. Visual data recorded in a VTS centre in the North of Europe;
3. Visual data from the ARGOS system [8] that monitors the naval traffic in the Grand Canal in the city of Venice, Italy;
4. Visual data recorded in a port in Asia.

A set of 100 randomly chosen images taken from 11 videos recorded with different light conditions and camera positions has been used to test the accuracy of the boat detection. To filter out false positives due to wakes (e.g., in the middle of Fig. 6a) and reflections (e.g., in Fig. 8), an additional weak-classifier has been created by means of a negative set made of 4000 images of wakes and other false positive detections obtained by the original classifier. The results are reported in Table 1, in terms of detection rate (DR) and false alarm rate (FAR)

$$DR = \frac{TP}{TP + FN} \quad FAR = \frac{FP}{TP + FP}$$

where TP are the correctly detected boats, FN is the number of not detected boats, and FP are the incorrect detections.

Table 1: Visual detection results on data coming from a real site.

Classifier	Coastline Detection	Detection Rate	False Alarm Rate
Without wake examples	NO	0.892	0.475
Without wake examples	YES	0.892	0.265
With wake examples	YES	0.928	0.251

4.2. VOC Data

The VOC database [9] is a publicly available dataset containing several images of interest for maritime surveillance applications. We considered important to include in the evaluation such database, because it contains a large variety of boats and it is thus very adequate to test in particular the detection phase of the system. Since the VOC database contains images belonging to different classes of objects, we selected a subset of the images containing boats, discarding images that are not suitable for a VTS application.

The results of the detection using our Haar-based classifier are summarized in Table 2.

Table 2: Visual detection results on data from VOC [9].

Coastline Detection	Detection Rate	False Alarm Rate
NO	0.872	0.332
YES	0.872	0.198

Results show good performance of our system and, in particular, the improvement in using the horizon detection in reducing false alarms (from 0.332 to 0.198).

4.3. Computational Speed

The detection method has been tested with two different CPUs:

1. Intel Core 2 U7300 1.30 GHz 4 GB RAM (2 cores);
2. Intel Core i7 3770 3.40 GHz 16 GB RAM (8 cores).

The computational speed has been measured using recorded data, live data coming from a commercial RTSP webcam, and live data coming from a EO camera on site. The tests demonstrated the capability of the system architecture of exploiting all the computational power available. The results are shown in Table 3.

Table 3: Computational load for the detection phase.

Frame Size	2 cores	8 cores
319 x 261	10	65
352 x 288	9	60
414 x 338	6	43
586 x 479	4	25
704 x 576	3	20

4.4. Discussion

The analysis of the results demonstrates that:

1. The Haar-based approach is an effective solution for boat detection with moving cameras in the maritime domain. An elevated DR can be obtained on real data and on benchmark images. It is worth noting that our detector has been built using images coming from real data recorded in a real site. Those data differ significantly from the VOC dataset data, thus a detection rate of 0.872 can be considered acceptable for validating its performance.
2. The horizon line detection is crucial in obtaining a lower FAR. Indeed, an approach based on Haar-like features for building the classifier inherently produces a high DR, with an elevated FAR. Thus the chance to lower it through the horizon line analysis allows to considerably improve the detection performance.
3. The approach is suitable for a real-time application. With a commercial CPU it is possible to achieve a real time (25 fps) processing speed for 586 x 479 images.

Moreover, the observations obtained using the boat detector are sent to the visual tracking module and then to the data fusion process. In such a way, it is possible to drastically reduce the FAR, as well as improve the DR, thanks to the temporal filtering performed by the tracker and to the context-based information coming from the VTS system. However, it is essential that frame-by-frame detection provides reliable results, since the presented framework is also conceived to be used in place of radar based systems at the occurrence.

5. Enhancing Air Traffic Control With Visual Data

The use of the described framework is not limited to the maritime scenario. Indeed, the same framework can be successfully applied also in the case of an air traffic monitoring system. Advanced surface movement guidance and control systems (A-SMGCS) aim at providing routing, guidance and surveillance for the control of aircraft and vehicles in order to maintain the declared surface movement rate under all weather conditions within the aerodrome visibility operational level (AVOL), while maintaining the required level of safety [???]. The necessary condition for all the algorithms (routing, guidance, and conflict avoidance) to work correctly is the provision of reliable surveillance data in all airport areas.

Usually, Surface Movement Radars (SMR) are used as the main source of information for estimating aircraft positions. However, in complex airport layouts (see Fig. 17), traditional radar-based surveillance systems are affected by limitations in their coverage due to reflections or shadows that are caused by buildings, equipment or other reflecting objects on the airport surface.

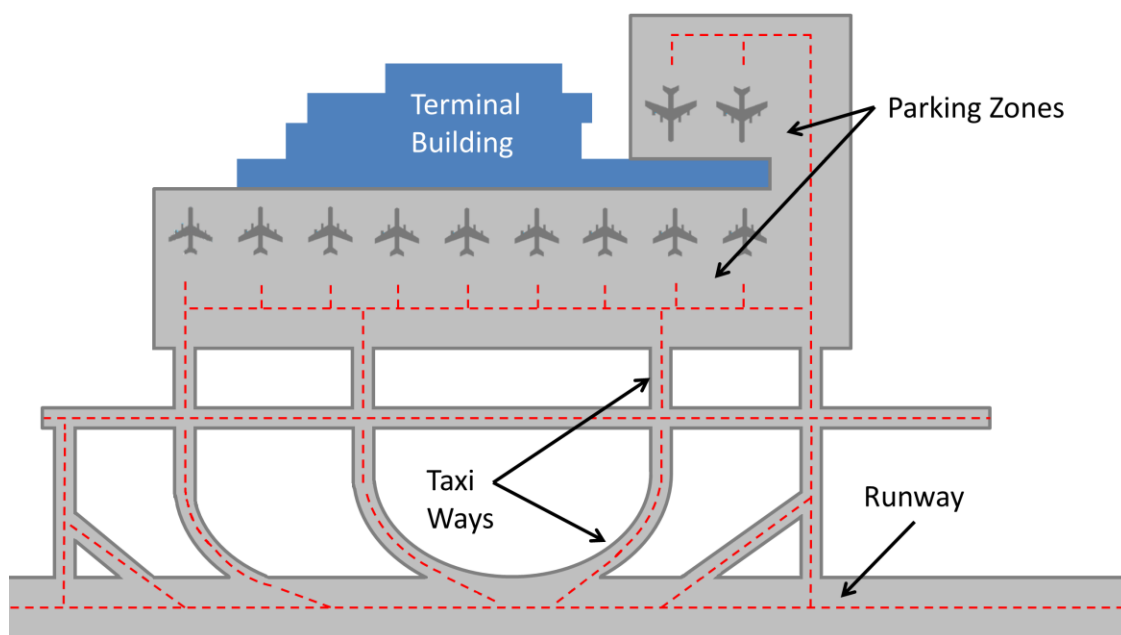


Fig. 17. A general airport layout. Shadows and blind spots can occur in the parking zones due to the presence of buildings and reflecting objects on the airport surface.

Cooperative systems for airport ground traffic surveillance, such as the Automatic Dependent Surveillance-Broadcast (ADS-B) system [???], can be involved for mitigating the problems. An ADS-B equipped aircraft determines its own position using a global navigation satellite system (GNSS) and periodically broadcasts its position as well as other relevant information to potential ground stations and to other aircrafts with ADS-B equipment. Nevertheless, the adoption of the ADS-B system does not provide reliable data, since the above mentioned limitations can also arise in the GPS signal propagation used by the aircrafts (or vehicles) to compute their position. Furthermore, not all the aircrafts are equipped with ADS-B transponders.

When the GPS signal is not available, it is possible to adopt a Multilateration (MLAT) approach. MLAT uses a grid of ground receivers to record ADSB reports and to perform a time-difference-of-arrival analysis to estimate the aircraft position. However, MLAT can only detect cooperative targets since the aircraft transponder has to be switched on. In case of malfunction or if the transponder is switched off, the system cannot detect the aircraft [???].

For all the above discussed reasons, it is very difficult to provide a reliable coverage in all the airport areas without increasing the number of surveillance sensors (SMR radars or MLAT sensors). Moreover, electromagnetic pollution caused by radars is something that must everywhere be reduced to a minimum, therefore a “green” technology should be used to get the exact position of all parked aircrafts reliably in order to provide the airport controllers a complete view of the situation at hand.

A possible solution for increasing the overall system coverage as well as for reaching higher levels of the A-SMGCS specification [??] is to introduce a system able to act as a “gap-filler” for the airport zones that traditional sensors fail to reliably cover. Such a system should be

1. less expensive than the electro-magnetic sensors used in the state-of-the-art systems;
2. radiation-free.

The framework discussed in this book chapter, where the main sensors are electro-optical cameras, can be used to identify the presence of aircrafts in the parking zones (see Fig. 17). The same assumption of using an uncalibrated pan-tilt-zoom (PTZ) camera (that can be freely moved by a human user) made in the maritime scenario is valid also when the goal is to detect and track aircrafts in parking areas. The information concerning the movements in the monitored area, e.g., the departure of an aircraft, can be merged with the data coming from the traditional radar-based system, thus increasing the overall system coverage.

5.1. A Framework for Ground Traffic Surveillance in Airports

Parking zones in airports (Fig. 17) represent a challenging scenario for automatic video surveillance due to [??]

- quite fast changes of illumination (e.g., due to cloud movement);
- reduced visibility caused by adverse meteorological conditions (e.g., rain, fog);
- presence of not moving objects of interest: relevant objects should be detected both while moving and while they are steady (e.g., parked aircrafts).

Moreover, in order to limit the number of sensors, usually parking zones are monitored by PTZ cameras. Since PTZ cameras can be moved (automatically or manually), the conventional foreground/background modeling approach is ineffective being based on the assumptions that all the objects of interest are moving and that the camera is static.

To overcome the above mentioned limitations, as in the above considered maritime scenario, the aircraft detection module has to be based on a classifier (see Fig. 18).

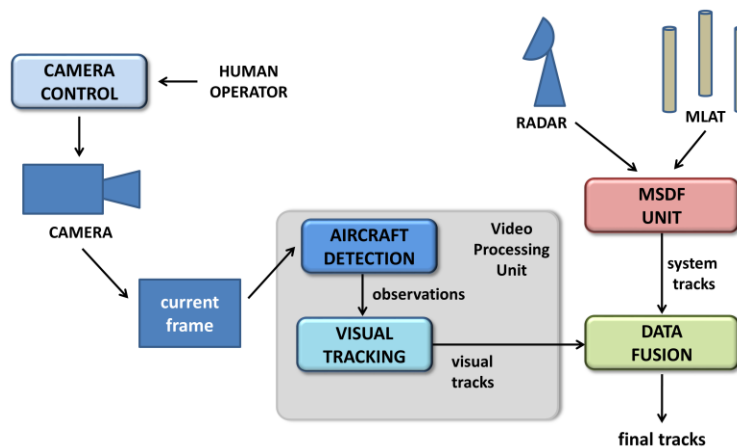


Fig. 18. The proposed framework. Visual information can be used in combination with existing radar systems to increase the overall detection accuracy.

The Video Processing Unit (VPU) is designed to detect and track the aircrafts. It takes as input a single image coming from the video stream (current frame). The Data Fusion (DF) module aims at associating the tracks coming from the VPU (visual tracks) and the tracks elaborated by the Multi Sensor Data Fusion (MSDF) unit (system tracks). Indeed, the MSDF module collects and computes all the data coming from the existing electro-magnetic sensors (ADS-B, SMR, MLAT).

5.2. Aircraft Detection

The same Haar-like feature-based approach used to detect boats can be applied to detect aircrafts. To verify the applicability of the classification-based method for aircraft detection, we created a binary aircraft classifier by using the OpenCV HaarTraining functions.

A set of 1000 images not containing aircrafts and a set of 550 images including different types of aircrafts (with different views) taken from the Internet have been used as input for the off-line training stage obtaining a 15 levels classifier (the training was stopped when the false alarm rate reached $5E10^{-6}$). We experimentally found that good detection results can be obtained using a search window of 90×30 pixels. It is worth nothing that since the aircrafts can be viewed by different angles, the dimension of the search window is a crucial parameter for creating an accurate classifier.

The output of the aircraft detection module is a list of observations. Each observation is a bounding box representing a detected aircraft. In Fig. 19 some examples of detection of different types of aircrafts are shown.

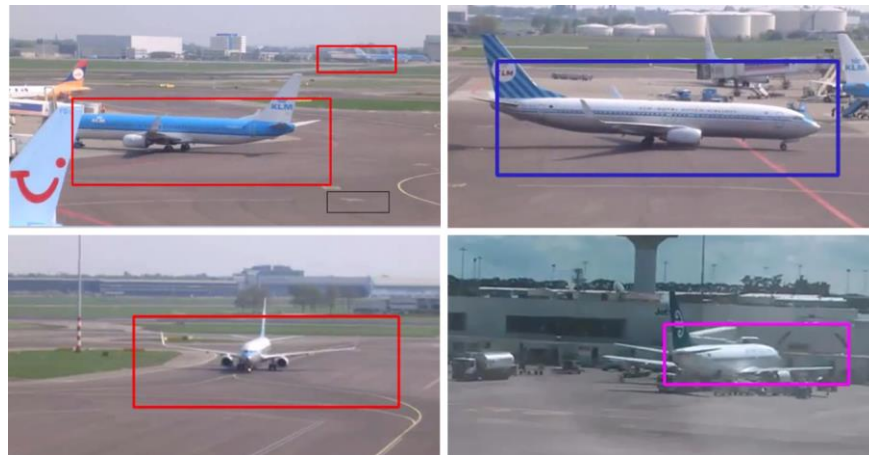


Fig. 19. Detection examples. The classifier is able to detect aircrafts given a video frame. The output is a list of bounding boxes (observations) containing the objects of interest. Images are from Youtube videos.

The results demonstrate that the proposed framework is a flexible and cost-effective solution for improving existing systems performance.

6. Conclusions

A framework for integrating visual information in an existing VTS system has been described. A major advantage of the proposed framework is that a global view (VTS information plus visual data) of the observed situation can be built, by adding a visual dimension to the traditional view combining radar and AIS data. Such an enhanced view is very effective for the user, that has an immediate visual evidence of the situation at hand. The framework is also designed to provide a feedback to the VTS system, in order to improve the radar detection by adjusting the parameters of the sensors to avoid misdetections.

The results on a real operational scenario demonstrate the robustness of the approach with respect to the presence of changing light conditions, reflections on the water surface, boat wakes, and targets that appear motionless (e.g., boats anchored off the coast). Moreover, the approach is robust to user controlled motion of the camera.

The results on two different data sets (including the publicly available data coming from the VOC data set), show that the proposed detection approach provides accurate results, while maintaining a real-time computational speed.

Furthermore, in this book chapter a publicly available data set of videos and images called MarDCT - Maritime Detection, Classification and Tracking, containing data recorded in different scenarios, using both EO and IR cameras, with varying light and weather conditions, has been presented. MarDCT contains also the data used to validate the proposed framework and it will be improved in the near future, by extending the data sets and acquisition scenarios and by adding vessel silhouettes for object recognition.

Overall, we believe that our framework substantially improves the effectiveness of surveillance in maritime domains. The underlying enabling technology is provided by exploiting a proper fusion of the information coming from different sensors. The key feature of the proposed approach is the use of cameras as the main sensors, differently from traditional VTS systems that use radars and AIS data.

References

- [1] I. Harre, AIS Adding New Quality to VTS Systems, *The Journal of Navigation* 3 (2000) 527-539.
- [2] Canadian Coast Guard, Vessel Trac Services (VTS) update study, vol. 159, 1991.
- [3] G. Saur, S. Estable, K. Zielinski, S. Knabe, M. Teutsch, M. Gabel, Detection and classification of man-made offshore objects in terrasax and rapideye imagery: Selected results of the demarine-deko project, in: *Proceedings of IEEE Oceans, Santander*.
- [4] W. Kruger, Z. Orlov, Robust layer-based boat detection and multi-target-tracking in maritime environments, in: *Waterside Security Conference (WSS), 2010 International*, pp. 1-7.
- [5] S. Fefilatyev, D. Goldgof, M. Shreve, C. Lembke, Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system, *Ocean Engineering* 54 (2012) 1-12.
- [6] M. D. Rodriguez Sullivan, M. Shah, Visual surveillance in maritime port facilities 6978 (2008) 11-19.
- [7] H. Wei, H. Nguyen, P. Ramu, C. Raju, X. Liu, J. Yadegar, Automated intelligent video surveillance system for ships 73061 (2009) 1-12.
- [8] D. Bloisi, L. Iocchi, ARGOS - a video surveillance system for boat traffic monitoring in Venice, *International Journal of Pattern Recognition and Artificial Intelligence* 23 (2009) 1477-1502.
- [9] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, *International Journal of Computer Vision* 88 (2010) 303-338.
- [10] Domenico D. Bloisi, Luca Iocchi, Andrea Pennisi, Luigi Tombolini. ARGOS-Venice Boat Classification, In 3rd Workshop on Vehicle Retrieval in Surveillance (VRS 2015) at AVSS 2015, pp. 1-6, 2015
- [11] D. Bloisi, L. Iocchi, M. Fiorini, G. Graziano, Automatic maritime surveillance with visual target detection, in: *Proc. of the International Defense and Homeland Security Simulation Workshop (DHSS), Rome, Italy*, pp. 141-145.
- [12] D. Bloisi, L. Iocchi, M. Fiorini, G. Graziano, Camera based target recognition for maritime awareness, in: *Proc. of the Fifteenth International Conference on Information Fusion (FUSION), Singapore*, pp. 1982-1987.
- [13] Allianz Global Corporate & Specialty AG, Safety and Shipping 1912-2012: From Titanic to Costa Concordia, 2012.
- [14] J. Leggat, T. Litvak, I. Parker, A. Sinha, S. Vidalis, A. Wong, Study on Persistent Monitoring of Maritime, Great Lakes and St. Lawrence Seaway Border Regions, Technical Report, DRDC CSS CR 2011-28, 2011.
- [15] Raytheon Canada, Surface based sensors for maritime domain awareness, *FrontLine Canada - Industry Special* 2 (2012) 37.
- [16] T. Nohara, A commercial approach to successful persistent radar surveillance of sea, air and land along the northern border, in: *IEEE international conf. on Technologies for Homeland Security*, pp. 276-282.
- [17] N. Pires, J. Guinet, E. Dusch, ASV: An innovative automatic system for maritime surveillance, *Navigation* 58 (2010) 1-20.
- [18] Thales Group, COMMANDER C3, <http://www.thalesgroup.com>, 2012.
- [19] DSIT Solutions, Harbor Surveillance System, <http://www.dsit.co.il>, 2012.
- [20] L-3 Klein, HarborGuard Command and Control, <http://www.l-3klein.com>, 2012.
- [21] B. J. Rhodes, N. A. Bomberger, M. Seibert, A. M. Waxman, SeeCoast: Automated port scene understanding facilitated by normalcy learning, in: *IEEE Conf. on Military Communications*, pp. 3656-3662.
- [22] M. D. Rodriguez Sullivan, M. Shah, Visual surveillance in maritime port facilities, in: *Proc. SPIE*, volume 6978, pp. 886-893.
- [23] R. G. Wijnhoven, K. van Rens, E. G. Jaspers, P. H. de With, Tracking ships from fast moving camera through image registration, in: *31st Symposium on Information Theory in the Benelux*, pp. 73-80.
- [24] N. Dalal, B. Triggs, Histogram of oriented gradients for human detection, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 886-893.
- [25] S. Fefilatyev, D. B. Goldgof, C. Lembke, Tracking ships from fast moving camera through image registration, in: *Int. Conf. on Pattern Recognition*, volume 4861, pp. 3500-3503.
- [26] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, *ACM Comput. Surv.* 41 (2009) 1-58.
- [27] O. Tuzel, F. Porikli, P. Meer, Pedestrian detection via classification on Riemannian manifolds, *PAMI* 30 (2008) 1713-1727.
- [28] P. Viola, M. J. Jones, Robust real-time face detection, *Int. Journal of Computer Vision* 57 (2004) 137-154.
- [29] OpenCV, <http://opencv.org>, 2012.
- [30] H. Bay, T. Tuytelaars, L. V. Gool, Surf: Speeded up robust features, in: *Proceedings of the Ninth European Conference on Computer Vision*, pp. 404-417.

[31] United Nation (UN), Convention on the law of the sea: Territorial sea and contiguous zone, 1982.