Camera based target recognition for maritime awareness

Domenico Bloisi, Luca Iocchi
Dipartimento di Ingegneria Informatica Automatica e Gestionale
Sapienza University of Rome
Rome, Italy
{bloisi,iocchi}@dis.uniroma1.it

Michele Fiorini, Giovanni Graziano
SELEX Sistemi Integrati S.p.A. Roma
Rome, Italy
{mfiorini,ggraziano}@selex-si.com

Abstract—**In this paper a framework for a camera based vessel recognition system is proposed. The framework is designed to enhance the functionalities of current Vessel Traffic Services (VTS) systems by adding a visual dimension to VTS data and the classification of non-cooperative targets. Furthermore, the framework can be suitable for the vessel traffic monitoring in populated areas where radar-based systems cannot be used due to electromagnetic radiation emissions. A quantitative evaluation of the detection performance on a publicly available dataset that validates the approach is provided.**

Keywords-coastal surveillance; target detection; object recognition; Vessel Traffic Services; maritime awareness.

I. INTRODUCTION

The control of vessel traffic in highly congested areas has become a critical requirement for safety and security. It is often correlated to environment protection issues, due to possible casualties and environmental disasters caused by tankers and vessels carrying dangerous goods. Moreover, the protection of the sea coasts is a requirement which has been given raising emphasis in recent years, due to the increasing threats coming from smugglers, refugees, intruders, and other non-conventional maritime forces (e.g., pirates).

The acronym VTS stands for Vessel Traffic Services and is used to name complex integrated systems for the acquisition, processing, and analysis of data in order to monitor the activities inland the Territorial Waters (12 nautical miles) [5]. Even if VTS systems are often equipped with long range surveillance cameras enslaved to the radar to enhance the early recognition of targets, the cameras are not used for automatic target recognition. Indeed, since the radar is the main sensor and cameras are used only as auxiliary sensors, target classification is usually possible only for cooperative targets equipped with Automatic Identification System (AIS) [6], while non-cooperative (non-AIS) targets are identified visually by the operator.

However, such a solution does not allow for target validation, i.e., identity verification of AIS targets (the AIS may be available or not, since it may not be activated or malfunctioning) and classification of non-AIS targets. Furthermore, it cannot be suitable for vessel traffic monitoring in populated areas where radar-based systems cannot be used due to electromagnetic radiation emissions. As an example, just consider the maritime traffic in the Venetian Laguna or in Venice down-town itself [1] and, more in general, in a touristic or residential environment where is not acceptable to live near a radar antenna.

The aim of this paper is to propose a framework for the realization of an environmental sustainable vessel traffic monitoring system able to automatically detect and recognize the targets with cameras as principal sensors. The framework is designed 1) to extend VTS functionalities, integrating the visual dimension and the management of non-cooperative targets 2) to be suitable for populated areas. An operative scenario where traditional VTS system can benefit from the proposed framework is presented.

The reminder of the paper is organized as follows. After analysing related work in Section 2, the proposed framework is detailed in Section 3. A possible operative scenario is described in Section 4 while in Section 5 a quantitative evaluation of the vessel detection module is provided. Conclusions are given in Section 6.

II. RELATED WORK

Maritime surveillance represents a challenge due to the complexity of the observed scene. Indeed, a huge amount of data coming from multiple and heterogeneous sensors needs to be fused in order to detect and track targets in wide area. Furthermore, the targets can have very different sizes (ranging from few to hundreds of meters in length) and their number can be very high (e.g., the number of small vessels during the summer months in a touristic location is typically very high).

In a recent study on persistent small vessel surveillance [7] several signal processing and information fusion solutions have been reviewed. They are briefly discussed in the following.

Raytheon Marine Small Target Tracker 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.

Accipiter Radar developed a radar surveillance network solution [8], 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.

SeeCoast [10] system detects, classifies and tracks vessels by fusing electro-optical (EO) and infrared (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 experiment 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.

Harbor Surveillance System[1] 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.

Thales Canada is developing a system called COMMANDER in which the Command, Control, and Communications (C3) nodes will be integrated through a satellite communications network to provide Canada-wide coverage. The system will enable real-time sharing of contact data, messages, and geo-referenced map overlays.

Data coming from radar, EO-IR, and sonar are used in the system called HarborGuard[2] 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.

ASV [9] is an automatic optical system for maritime safety using IR, GPS, and AIS. To detect relevant objects, the sea area 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.

Maximum average correlation height (MACH) filters are employed for vessel classification in [11]. Vessel detections are cross-referenced with ship pre-arrival notices in order to verify the vessel's access to the port. As reported by the Authors, such an approach tends to misclassify small vessels like speed boats and fishing boats.

As discussed in [7], none of the above presented systems addresses the requirements for performing in high clutter conditions, for tracking targets in ambiguous situations and for reporting suspicious activity. Moreover, the cited systems use the radar as the main sensor and such a solution cannot be suitable for populated areas.

In this paper, we propose a camera based framework in which the electro-optical cameras are the main sensors that can be used in substitution of the radar in particular situations and operative scenarios (see Section IV).

[1] http://www.dsit.co.il/siteFiles/1/84/5379.asp
[2] http://www.l-3klein.com/?page_id=13

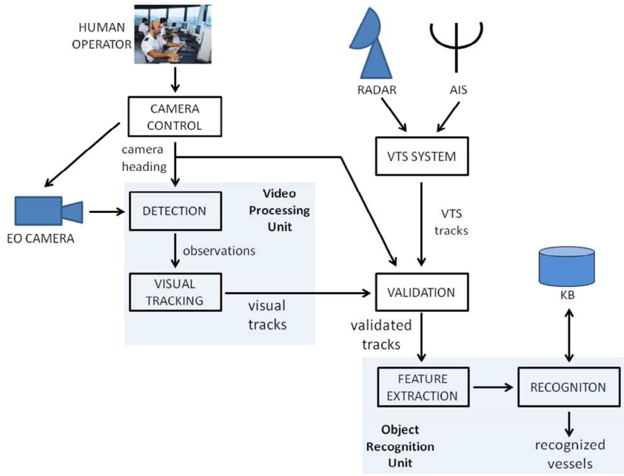

Figure 1. The proposed framework.

In addition, the proposed approach takes into account a series of criticisms related to automatic video surveillance in a port scenario [2]:

- the use of Pan-Tilt-Zoom (PTZ) cameras;
- the presence of targets having very different size;
- reflections and wakes on the water surface;
- the presence of apparently motionless boats anchored off the coast.

III. Framework for Camera Based Recognition

The functional architecture of the proposed framework is shown in Fig. 1.

The EO camera is the main sensor and can be moved by a human operator through a control module that provides orientation and field-of-view of the camera. The video processing unit is designed to detect and track vessels using also camera heading information. The detection task suffers from an additional difficulty since the camera can be freely moved by the operator, thus creating an highly dynamical background.

The validation module fuses information coming from the video processing unit and the VTS system. It aims at associating the visual tracks coming from the video analysis with the tracks provided by the VTS system. In this way, it is possible to provide the user with a new visual dimension in addition to the traditional geo-referenced VTS view (see Fig. 5). Once a track has been validated, the object recognition unit classifies the target according to its visual features.

In the reminder of this Section, each module of the framework is described in details.

A. Detection

The detection module processes the current video frame searching for possible targets. It is the most critical part of the system, since detection accuracy must be as high as possible, while maintaining an acceptable computational load.

Since the camera can be moved by the user, a foreground/background modeling approach to detect vessels is ineffective. A possible solution consists in using a classifier for detection. Different methods with different computational loads can be used to create the classifier [11,4]. In order to obtain real-time performance, we discarded computationally expensive methods (e.g., [3,12]) and adopted an approach based on Haar-like [13]. However, such an approach has been originally designed for face recognition, thus we verified the applicability of the method for boat detection.

Figure 2. Detection examples.

The output of the detection module is a list of observations, each observation being a bounding box representing a detected boat (see Fig. 2). Our classifier [2] is able to detect targets of different sizes, with different light conditions, and in presence of reflections on the water surface.

Along with the boats, it can be useful also to detect the limit of the sea surface (Fig. 3). Depending upon the heading of the camera, the system differentiates between sea-coast line (Fig. 3b) and sea-sky line (Fig. 3c). Since in presence of the coast the probability of finding false positives increases (Fig. 3a), it is possible to filter out the detections above the sea-coast line [2].

Figure 3. Sea-sky line and sea-coast line.

B. *Visual Tracking*

The visual tracking module takes in input the image flow and the output of the detection module and returns a set of visual tracks, i.e., bounding boxes with an identification number (see top left of Fig. 4). Its main role is the temporal filtering of the false positives.

The association between tracks and observations is made on the basis of a nearest-neighbor 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 (we set this threshold to 10): such a policy allows to filter out a high number of false positive detections.

Occlusions can occur sometimes when boats are aligned with respect to the camera view or when they are close to each other. Those situations cause considerable difficulty for visual tracking.

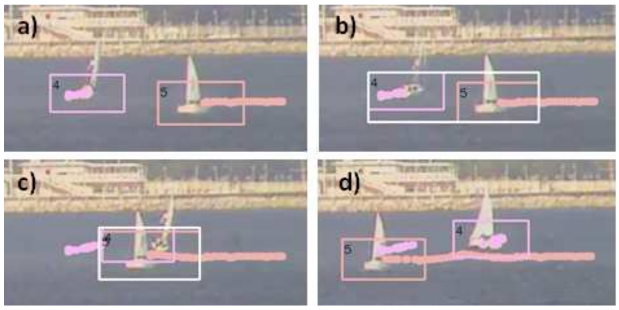

Figure 4. Group tracking example.

A possible solution is to consider collapsing tracks to form a group instead of a single track (see Fig. 4). When two or more tracks have bounding boxes moving closer to each other, the tracker starts considering them to form a group. Before the creation of the group, the HSV histograms of the involved tracks are saved. When the objects become visible again, the saved histograms are used to re-assign the correct identification number to the targets.

C. *Validation and Data Fusion*

The validation module aims at providing the user with a real-time visual image for the tracks, which is not available in today's VTS systems. The following steps are performed by the validation module:

1. Synchronization. Time synchronization between VTS data and video data is required to perform a consistent data association.

2. Localization. VTS tracks navigating in the FOV of the camera are individuated (see Fig. 5).

3. Rotation. VTS tracks are rotated with respect to the camera heading.

4. Projection. Visual and VTS tracks are projected in a common space (see Fig. 6).

5. Association. A probabilistic association between VTS and visual tracks is carried out.

Data fusion between video and VTS data is performed on a probabilistic basis. VTS and visual tracks are projected onto a two dimensional common space in order to perform the association. For the video data, the first dimension (*x*) is the distance (in pixels) of the bounding box from the left margin of the frame and the second one (*y*) is the distance from the bottom of the frame. For the 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 projection represent the *x* and *y* dimensions respectively (see Fig. 6).

Since the video frame and the geographic view present different scales, the dimensions are normalized with respect to the common space width and height. The projected visual and VTS tracks are associated on the basis of a nearest-neighbor policy.

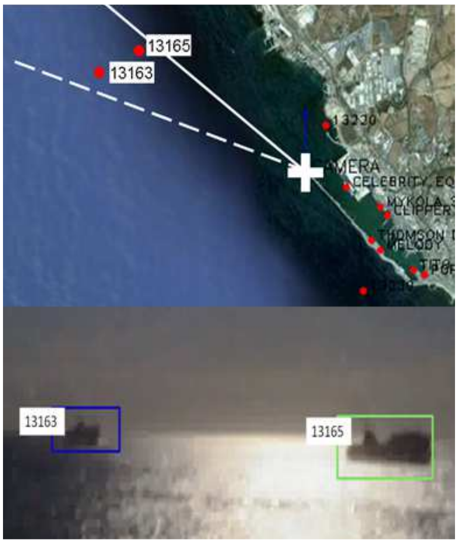

Figure 5. A new visual dimension is added to VTS geographical view.

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 (see bottom right of Fig. 6). This is the case shown in the bottom of Fig. 5, corresponding to the geographic view in the top of Fig. 5. The final output presented to the user is a global view of the situation at hand.

Figure 6. Visual and VTS track data fusion.

D. *Object Recognition*

A novel approach has been developed in order to recognize the tracked boats. The method is be able to deal with non-closed curves and low information images. The algorithm has been called “barcode matching” and it is detailed in the reminder of this Section.

The first steps of the “barcode matching” algorithm are the following (see Fig. 7).

1. The bounding box around the target is obtained by the Haar-based classifier (Fig. 7a).

2. The edges inside the bounding box are extracted, filtering out the horizon line (sea-sky line) if present. In this way, it is possible to obtain the target silhouette.

3. The target silhouette is resized in order to match a predefined model silhouette dimension (Fig. 7b). This step is crucial in order to compute the following ones.

4. An histogram is computed starting from the target silhouette. Each bin value is the distance of the target silhouette from the top of the image (Fig. 7c).

5. An histogram is computed starting from a model silhouette stored in a knowledge base. Each bin value is the distance of the model silhouette from the top of the image (this step can be performed off-line).

6. The target and the model histogram are compared on a distance metric.

The Bhattacharrya distance has been chosen to compare the target and the model histograms. The more similar the two histograms are, the lower the 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 rely on it completely to recognize an object. Indeed, it is very difficult to determine a threshold to recognize a target given a model histogram.

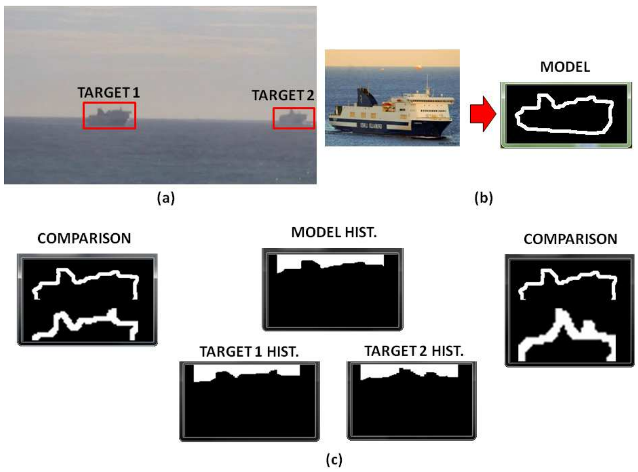

Figure 7. Target silhouettes extraction and comparison.

Thus, in order to improve the discriminative power of the method, an additional set of features can be used. Computing the mean value μ of the histogram, it is possible to identify regions in the histogram where the values are below μ (see Fig. 8). As an example, those regions can be produced by the chimney of the boat or by antennas. Comparing the number, the dimension and the position of the regions identified by the mean value filtering, it is possible to discard non-correct matchings. The idea is to obtain a sort of bar code for each histogram that can highlight particular features of the silhouette of the vessel.

For example, in Fig. 8 a difference in the position of the filtered regions can be noted. Indeed, the chimney of the first vessel is towards the bow, while the other one has its chimney in the middle. The mean filtering and the comparison of the obtained regions are the last two steps of the barcode matching.

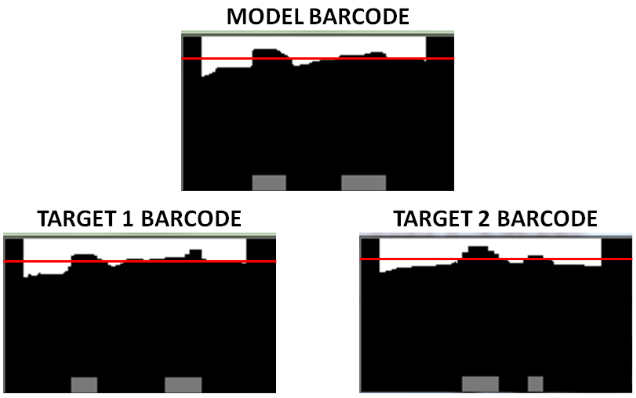

Figure 8. Barcode matching.

IV. OPERATIVE SCENARIO

Automatic systems for target tracking by camera slaved to radar are quite commonly present in the modern VTS or coastal surveillance systems in various configurations. They may be activated manually by the operator on a specific target or be fully automatic by relying upon an unique identification number such as IMO (Inter number), MMSI (Maritime Mobile Service Identifier) or even a unique system track number. The command and control software constantly processes the data associated to the track in order to identify the coordinates that must be passed on to the camera to control its motion. The above described process works with any identified track no matter which source has been used to identify it.

The drawback of this process is that the camera is slaved to the sensor or sensors used to track the target and stops if the track, for any reasons, is lost. The visual tracking removes the above limitation allowing the system to create, track and identify the target (i.e. the track) by means of cameras only. Of course this method is strongly affected by the field of view (FOV) of the camera and in particular by the angle of the target with respect to camera position. However, this last variable is rather stable in the real environment because there is always an approaching corridor to enter and to leave a port or transit area.

In Fig. 9 a tugboat is carrying some material going out of a port area. The screen-shot has been captured on an operative scenario where the tug is identified (named “ODYS” and marked with the green dotted square) because it was equipped with AIS and therefore detected by AIS and radar (fused track), while the material dragged is detected by radar only. In the same area there is also a third small vessel very near to the edge of the channel tracked by the radar. The photo, available on world wide web, is of the tugboat, but on a different context.

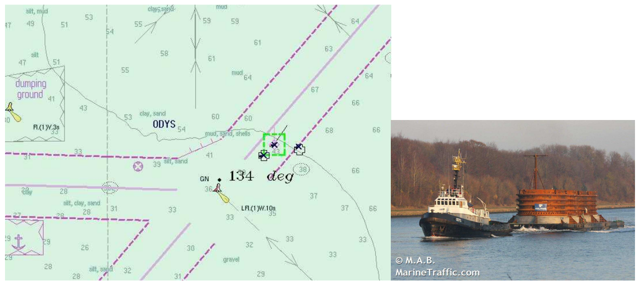

Figure 9. Operative scenario.

The above representative example demonstrates that in operative scenarios the angle between targets and cameras are de-facto almost the same (and generally they are the best possible view), because the camera is fixed on site and the targets follow almost the same path.

V. EXPERIMENTAL RESULTS

In order to quantitative evaluate the detection performance a set of test images as been used. Such a set has been built starting from the publicly available VOC database [4]. VOC database includes images containing different classes of objects. We used the images containing boats, discarding images that are not suitable for a VTS applications. The resulting set is composed of 163 images. The complete list can be found at

http://www.dis.uniroma1.it/~bloisi/papers/FUSION2012/list.txt

TABLE I. VISUAL DETECTION RESULTS.

Coastline Detection	DR	FAR
NO	0.872	0.332
YES	0.872	0.198

The results are reported in Table 1, in terms of detection rate (DR) and false alarm rate (FAR)

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

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

It is worth noting that visual tracking and data fusion modules can drastically reduce the FAR as well as improve the DR of the whole system thanks to temporal filtering and radar data. However, it is fundamental that frame-by-frame detection provides reliable results.

The analysis of the results shows that the coastline detection is crucial in obtaining a lower false alarm rate. 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 provided by the coastline analysis allows to improve considerably the system performance.

Furthermore, the tested detector has been built using images coming from real data recorded in a real site [2]. Those data differ significantly from the VOC dataset data, thus a detection rate of 0.872 can be considered acceptable for validating the detector performance. A higher detection rate of 0.928 has been obtained on real data [2].

The computational speed for the detection process using an Intel Core 2 Duo SU7300 CPU, 4 GB RAM is 10 fps computing 320x240 images.

VI. Conclusions and Future Work

In this paper a framework for a camera based vessel recognition system has been presented.

Differently from traditional VTS systems, the framework uses the EO camera has the main sensor. In this way, it should be possible to deploy automatic maritime surveillance systems in populated areas.

One of the major advantages of the proposed system concerns the possibility to provide the user with a global view of the situation at hand adding a visual dimension to VTS and AIS data.

The results on publicly available data coming from the VOC dataset show the effectiveness of the proposed detection approach, maintaining a 10 fps computational speed. The system is able to deal with user controlled motion of the camera, targets having very different size, reflections and wakes on the water surface, and apparently motionless boats anchored off the coast.

As future work, we intend to add the IR data to the data fusion scheme, to create a database with vessel silhouettes for the object recognition module and to perform the evaluation of the whole system.

References

[1] D. Bloisi and L. Iocchi, “Argos - a video surveillance system for boat traffic monitoring in Venice,” *IJPRAI*, vol. 23 (7), pp. 1477–1502, 2009.

[2] D. Bloisi, L. Iocchi, M. Fiorini, and G. Graziano, “Automatic maritime surveillance with visual target Detection,” In Proc. of the International Defense and Homeland Security Simulation Workshop (DHSS), pp. 141–145, 2011.

[3] N. Dalal and B. Triggs, “Histogram of oriented gradients for human detection,”, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893, 2005.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zissermann, “The PASCAL visual object classes (VOC) challenge,” International Journal of Computer Vision, 88(2), pp. 303–338, 2010.

[5] Canadian Coast Guard, “Vessel Traffic Services (VTS) update study”, OM: Canadian Coast Guard, Marine Navigation Services, vol. 159, 1991.

[6] I. Harre, “AIS adding new quality to VTS systems,” The Journal of Navigation. Cambridge Univ Press, vol. 03(53), pp. 527–539, 2000.

[7] J. Leggat, T. Litvak, I. Parker, A. Sinha, S. Vidalis, and A. Wong, “Study on persistent monitoring of maritime, Great Lakes and St. Lawrence seaway border regions,” Contract Report DRDC CSS CR 2011–28, 2011.

[8] T. Nohara, “A commercial approach to successful persistent radar surveillance of sea, air and land along the northern border,” IEEE international conf. on Technologies for Homeland Security, pp. 276–282, 2010 .

[9] N. Pires, J. Guinet, and E. Dusch, “ASV: an innovative automatic system for maritime surveillance,” Navigation, vol. 58(232), pp. 1–20, 2010.

[10] B. J. Rhodes, N. A. Bomberger, M. Seibert, and A. M. Waxman, “SeeCoast: Automated port scene understanding facilitated by normalcy learning,” Proceedings of IEEE Conf. on Military Communications, pp. 3656–3662, 2006.

[11] M. D. Rodriguez Sullivan, M. Shah, “Visual surveillance in maritime port facilities,” Proc. SPIE 6978, pp. 697811–8, 2008.

[12] O. Tuzel, F. Porikli, and P. Meer, “Pedestrian detection via classification on Riemannian manifolds,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30 (10), pp. 1713–1727, 2008.

[13] P. Viola and M. J. Jones, “Robust real-time face detection,” Int. Journal of Computer Vision, vol. 57(2), pp. 137–154, 2004.