AUTOMATIC MARITIME SURVEILLANCE WITH VISUAL TARGET DETECTION

Domenico Bloisi[(a)], Luca Iocchi[(b)], Michele Fiorini[(c)], Giovanni Graziano[(d)]

[(a)(b)] Department of Computer and System Sciences Sapienza University of Rome
[(c)(d)] SELEX Sistemi Integrati S.p.A. - Roma

[(a)]bloisi@dis.uniroma1.it, [(b)]iocchi@dis.uniroma1.it, [(c)]mfiorini@selex-si.com, [(d)]ggraziano@selex-si.com

ABSTRACT
In this paper an automatic maritime surveillance system is presented. Boat detection is performed by means of an Haar-like classifier in order to obtain robustness with respect to targets having very different size, reflections and wakes on the water surface, and apparently motionless boats anchored off the coast. Detection results are filtered over the time in order to reduce the false alarm rate. Experimental results show the effectiveness of the approach with different light conditions and camera positions. The system is able to provide the user a global view adding a visual dimension to AIS data.

Keywords: coastal surveillance, data fusion, border control, territorial waters.

1. INTRODUCTION
The increasing requests for safety, security and environmental protection, reveal the deficiency of the classic on-shore traffic management centres to satisfy the needs and requirements of the maritime domain. The traditional devices (Aids to Navigation, AtoN) and sensors (radar and Automatic Identification System, AIS) based on microwave tubes are starting to be replaced by solid-state equipments (Amato, Fiorini, Gallone, and Golino 2010) .

The centres are, as often as not, equipped with long range surveillance cameras enslaved to the radar to enhance the early recognition of targets. This approach is used in many Vessel Traffic Services (VTS) systems installed by SELEX Sistemi Integrati in Europe and abroad. It is useful both for monitoring vessels traffic, increasing safety at sea and facing illegal and hostile activities along the coastline.

SELEX Sistemi Integrati installed VTS and coastal systems in Yemen, Turkey, Serbia (fluvial), Panama, Poland, Russia (St. Petersburg), China and Italy.

In the VTS context, the target classification is done automatically by the system only for the cooperative targets equipped with AIS, while remaining targets should be identified manually by the operator.

Cameras are used for monitoring and providing pictures of the tactical situation inland the Territorial Waters (12 nautical miles).

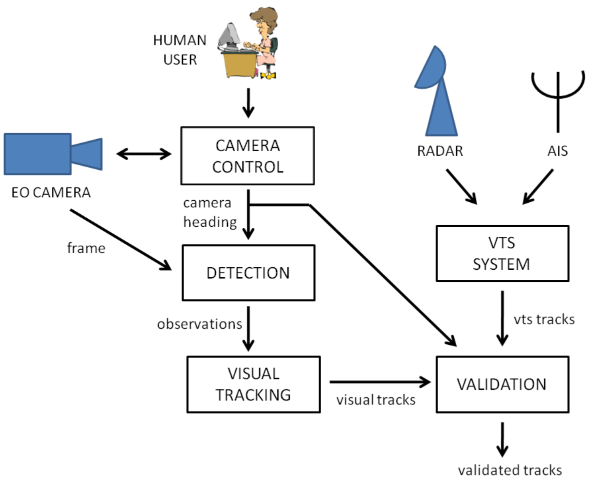

Figure 1: Architecture of a VTS/Coastal Surveillance Site

The cameras can operate in manual (under the operator control) or automatic (enslaved to a radar track) mode, but are not used for target validation, i.e., identity verification of AIS targets and classification of non-AIS targets.

The main contributions of the paper concern the description of a complex implemented system and the evaluation of its critical components. In particular, the detection approach is specifically designed in order to deal with the characteristic of the scenario.

The paper is organized as follows: after discussing related work in Section 2, the general architecture of the system is described in Section 3. In Section 4 we detail the visual detection module, while Section 5 shows quantitative results obtained by our approach on real data. Section 6 provides conclusions and future directions.

2. RELATED WORK
Maritime environment represents a challenging scenario for automatic video surveillance due to gradual and sudden illumination changes (e.g., clouds), motion changes (e.g., camera jitter), high frequency background objects (e.g., waves, raindrops) and reflections.

Although it is possible to develop systems based on electro-optical (EO) sensors only (e.g., ARGOS (Bloisi and Iocchi 2009)), the use of additional sensors such as infrared (IR), radar, Global Positioning System (GPS), and AIS can reduce the false positive detections thus improving the performance.

SeeCoast (Rhodes, Bomberger, Seibert, and Waxman 2006) is a system for coastal surveillance using EO and IR cameras, radar, and AIS. 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.

ASV (Pires, Guinet, and Dusch 2010) 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, due to wakes, such an approach can produce false positives.

A method for visual surveillance in maritime domain with non-stationary camera installed on an untethered buoy is presented in (Fefilatyev, Goldgof, and Lembke 2010). After the detection of the horizon line, a color gradient filter is applied to obtain a gray-scale image with intensities corresponding to the magnitude of color changes. Detection of the objects of interest is performed through thresholding of such gray-scale image into a binary map. The algorithm is limited by the assumption that all marine targets are located above the horizon line.

An object detection system for finding ships in maritime video is detailed in (Wijnhoven, van Rens, Jaspers, and de With 2010). The used approach is based on the Histogram of Oriented Gradients (HOG) (Dalal and Triggs 2005). 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.

Maximum average correlation height (MACH) filters are employed for vessel classification in (Rodriguez Sullivan and Shah 2008). 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.

The analysis of the literature shows a series of criticisms related to the port scenario:

- 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.

Our approach takes into account all of the above mentioned problems.

Figure 2: Boat detection

3. SYSTEM OVERVIEW

The general architecture of the system is depicted in Fig. 1.

3.1. Camera Control Module

The EO camera is a PTZ FLIR SR-TV with a 26x optical zoom, that can be moved by a human user through a control module. Furthermore, the control module is able to provide camera orientation and field-of-view.

3.2. Detection Module

The detection module takes as input the current frame acquired by the camera and the current heading of the camera. It is the most critical part of the system, since detection accuracy must be as high as possible while maintaining an acceptable computational load. The output of the detection module is a list of observations.

Each observation is a bounding box representing a detected boat (Fig. 2). All the details about the detection method are reported in Section 4.

3.3. Visual Tracking Module

The visual tracking module principal role concerns the temporal filtering of the false positives. The output of the tracking module is a set of visual tracks, i.e., bounding boxes with an identification number as in Fig. 3. Only tracks that present a sufficient number of observations are considered of interest (we set this threshold to 10).

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 and observation as measure.

We experimented also a CamShift (Bradski 1998) based approach obtaining poor results, probably due to the fact that CamShift creates color models taking histograms only from the hue channel in the HSV space which alone is not discriminative in the maritime scenario.

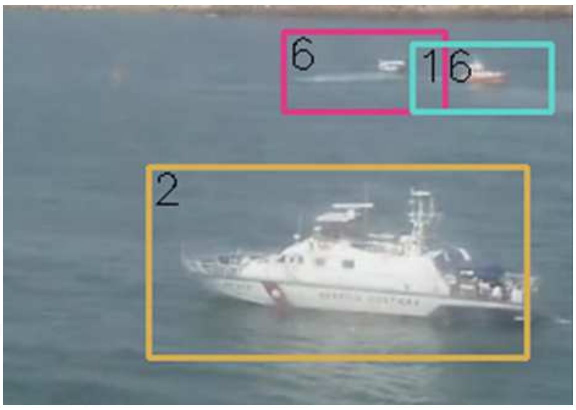

Figure 3: Visual tracking

Figure 4: Geographic projection of the VTS tracks. The camera (white +) is located at the entrance of the port

3.4. VTS Systems

A Coastal Surveillance Site (CSS) is generally responsible for detecting and tracking vessels in a range up to 12 nautical miles (NM), the defined “Territorial Waters” (United Nation 1982). Radar information are merged with AIS data in order to obtain a geographic view (see Fig. 4).

3.5. Validation Module

The validation module aims to give the user a real-time visual image for the tracks (Fig. 8), which is not available for AIS and non-AIS targets.

Data fusion between video and radar data is performed on a probabilistic base. VTS and visual tracks are projected onto a two dimensional common space in order to perform the association (see Fig. 9c). 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 radar data, the distance (in pixel) of the VTS track from the left side of the field of view and the distance from the camera position in the geographic projection represent the x and y dimensions respectively.

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.

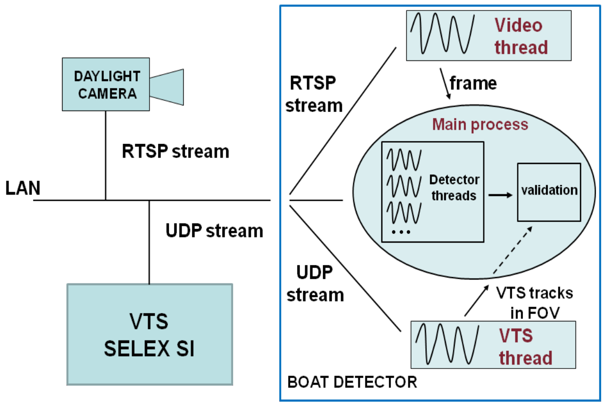

Figure 5: Communication between modules.

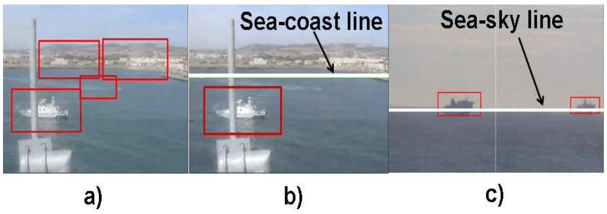

Figure 6: Sea-sky line and sea-coast line.

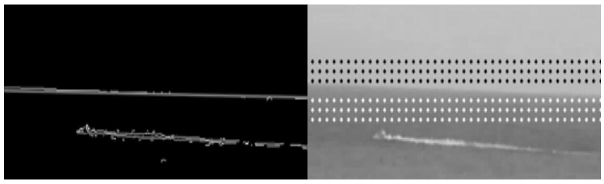

Figure 7: Sea-sky line detection

The proposed data fusion scheme is justified by the fact that the boats are moving on a planar surface, thus a boat closer to the camera than another will appear in a lower position in the image. This is the case depicted in Fig. 8, corresponding to the geographic view in Fig. 4.

In Fig. 9 the output for the detection (Fig. 9a), the visual tracking (Fig. 9b), and the data fusion (Fig. 9c) in case of partially overlapping targets are shown.

3.6. Inter-module Communication

The inter-module communication scheme is reported in Fig. 5. The VTS system, the EO camera, and the detector are connected through a LAN. Video frames are transmitted by the camera using RTSP, while VTS data are transmitted by the SELEX Sistemi Integrati’s VTS system via UDP.

The detector is made of a main process managing the detection algorithm (performed by a set of threads in order to speed up the detection) and the validation module and two threads managing the different refresh rates of the incoming data. Indeed, video data are transmitted at 25 fps, VTS data have a refresh period of about 2 seconds, and the detector is able to compute 10 fps (see Section 5).

In the rest of the paper we analyze in detail the detection algorithm, that represents the most innovative part of the system.

Table 1: Visual detection results

Classifier	Coastline Detection	DR	FAR
Without wake examples	NO	0.892	0.475
Without wake examples	YES	0.892	0.265
With wake examples	YES	0.928	0.251

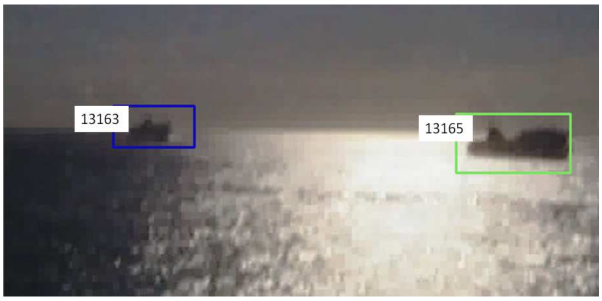

Figure 8: Validated tracks. Visual and tracking data are fused in a single view

4. VISUAL DETECTION

The detection module aims to find the boats in the current frame obtained from the PTZ camera. Since the camera is frequently moved by the user, a foreground/background modelling approach to detect vessels is ineffective. Thus, we decided to adopt a classifier based detection. In order to obtain real-time performance, computationally expensive methods (e.g., (Dalal and Triggs 2005; Tuzel, Porikli, and Meer 2008)) were discarded and a Haar-like features based approach (Viola and Jones 2004) has been adopted. Indeed, its main advantages are the computational speed and the possibility of combining different classifiers in a cascade.

However, Haar-like classifier has been originally designed for face recognition, thus we verified the applicability of the method for boat detection. At this aim, we used the OpenCV HaarTraining functions.

A set of 4000 images not containing boats and a set of 1500 images depicting different types of boats taken from the internet have been used as input for the offline training stage obtaining a 24 level classifier (the training was stopped when the false alarm rate reached 5×10^{-6}).

Given a single frame, the boats in the image are detected. Along with the boats, also the limit of the sea surface is detected. Depending by the heading of the camera, the system differentiates between sea-coast line (Fig. 6b) and sea-sky line (Fig. 6c).

Since in presence of the coast the probability of finding false positives increases (top of Fig. 6a), it is possible to filter out the erroneous detections laying above the sea-coast line (Fig. 6b).

In order to detect the limit of the sea surface, the Hough transform is applied to the edge map of the frame and the candidate lines are validated with respect to a set of points belonging to a rectangular region containing the line (right part of Fig. 7). If at least the 90% of the corresponding points above and under the line present different intensity values, then the line is considered valid. In this way, it is possible to filter out misdetections due to long wakes (left part of Fig. 7).

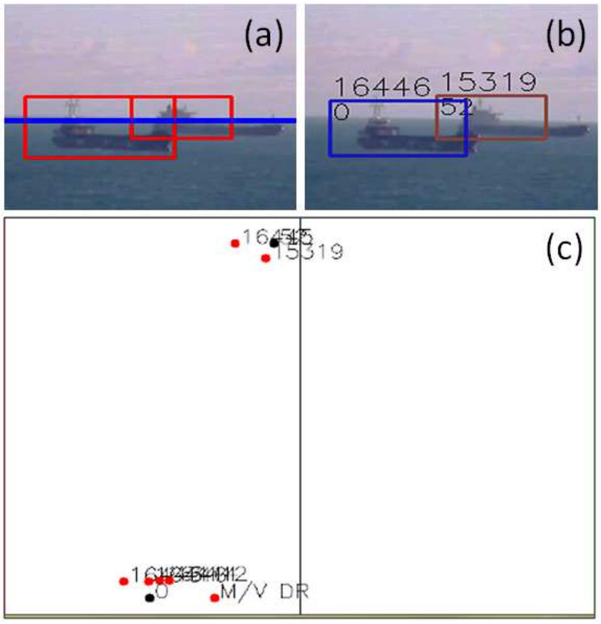

Figure 9: Visual analysis can help in correcting erroneous detection made by the radar. Detection results (a) are filtered over the time to obtain visual tracks (b) that are projected over a common space together with radar and AIS tracks (c)

5. RESULTS ON VISUAL DETECTION

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 Haar 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 24 level classifier. 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.

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.

Visual analysis can help also in correcting radar errors. In presence of big targets (e.g., oil tankers in Fig. 9), it is possible to obtain multiple radar tracks for the same object (red dots in Fig. 9c). Those multiple radar detections can be merged if a single visual track has been individuated (the black dots in Fig 9c referring to the visual tracks in Fig. 9b). The computational speed for the system using an Intel Core 2 Duo SU7300 CPU, 4 GB RAM is 10 fps computing 320x240 images.

6. CONCLUSIONS

In this paper an automatic video surveillance system for maritime environment has been presented.

The system is able to deal with user defined movements of the cameras, targets having very different size, reflections and wakes on the water surface, and apparently motionless boats anchored off the coast.

The main goal of the system is to provide the user a global view of the situation at hand adding a visual dimension to AIS data.

The results on real data show the effectiveness of the proposed detection approach maintaining a 10 fps computational speed.

As future work, we intend to complete the evaluation of the whole system (that is currently underway) and to add the IR data to the data fusion scheme in order to further improve the performance of the system.

REFERENCES

[1] Amato, F., Fiorini, M., Gallone, S., Golino, G., 2010. Fully Solid State Radar for Vessel Traffic Services. *Proceedings of International Radar Symposium (IRS)*, pp. 1-5. June 16-18, Vilnius, Lithuania.

[2] Bloisi, D., Iocchi, L., 2009. Argos - a Video Surveillance System for boat Traffic Monitoring in Venice. *IJPRAI*, vol. 23 (7), pp. 1477-1502.

[3] Bradski, G. R., 1998. Computer video face tracking for use in a perceptual user interface. *Intel Technology Journal, Q2*.

[4] Dalal, N., Triggs, B., 2005. Histogram of oriented gradients for human detection. *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 886-893. June 20-26, San Diego, CA, USA.

[5] Fefilatyev, S., Goldgof, D.B., Lembke, C., 2010. Tracking Ships from Fast Moving Camera through Image Registration. *Proceedings of the Int. Conf. on Pattern Recognition*, pp. 3500-3503. August 23-26, Istanbul, Turkey.

[6] Pires, N., Guinet, J., Dusch, E., 2010. ASV: An Innovative Automatic System for Maritime Surveillance. *Navigation*, vol. 58 (232), pp. 1-20.

[7] Rhodes, B.J., Bomberger, N.A., Seibert, M., Waxman, A.M., 2006. SeeCoast: Automated Port Scene Understanding Facilitated by Normalcy Learning. *Proceedings of IEEE Conf. on Military Communications*, pp. 3656-3662, October 23-25, Washington, DC, USA.

[8] Rodriguez Sullivan, M.D., Shah, M., 2008. Visual surveillance in maritime port facilities. *Proc. SPIE 6978*, pp. 697811-697811-8.

[9] Tuzel, O., Porikli, F., Meer, P., 2008. Pedestrian Detection via Classification on Riemannian Manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30 (10), pp. 1713-1727.

[10] Viola, P., Jones, M.J., 2004. Robust Real-Time Face Detection. *Int. Journal of Computer Vision*, vol. 57 (2), pp. 137-154.

[11] Wijnhoven, R.G.J., van Rens, K., Jaspers, E.G.T., de With, P.H.N., 2010. Online Learning for Ship Detection in Maritime Surveillance. *Proceedings of the 31st Symposium on Information Theory in the Benelux*, pp. 73-80. May 11-12, Rotterdam, The Netherlands.

[12] United Nation (UN), 1982. Convention on the Law of the Sea: Territorial Sea and Contiguous Zone.

AUTHORS BIOGRAPHY

Dr. Domenico Bloisi received his Ph.D., M.Sc., and B.Sc. degrees in computer science engineering from Sapienza University of Rome, Italy in 2010, 2006, and 2004 respectively. He is currently a post-doc in the Department of Computer and System Science, Sapienza University of Rome, Italy. His main research interests are related to object detection, visual tracking and data fusion in automatic video surveillance.

Dr. Luca Iocchi received his Master (Laurea) degree in 1995 and his Ph.D. in 1999 from Sapienza University of Rome. He is currently Assistant Professor in the Department of Computer and System Science, Sapienza University of Rome, Italy. He is the author of more than 100 referred papers in international journals and conferences. His main research interests are in the areas of cognitive robotics, action planning, multi-robot coordination, robot perception, robot learning, stereo vision and vision based applications.

Dr. Michele Fiorini was educated at Ancona (Italy) and Bath (UK) Universities, receiving his PhD in 2006, and is now a senior systems engineer at SELEX Sistemi Integrati, a Finmeccanica company, in Rome, Italy. His working interests are mainly concerned with command, control, communications, computers, intelligence, surveillance and reconnaissance (C4ISR), homeland protection and vessel traffic service (VTS) as well as in electromagnetism theory and antenna design.

Giovanni Graziano received his Master (Laurea) degree in 2007 from Federico II University, Naples, Italy. Since 2007 he is a systems engineer at SELEX Sistemi Integrati, a Finmeccanica company, in Rome, Italy. His working interests are related to localization systems and tracking algorithms for radar and electro-optical sources.