Optical Target Recognition for Drone Ships

M. Fiorini
Leonardo S.p.A., Rome, Italy

A. Pennisi
Katholic University Leuven, Leuven, Belgium

D.D. Bloisi
University of Verona, Verona, Italy

ABSTRACT: Remote controlled drone ships without crews on board are expected by the end of the decade. To achieve the goal of developing (semi-)autonomous boats, reliable vision-based methods for vessel detection, classification, and tracking are needed. In this paper, we present a machine learning approach for vessels detection from a moving and zooming camera. In particular, the proposed method is supervised and derives from a fast and robust people detection algorithm. Quantitative experimental results have been obtained on a publicly available data set, which contains images from real sites, demonstrating the effectiveness of the approach. Ground truth annotations and the code of the proposed algorithm are both released for the community.

1 INTRODUCTION

Drones are widely attracting attention in the media after the Internet activist Pirate Party has managed to interrupt Chancellor Angela Merkel and Defence Minister at a CDU campaign event in Dresden, Germany on September 16, 2013 making use of a miniature drone started circling above the audience.

UAVs (Unmanned Aerial Vehicles) technology is widely available to hobbyists and environmental scientists at affordable prices. Applications include land management, animal conservation, crop monitoring, and disaster mapping. The marine and maritime sector is not excluded from the possibility of using UAVs based solution. For example, underwater robots are able to gather environmental information, scour and sediment transport analysis, meanwhile drone ships will enable new business models in near future. According to Oscar Levander, head of the innovation marine unit at Rolls-Royce, remote controlled drone ships without crews on board may generate, by the end of the decade, a similar disruptive effect as the one provoked by Uber, Spotify and Airbnb on other industries. Totally autonomous ships are just a step ahead.

Optical tracking features are now present at different stages of development and integration in almost all surveillance applications, fixed or mobile, equipped with cameras. However, in order to allow those technologies to be used for autonomous vessels, targets recognition, i.e., classification, are needed. The target identification process, coupled with a decision support software module, allows to rise warning issues for potential collisions and to modulate speeds. Moreover, in the context of maritime boarder control and Search And Rescue (SAR), vessels patrolling are still a widely used procedure. These operations require considerable effort and resources, which could be considerably reduced by autonomous patrolling vessels.

The aim of this work is to give a general overview of existing optical based recognition solution in the maritime context and to present a machine-learning based approach for vessel classification and detection, which is a fundamental requirement to achieve autonomous navigation.

The remain of the paper is organized as follows. Section 2 provides an overview of existing techniques for vessels, humans, and floating objects detection. The proposed approach is presented in Section 3. Qualitative and quantitative experiments are shown in Section 4. Finally, conclusions are drawn in Section 5.

2 RELATED WORK

Maritime environment represents a challenging scenario for automatic object detection due to the complexity of the observed scene: High frequency background objects (e.g., waves on the water surface), boat wakes, and weather issues (e.g., heavy

raindrops) contribute to generate a highly dynamic scenario (Bloisi et al., 2014).

2.1 *Vessel Detection and Tracking*

Vessel detection in the maritime scenario requires the monitoring of large areas at different resolution levels. Indeed, the objects of interest can have very different size, ranging from few to hundreds of meters in length. SeeCoast system (Rhodes et al., 2006) detects, classifies, and tracks vessels by fusing electro-optical (EO) and infrared (IR) video data with radar and AIS data. 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 by Sullivan & Shah (2008). 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 boats. Fefilatyev et al. (2010) propose a system exploiting a non-stationary camera installed on an untethered buoy. After detecting the horizon line, a color gradient filter is applied to obtain a grayscale image with intensities corresponding to the magnitude of color changes, then thresholding on the grayscale image is used to find objects of interest. The algorithm is limited by the assumption that all marine targets are located above the horizon line. ASV (Pires et al., 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. An object detection system for finding ships in maritime video based on Histogram of Oriented Gradients (HOG) is described by Wijnhoven et al. (2010). 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.

2.2 *Floating Objects Detection*

Floating objects detection plays an important role in USVs (Unmanned Surface Vehicles). Indeed, obstacles in the operational environment can be floating pieces of wood or other debris, which presents a significant challenge to continuous detection from images taken on-board (Kristan et al., 2016). Snyder et al. (2004) state that transitory obstacles, such as floating debris, are best detected and analyzed at navigation time with visual means. They propose a system for fully autonomous navigation in a river scenario. Obstacles and objects of interest are tracked across multiple cameras (by using feature clusters in aspect-elevation space) and then mapped onto the 3D world. Distant moving objects are detected and tracked by clustering feature points, while nearby movers are detected with motion blobs. Stereo rigs are used by Huntsberger et al. (2011) for obstacle and moving objects detection on a USV. However, since a large baseline is required for granting a large field of view coverage, this can create instability for small vessels. A method for detecting water regions in videos by clustering color and texture features is proposed by Santana et al. (2012). Fefilatyev et al. (2009) use the horizon line position to eliminate all edges not belonging to floating objects: it is assumed that within an image, all objects of interest lie above the horizon line. A similar idea is exploited by Wang et al. (2011): They first detect the horizon line and then search for a potential obstacle in the region below the horizon.

The main drawback of approaches based on the horizon line detection is that situations in coastal waters, close to the shoreline, cannot be easily handled, since the edge of water does not correspond to the horizon.

2.3 *Humans Detection*

In many USV applications, humans are considered just a special case of obstacles to be avoided (Almeida et al., 2009). However, distinguishing between human and non-human detections is important in order to devise different and more opportune navigation strategies, for example in case of casualty detection for search and rescue operations. Differently from a static floating objects, a person may be swimming or diving, possibly without being aware of an approaching USV, and this poses serious concerns for human safety. In recent years, the use of thermal images from Forward-Looking Infrared (FLIR) cameras for target detection and tracking has become popular in various application domains (Sanna & Lamberti, 2014).

Independent FLIR cameras have also been mounted on vessels and used for target detection and tracking at sea (Kim & Lee, 2014). Martins et al. (2013) uses a combination of thermal and colour images to detect casualties from a USV in the context of the FP7 project ICARUS. A simple horizon detection algorithm was applied to the thermal image to limit the search space to the water surface only, then targets detected by both (fixed and calibrated) cameras were tracked using Kalman

filtering to deal with false positives and temporary false negatives. Although computationally efficient, the detection coverage of the proposed solution is limited by the narrow field of view and by the minimum focusing distance of the thermal camera, which made the system unsuitable for tracking humans in proximity of the USV. To increase the water surface area covered by the cameras, the latter can be mounted on a stabilized pan-tilt unit (PTU), as proposed by Bibby & Reid (2005). Underwater human detection can be achieved instead by using multibeam or mechanically-rotated sonars, similarly to what is already done for obstacle detection with autonomous surface vehicles (Heidarsson & Sukhatme, 2011).

3 PROPOSED APPROACH

To detect the boats, a method based on Aggregated Feature Channels (AFC) has been adopted. The method has been presented by Dollar et al. (2014) and is made of three main steps:

1 Feature extraction;
2 Pyramid computation;
3 Classification.

Each step is detailed below.

3.1 *Feature Extraction*

To represent a target, a set of feature channels is extracted from the images. Given an image I and a function Ω that represents the process of extracting features, we indicate a channel as $C = \Omega(I)$. Then, the information contained in each channel is aggregated over multiple pixel by summing (Σ) every block of pixels in C, and smoothing the resulting lower resolution channels. In such a way, a feature is a single pixel lookups in the aggregated channels. The process is shown in Figure 1.

The feature vector contains the following features: normalized gradient magnitude, histogram of oriented gradients (6 channels), LUV color channels, and integral image. The channels are divided into 4×4 blocks and the pixels in each block are summed. Finally, the channels are smoothed by applying a Gaussian filter.

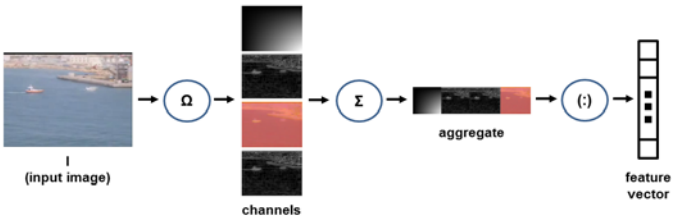

Figure 1. Functional architecture of the process for generating the feature vector. The input image is split in channels, then the channels are aggregated using a lower resolution. As in (Dollar et al., 2014), the feature vector is composed by single pixel lookups in the aggregated channels.

3.2 *Pyramid Computation*

A feature pyramid is a representation of an image at multiple scales. Usually, the number of scales s is between 4 and 12 in a log-space starting from $s = 1$. The common approach to build a pyramid is to compute a set of feature channels at each scale. However, this can be computationally expensive. Instead of computing for each scale a large amount of features, Dollar et al. (2014), propose a way to approximate the features within channels. Being I_s the scaled image of I, the feature channels C_s is computed by using an approximation of the information contained in $C = \Omega(I)$. In particular,

$$C_s \approx R(C,s)s^{\lambda\Omega} \quad (1)$$

where $R(C,s)$ represents the resampled features by s, while λ is a constant. Equation 1 is valid not only for the images, it is valid also for any corresponding window w_s and w in I_s and I, respectively.

C_s is computed at one scale per *octave*, where an octave is the interval between one scale and another with half or double its value. While, at the intermediate scales, it is computed as

$$C_s \approx R(C_{s'}, s/s')\,(s/s')^{-\lambda\Omega} \quad (2)$$

where $s' = \{1, ½, ¼, \ldots\}$ is the nearest scale for which $C_{s'} = (\Omega I_{s'})$.

The above described approach represents a good compromise between speed and accuracy. Indeed, the cost of evaluating for the approximated scales is within 33% of computing $\Omega(I)$ at the original scale, and moreover, the channels do not need to be approximated more than half an octave.

1.1 *Classification*

A boosted tree classifier is used for detecting the boats. The classifier combines 2048 depth-two trees over all the candidate features (i.e., the channel pixel lookups) in each window. Since a vessel can be contained into a bounding box with height x and width $2x$, the size of the window is equal to 128×64 pixels. The detector has a step size of 4 pixels and 8 scales per octave.

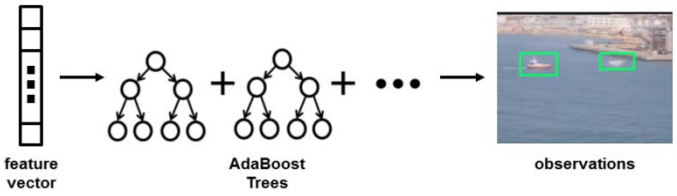

Figure 2. Adaptive Boosting is used to obtain the observations about the objects of interest. As in (Dollar et al., 2014), a set of decision trees is built over the feature vector provided by in order to distinguish object from background.

Figure 2 shows the process for obtaining the final observations, where multiple vessels can be detected in the same image. For training the classifier, a total of 288 boat samples from four different videos has been collected. The annotated images and the videos used for the training stage, as well as the video sequences used for the testing phase (including ground-truth data) are available at: https://github.com/apennisi/fast_vessel_detection

4 EXPERIMENTAL RESULTS

An experimental evaluation has been carried out on visual data coming from a real site to validate the proposed approach. In particular, we have decided to use a video from the publicly available Maritime Detection, Classification, and Tracking (MarDCT) data base (http://www.dis.uniroma1.it/~labrococo/MAR/). MarDCT (Bloisi et al., 2015) contains a collection of videos and images captured with different camera types (i.e., fixed, moving, and Pan-Tilt-Zoom cameras) and in different scenarios. The aim of MarDCT is to provide visual data that can be used to help in developing intelligent surveillance system for the maritime environment.

Figure 3. Boat detection using the proposed method on different videos from the MarDCT data base. The algorithm runs at multiple scales and can detect vessels of different size with varying lighting conditions.

Qualitative Evaluation. The proposed approach has been applied to several sequences. In particular, we tested 6 different videos from a moving and zooming camera. The experiments show (see Fig. 3) that our approach achieves good results in most of the videos, also considering that a small data set (288 samples) has been used for training the classifier. In the top left image of the Figure 3, the big cruise boat is not detected since no samples of similar boats are contained in the training set.

Quantitative Evaluation. In order to evaluate the robustness of the detection module, we have carried out quantitative quality measurements by calculating the Precision, the Recall (or True Positive Rate - TPR), and the F1-score. The used metrics are defined as follows.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\text{-}score = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{5}$$

where TP is the number of the true positive observations, FP is the number of false positives, and FN is the number of false negatives. F1-score gives a weighted average of the precision and recall.

The results of the experiments are shown in Table 1 and Table 2 and are totally reproducible, since the algorithm's code and the testing video are provided.

Table 1. The proposed approach has been tested extracting 116 frame samples from a video from a moving and zooming camera

Num. Frames	Num. Samples	TP	FP	FN
116	82	66	30	14

Table 2. Quantitative Results. Three different metrics has been used to measure the quality on the proposed approach on a real video.

Precision	Recall	F1-score
0,688	0,825	0,375

Runtime Performance. The detector has been implemented by using MATLAB for the training phase of the classifier and C++ for the testing stage. The training phase is offline and should be computed once, while the code of the testing stage has been realized to allow a real-time computation. In particular, for 640×480 images, the complete pipeline runs at about 30 fps (frames per second).

5 CONCLUSIONS

In this paper we have presented a fast and robust method for vessel detection from moving and zooming cameras. The proposed algorithm represents a valid baseline for building a Maritime Unmanned Navigation system. In particular, our algorithm is derived from a solid supervised method

originally conceived for people detection, which runs at real-time speed (Dollar et al., 2014).

A quantitative experimental evaluation has been carried out on videos coming from the publicly available database MarDCT, containing data from different real sites captured with varying lighting conditions.

As future work, we intend to extend the method to run with images coming from omni-directional (360°) cameras and to add a module for the coastline detection. Indeed, tracking the coastline can provide useful information for the heading (yaw) of the drone ship, while pitch and roll values could be obtained by inertial sensors on board.

REFERENCES

Almeida, C. et al. 2009. Radar based collision detection developments on USV ROAZ II. In *Oceans - Europe*, 1-6.

Bibby, C. & Reid, I. 2005. Visual Tracking at Sea, In Proc. of *IEEE Int. Conf. on Robotics and Automation*, 1841-1846.

Bloisi, D. D. Pennisi, A. & Iocchi, L. 2014. Background modeling in the maritime domain. Machine Vision and Applications 25(5): 1257-1269.

Bloisi, D. D. Iocchi, L. Pennisi, A. & Tombolini, L. 2015. ARGOS-Venice Boat Classification. In *IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, 1-6.

Dollar, P. Appel, R. Belongie, S. & Perona, P. 2014. Fast feature pyramids for object detection, IEEE Trans. Pattern Anal. Mach. Intell. 36: 1532–1545.

Fefilatyev, S. Goldgof, D. B. & Lembke, C. 2009. Autonomous buoy platform for low-cost visual maritime surveillance: design and initial deployment. In Proc. SPIE 7317, Ocean Sensing and Monitoring, 73170A.

Fefilatyev, S. Goldgof, D. B. & Lembke, C. 2010. Tracking Ships from Fast Moving Camera through Image Registration. In *Proc. of the Int. Conf. on Pattern Recognition*, 3500-3503.

Heidarsson, H. & Sukhatme, G. 2011. Obstacle detection and avoidance for an Autonomous Surface Vehicle using a profiling sonar. In *IEEE Int. Conf. on Robotics and Automation*, 731-736.

Huntsberger, T. Aghazarian, H. Howard, A. & Trotz, D. C. 2011. Stereo vision based navigation for autonomous surface vessels. JFR 28(1): 3–18.

Kim, S. & Lee, J. 2014. Small Infrared Target Detection by Region-Adaptive Clutter Rejection for Sea-Based Infrared Search and Track. Sensors 14(7): 13210-13242.

Kristan, M. Sulić Kenk, V. Kovačič, S. & Perš, J. 2016. Fast Image-Based Obstacle Detection From Unmanned Surface Vehicles. IEEE Trans. on Cybernetics 46(3): 641-654.

Martins, A. et al. 2013. Field experiments for marine casualty detection with autonomous surface vehicles, In *Oceans - San Diego*, 1-5.

Pires, N. Guinet, J. & Dusch, E. 2010. ASV: An Innovative Automatic System for Maritime Surveillance. Navigation 58(232): 1-20.

Rhodes, B. J. Bomberger, N. A. Seibert, M. & Waxman, A. M. 2006. SeeCoast: Automated port scene understanding facilitated by normalcy learning. In *Proc. IEEE Military Communications Conference*, pp. 1–7.

Sanna, A. & Lamberti, F. 2014. Advances in Target Detection and Tracking in Forward-Looking InfraRed (FLIR) Imagery. Sensors 14(11): 20297-20303.

Santana, P. Mendica, R. & Barata, J. 2012. Water detection with segmentation guided dynamic texture recognition. In *IEEE Robotics and Biomimetics.*

Snyder, F. D. Morris, D. D. Haley, P. H. Collins, R. T. & Okerholm, A. M. 2004. Autonomous river navigation. In *Proc. SPIE 5609, Mobile Robots XVII*, 221.

Sullivan, M. D. R. & Shah, M. 2008. Visual surveillance in maritime port facilities. In *SPIE Optics and Photonics*, 697 811–697 811.

Wang, H. Wei, Z. Wang, S. Ow, C. Ho, K. & Feng, B. 2011. A vision based obstacle detection system for unmanned surface vehicle. In *Int. Conf. Robotics, Aut. Mechatronics*, 364–369.

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. In *Proc. of the 31st Symposium on Information Theory in the Benelux*, 73-80.