# Practical Results for Buoy-Based Automatic Maritime IR-Video Surveillance

**Zigmund Orlov / Wolfgang Krüger / Norbert Heinze**
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB
Fraunhoferstraße 1, 76131 Karlsruhe
Germany

email: zigmund.orlov@iosb.fraunhofer.de / wolfgang.krueger@iosb.fraunhofer.de / norbert.heinze@iosb.fraunhofer.de

## *ABSTRACT*

*Criminal activities at sea have been a reality for past years. In such activities, predominantly small maritime vessels are used, which are difficult to detect. Until now, border agencies observe and protect the critical maritime areas by ships, planes or helicopters. Therefore, surveillance is expensive and full coverage is difficult to obtain.*

*To improve this situation, the European research project AMASS (Autonomous Maritime Surveillance System) investigates to use a network of unmanned surveillance platforms, equipped with different sensors. An important sensor is an uncooled thermal imager. In the AMASS project, Fraunhofer IOSB developed an image exploitation module for maritime surveillance systems, which was described in [1] and [2].*

*In this paper, we will present the practical experience, challenges and performance results of the buoy-based system achieved under real conditions in shallow water of the Atlantic Ocean in the Melenara Bay on Gran Canaria. Despite of rough conditions, the image exploitation provided respectable results.*

## 1 INTRODUCTION

Criminal activities at sea such as illegal immigration, piracy or trafficking of drugs, weapons and illicit substances have been a reality for past years. In such activities, predominantly small maritime vessels are used, which are difficult to detect. Until now, border agencies observe and protect the critical maritime areas by ships, planes or helicopters. Therefore, surveillance is expensive and full coverage is difficult to obtain.

To improve this situation, the European research project AMASS (Autonomous Maritime Surveillance System) investigates to use a network of unmanned surveillance platforms. The platforms are equipped with different sensors. An important sensor is an uncooled thermal imager. In order to exploit the data delivered by this thermal imager, detection and tracking algorithms, which are able to work with a moving sensor under a variety of weather and visibility conditions, are required.

In the AMASS project, Fraunhofer IOSB developed an image exploitation module for maritime surveillance systems. It automatically detects and tracks distant vessels in the images generated by thermal imagers on an autonomous mobile platform such as a buoy or a ship.

There is only little information in the literature about visual surveillance from moving autonomous platforms deployed at sea. An example is [3] where the authors describe an un-tethered autonomous buoy that stations itself on the sea floor and is able to ascend to the surface when needed. The installed optical surveillance unit is based on a low-power minicomputer and processes colour images from a web-camera.

The camera is more or less at sea-level and no physical stabilization is available. Ship detection and tracking is based on previous work of the authors [4], [5]. Due to limitations in computing power, video data is collected in an online phase and image exploitation has to be done offline afterwards.

Compared to other approaches, Fraunhofer IOSB developed a robust multi-algorithm solution, exploiting complementary image cues and integrated them in flexible multi-layer software architecture, which was described in [1] and [2]. The system is robust with respect to variations of boat appearance, image quality, and environmental conditions. In this solution each algorithm can be developed individually or can be easily substituted by another module. The fundament of the image exploitation is a detection layer which provides the results of several detection algorithms in a motion-stabilized scene-coordinate frame aligned with the estimated horizon line. In the autonomous system, detections are used to trigger alarms and to facilitate tracking.

In this paper, we will present the practical experience, challenges and performance results of the buoy-based system achieved under real conditions in shallow water of the Atlantic Ocean in the Melenara Bay on Gran Canaria. Despite of rough conditions, the image exploitation provided respectable results.

## 2 SYSTEM ARCHITECTURE

Figure 1 depicts the hardware architecture of the image exploitation on an autonomous platform. There are four components: camera, pan-tilt unit (PTU), the inertial measurement unit (IMU) and the image-exploitation computer (IE-PC)

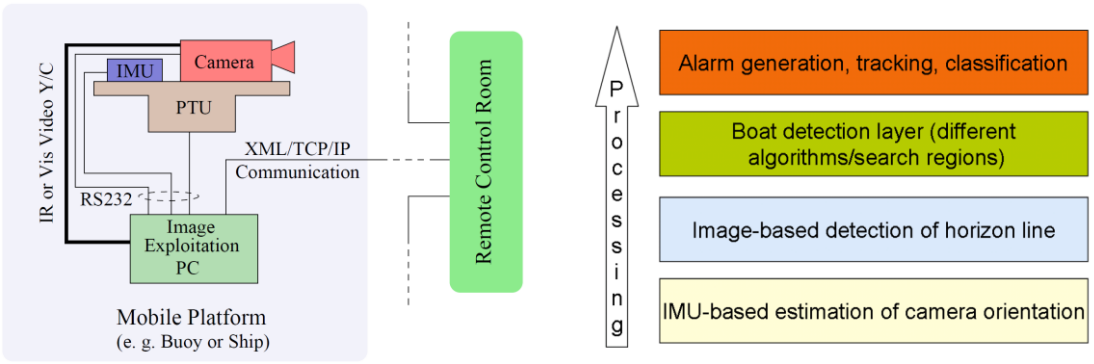


**Figure 1: Overall architecture (left), layer structure of the image exploitation algorithms (right).**

The IMU and the camera are located together on the turret of the platform. Both components are mounted on the PTU. The task of the IMU is to measure the angular orientation of the camera with respect to a world coordinate frame. The measured roll and pitch angles are needed by the image exploitation software in order to estimate position and slope of the horizon line for processed camera images.

The IE-PC is mounted in a waterproof rack. It is connected to the Ethernet switch, the IMU, the PTU and the camera. Its task is to run image exploitation algorithms using received IMU data, to control camera and pan-tilt unit, and to perform communication with a remote control room by using XML-based messages over TCP/IP.

An operator in the remote control room has to evaluate the information received from sensor system and to make decisions how to act.

The image exploitation software is structured according to the layer model in Figure 1. Each layer builds on the results of the previous layer and data processing is carried out from bottom to top:

- The first and lowermost layer is the estimation of camera orientation using an inertial measurement unit (IMU). Input for this layer is the measurement data generated by sensors (accelerometers, gyros, and magnetometers) inside the IMU. An extended Kalman filter is used to estimate the time-varying pitch, roll, and yaw angles of the camera to which the IMU is attached.
- The next layer uses the estimated pitch and roll angles to find and improve the localization of the horizon line in the captured camera images. The horizon line is determined by a robust fit to edge features (local jumps of image brightness) extracted from the images. Pitch and roll angles from the IMU-layer are used to narrow the search areas for feature extraction.
- The third software layer is the boat detection layer. Since the aim is to detect small distant boats and those boats will appear near the horizon line, the boat detection layer uses the information about the horizon line in an image to set up search areas for the implemented detection algorithms. The search areas are fixed relative to the horizon line. The boat detection algorithms are based on searching for temporally stable image features. (e.g. bright blobs in thermal images). The challenge is to separate detections at boats from those at sea clutter.
- The final and topmost software layer uses the generated detections to compute internal alarms or to perform tracking. In order to do this, results from the detection layer are fused and classifiers are used to separate relevant detections at boat/ships from false or irrelevant detections.

# 3 TEST ENVIRONMENT

## 3.1 Scenario

The test was carried out under real conditions in shallow water of the Atlantic Ocean in the Melenara Bay on Gran Canaria.

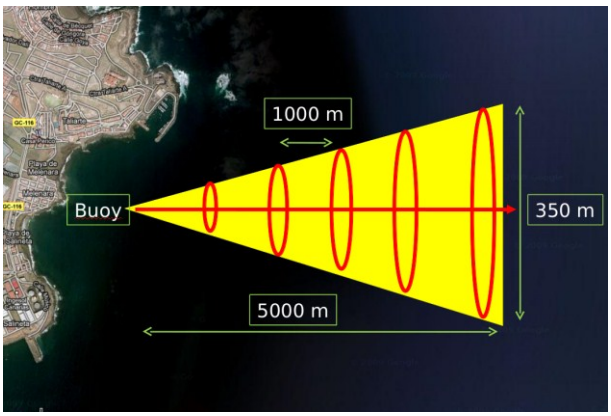


**Figure 2: Sketch of the course of the rubber boat in the Melenara Bay test (not drawn to scale).**

It was performed at about noon. A rubber boat (5.8 m long, 2.3 m wide) with four occupants travelled from the buoy up to 5 km in eastward direction performing a loop after every singe kilometre (Figure 2) After having reached the final distance of 5 km, the boat turned towards the buoy and travelled back with similar intermediate loops. The speed of the boat was about 10 knots. The loops were scheduled in order to provide different aspects of the boat and to have a visual distance indication. The boat was equipped with GPS which was used to measure the distances to the buoy with the camera. The state of the sea was

normal. In total about two hours of image and motion data was recorded (119 sequences with duration of one minute).

## 3.2 Data Analysis

Due to stringent requirements on hardware costs and power consumption, no active camera stabilization could be used. Therefore, the data showed a pronounced angular motion causing large vertical displacements (pitch motion around horizontal image axis) as well as horizontal displacements (yaw motion around vertical image axis) in the images. Such a large angular motion had the consequence that the number of image frames having a boat in the field of view was low. Large pitch motion caused many images to show only water or sky, and large yaw motion led to many images with a boat located outside of the left or right image borders. Additionally, there were images with strong motion blur.

### 3.2.1 Movement Data

The distribution of the pitch angle values during the whole test is depicted in Figure 3. The camera has a vertical field of view of about 3.3 degrees. The pitch angle of 0° means that the horizon line is in the middle of the camera image. The interval of 3.3 degrees is marked around the 0° by two red dashed lines in the figure. 28.2% of pitch positions, measured in the test, were outside this interval (only-water or only-sky images, see remark in Table 1.

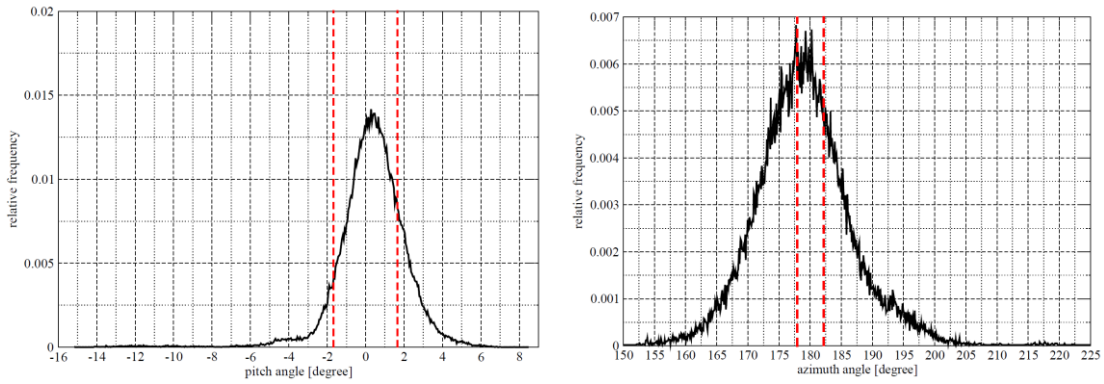


**Figure 3: Histogram of pitch angle (left) and azimuth angle (right).**

Figure 3 shows the distribution of the azimuth angle. The camera has an horizontal field of view of about 4.4 degrees. The angle set in the test according to the agreed boat heading was 180°. The camera direction was corrected, if necessary, during the test within moderate time intervals. Around the 180°, the interval of 4.4 degrees corresponding to the horizontal view of field is marked by two red dashed lines in the figure. About 75% of azimuth positions, measured in the test, were outside this interval (see remark in Table 1). The values of roll angles were not critical. Statistic values of angle distributions are shown in Table 1.

This large camera movement had high dynamic (see Figure 4). Sometimes, the angular velocity of the pitch motion was about 10 degrees per second. During the integration time of the image sensor (16 ms) we find that we must have had an angular motion of 9.6 arc-minutes. This fact led to the images with strong motion blur.

**Table 1: Statistic values of measured camera angles.**

| | pitch angle $\beta$, ° | azimuth angle $\alpha$, ° | roll angle, ° |
|---|---|---|---|
| Number of samples | 178500 | 178500 | 178500 |
| Minimum | -15.17 | 150 | -6.953 |
| Maximum | 8.534 | 225.6 | 8.861 |
| Mean | 0.3479 | 179 | -0.1269 |
| Standard deviation | 1.845 | 7.682 | 1.617 |
| Variance | 3.404 | 59.01 | 2.616 |
| Coefficient of variation | 5.303 | 0.04292 | 12.75 |
| Median | 0.4157 | 178.7 | -0.2309 |
| 1st quartile | -0.5255 | 174.2 | -1.216 |
| 3rd quartile | 1.382 | 183.3 | 0.8756 |
| Remark | $p(\beta < -1.65) = 0.0847$<br>$p(\beta > 1.65) = 0.1977$ | $p(\alpha < 177.8) = 0.4489$<br>$p(\alpha > 182.2) = 0.2993$ | |

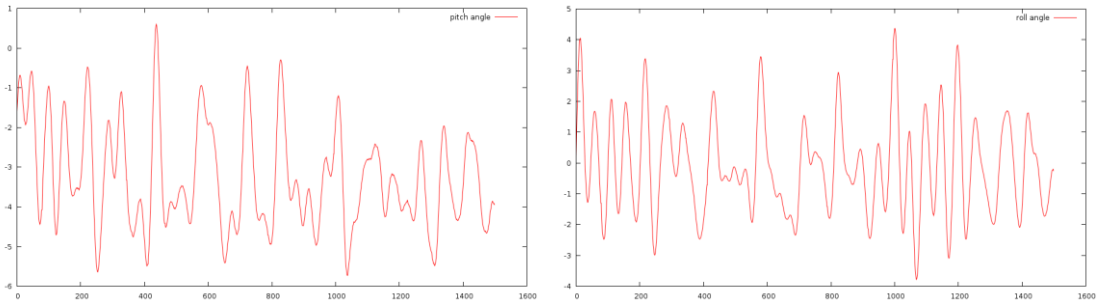


**Figure 4: Pitch angle (left) and roll angle (right) estimated by the adapted Kalman filter. Angles are in degrees and the horizontal plot axis is the number of the corresponding image frame.**

### 3.2.2 Boat Occurrence

The large angular motion, its dynamic and occasional deviation of heading of boat due to sea streams led to low number of image frames having a boat in the field of view. An analysis of the data confirmed this by showing that in only 6.6% of the total number of images the boat was in the field of view.

Figure 5 shows the relative time amount of boat observation in each sequence recorded. We see that there are sequences without boat occurrences. The boat is visible in 76 sequences.

Table 2 statistically shows how long the boat was visible without interruptions in the relevant 76 sequences. Mean number of frames, in which the boat was continuously visible, amounts to 30 (frame rate is 25 fps). The median is 22 frames. The whole histogram of the continuous boat observation is presented in Figure 5.

**Table 2: Statistical data of the continuous boat observation time.**

| Number of observations | Minimum, frames | Maximum, frames | Mean, frames | 1st quartile, frames | 2nd quartile, frames | 3rd quartile, frames |
|---|---|---|---|---|---|---|
| 398 | 1 | 713 | 30.03 | 12 | 22 | 36 |

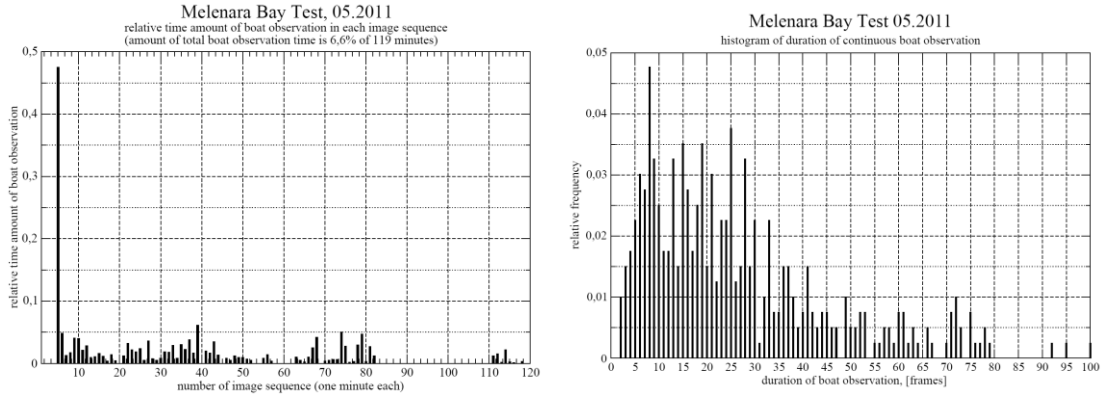


**Figure 5: Relative time amount of boat observation in each sequence (left) and histogram of the duration of continuous boat observation (right).**

# 4 SOFTWARE ADJUSTMENTS AND SETTINGS

The software described in Section 2 and in [1] - developed with North Sea (Helgoland) data recorded from a ship - had to be adapted to the buoy scenario with the rough test conditions in the Atlantic Ocean.

The first adjustments had to be made at the IMU-layer in order to adapt the signal processing (Kalman filter) to the more agile motion. The filter designed on the basis of the available data from a different buoy was not appropriate. The adaption was based on image and motion data from a pre-test in order to have the already improved signal processing available for the test phase involving boats (see Section 3.1). Optimizing the parameters of the Kalman filter required "ground truth" data for angular motion which was derived by estimating pitch angles from the position of the horizon line in the images.

The parameterization and robustness of the image-based detection of the horizon line had to be improved because the Melenara Bay images had a much higher amount of noise than Helgoland data which was available for algorithm development. In addition, the algorithms were improved to better cope with multiple contrast changes at the imaged horizon line.

The large pitch motion and offsets in IMU signal processing made it necessary to design and implement a fusion procedure which gives priority to the image measurements and uses relative IMU measurements to eliminate bias. In addition, a very large search area for the image based detection of the horizon line had to be used which increases the risk of detection errors.

The large angular camera motions also had consequences for the boat detection algorithms:

- Stronger variation of boat positions between image frames made it necessary to improve the prediction of image frame motion from the estimated horizon line and all angular IMU measurements (roll, pitch, and yaw).
- Exploitation of temporal consistency to separate detections in sea clutter from those at boats is not as effective because boats are too shortly in the field of view.
- Very large search areas for boat detection became necessary which in conjunction with the reduced effectiveness of clutter suppression by temporal consistency leads to more false detections. This was moderated by an improved temporal filtering of detection thresholds.

The higher fixed pattern noise (vertical stripes) than in the previously available Helgoland data made it

necessary to design a non-uniformity correction algorithm. The compass data proved as too coarse and inaccurate for fusion of detection results during alarm generation. Therefore, an appropriate fusion of compass data with IMU yaw measurements had to be developed. The classifier described in [2] was trained with Melenara Bay data from the pre-test.

The tracking algorithm, developed on the video material from Helgoland, operates on the results of the detection layer in a symbolic space (scene data structure) that is stabilized relating to the roll and pitch movements of the sensor. The detection symbols can be tracked using spatial and temporal detection information in the stabilized scene data structure (with the same related north direction) within one basic tracking step (sensor could be considered as stabilized relating to the north direction). Due to a strong sensor movement in the Melenara Bay test data resulting in a seldom occurrence of the observed boat in the field of view, a long observation time is needed to perform one basic tracking step. Within this time, strong and quick changes of the sensor's north direction (more than one image width) can happen. The tracking decision must be made within one basic step consisting of multiple observations, where each of them has its own horizon-stabilized scene data structure with a different north direction reference. Therefore, relevant changes relating to the tracking program flow, program states, data structures and parameters were necessary and have been performed.

# 5 RESULTS

The image exploitation algorithms were able to detect the rubber boat having a length of 5.8 m and a width of 2.3 m up to the maximum distance of 5 km. Figure 6 and 7 show examples of detection results at the various boat distances measured by the GPS on the rubber boat.

It can be seen that the rubber boat becomes very small at larger distances. It is difficult to achieve stable detections at distances of 4 km and 5 km. There were also some false detections at waves. From the detection results at the various distances it can be seen that the boat detection algorithms are able to handle the quite different scales of the imaged boat.

For alarm generation two different algorithms were used. The first one is a voting procedure fusing the single detections collected from a small time window. The second algorithm is tracking based and determines stable tracks of a boat before it generates an alarm.

In order to suppress false detections at sea clutter, it is advantageous for the voting procedure to use the largest time window possible. Unfortunately, the unexpectedly large angular motion of the camera considerably shortens the number of image frames for which boats are continuously in the field of view. Therefore, only a very short time window of 16 image frames (i.e. 0.64 seconds) had to be used.

The total number of processed image frames was 178500 (119 sub-sequences, each having a length of 1500 image frames (one minute)) and the voting procedure was executed 11067 times. In total, 468 internal alarms were generated, which corresponds to an alarm rate of 4.2%. Remembering that according to our visual analysis the boat was only 6.6% of the recording time in the field of view, that alarm rate is reasonable. A visual inspection showed, that if the 468 generate internal alarms had been generated, 415 of the alarms would have been be correct with detections at the rubber boat, and 53 alarms would have been false detections at waves. This corresponds to a true positive rate of 89%. With the classifier described in [2], it was possible to decrease the number of false positive alarms from 53 to 9, which corresponds to an increase of true positive rate from 89% to 98%.

The advantage of the voting procedure is the ability to detect almost all boat occurrences (also very short occurrences). The disadvantage is the stateless nature of the algorithm and its short observation time, which results in many multiple internal alarms and increased CPU load to eliminate them or increased

bandwidth requirements to transmit them to the land station. Another consequence is the limited ability to suppress false detections at sea clutter in cases of short boat occurrences. To avoid this disadvantage, a tracking-based alarm generation algorithm was developed that on his part is not able to detect very short boat occurrences due to the tracking principle.

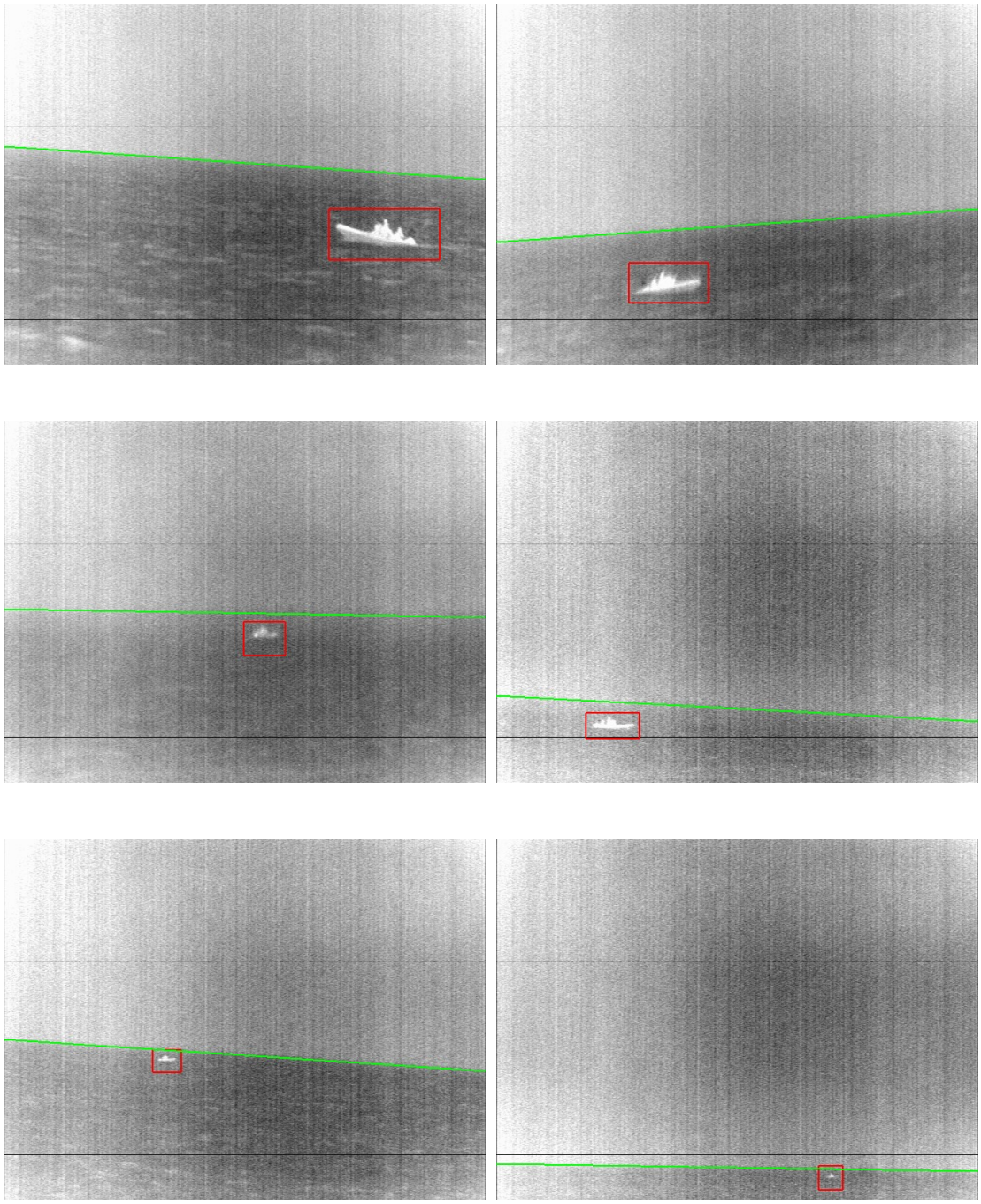

**Figure 6: Boat detections (red bounding box) and estimated horizon line (green) at a distance of about 500 m (top), 1 km (middle), and 2 km (bottom).**

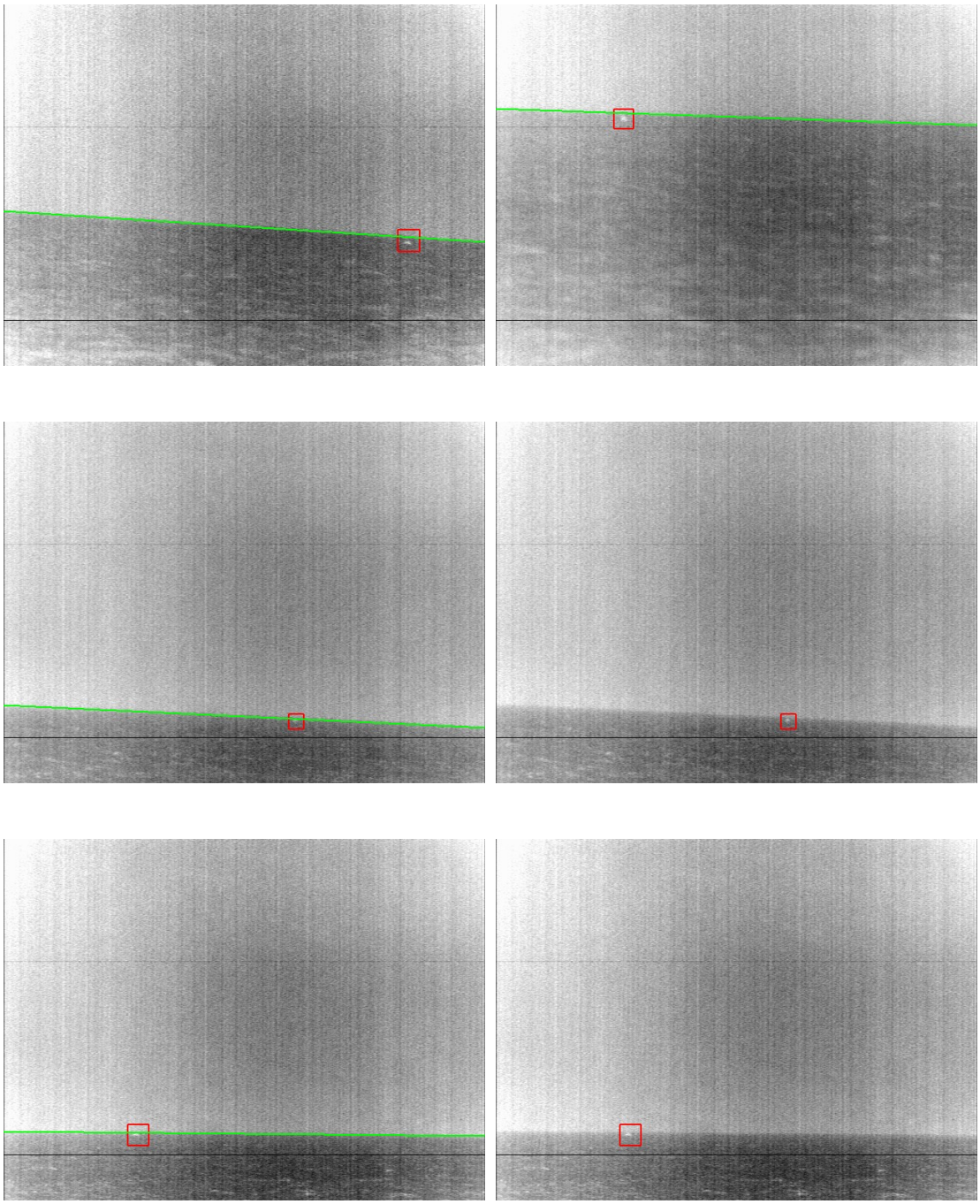

**Figure 7: Boat detections (red bounding box) and estimated horizon line (green) at a distance of about 3 km (top), 4 km (middle), and 5 km (bottom). For distances 4 km and 5 km the original images are shown for better visualization on the right hand side.**

The voting algorithm for the alarm generation was used with a fixed duration of the sea observation (16 frames, see the explanation before). The tracking-based alarm generation was investigated with three values of a threshold for the alarm generation. The threshold is the number of frames with the same detected boat. The thresholds were chosen based on the parameter of the continuous observation time distribution (see Table 2).

As explained in Section 3.2.2, the boat was visible in 76 sequences (from 119). As the alarm generation with tracking method is applied for each sequence without knowledge of past sequences, 76 alarms should be generated in an ideal case, namely one alarm for each sequence with a boat occurrence. Among these 76 sequences, there are 12 sequences with only singular and very short (5 to 22 frames) occurrence of the observed boat. We do not consider these sequences as sequences where the boat must be detected. Therefore we only have 64 sequences with relevant boat occurrences. Consequently, the ideal number of alarms should be 64.

Tracking-based alarm generation, where the minimum number of frames (for a new alarm) with the same detected object is 16 / 24 / 30 / 38, is called here 16f- / 24f- / 30f- / 38f-tracking. The number 24 is chosen based on the median value for the distribution of continuous boat observation time (Table 2). The number 30 equates to the mean value. The number 38 corresponds around to the 75% quantile.

**Table 3: Comparison of alarm generation approaches and their settings.**

| | Voting, 16f | 16f-tracking | 24f-tracking | 30f-tracking | 38f-tracking |
|---|---|---|---|---|---|
| Total number of alarms | 468 | 114 | 84 | 72 | 50 |
| Number of true positive alarms | 415 | 91 | 77 | 71 | 49 |
| Rate of true positive alarms | 89% | 79.8% | 91.7% | 98.6% | 98% |
| Number of false positive alarms | 53 | 23 | 7 | 1 | 1 |
| Rate of false positive alarms | 11% | 20.2% | 8.3% | 1.4% | 2% |

Table 3 compares the tracking-based alarm generation with different threshold settings and the voting-based alarm generation. We see that increase of observation duration threshold from 16 to 38 strongly reduces the rate of false positive alarms. On other hand, the sensitivity of the alarm generator decreases. The number of true positive alarms falls from 91 (for 16f) to 49 (for 38f). This table shows the effective suppression of false positive alarms (detection on waves) and elimination of multiple alarms in comparison to the voting approach. Nevertheless, there are still multiple alarms. One reason is the fact that long interruptions in the boat observation lead to multiple alarms because of a changed boat position. The other possible reason is inaccuracy of measuring instruments (compass and IMU).

**Table 4: Alarm rates for all sequences (119) with boat visibility larger than 900 milliseconds (22 frames) during 1 minute long sequences. Relevant boat occurrences were in 64 sequences.**

| | 16f-tracking | | 24f-tracking | |
|---|---|---|---|---|
| | Alarm-true | Alarm-false | Alarm-true | Alarm-false |
| Truth-true (64) | 58<br>**90.6%** | 6<br>**9.4%** | 54<br>84.4% | 10<br>15.6% |
| Truth-false (119) | 17<br>14.3% | 102<br>85.7% | 4<br>3.4% | 115<br>96.6% |

Alarm rates presented in Table 4 are measured for the system that produces alarms with a one-minute-tact with following model: the multiple true positive alarms *within one sequence* are counted as one true

positive alarm; all the false positive alarms *within one sequence* are counted as one false positive alarm.

Table 4 presents results for the case, in which only relevant boat occurrences were used for the alarm rate computation. As already said before, 64 relevant sequences were identified. Then, the ideal result should be 64 alarms. We see that 16f-tracking yields the best result with 58 alarms (90.6%). On the other hand, 16f-setting leads to 14.3% of false positive alarms. But they can be efficiently suppressed by the following classification. The reason, why 16f-setting does not reach higher true positive alarms is the inaccuracy of measuring instruments, because tracking relies on the compass and IMU data to make a tracking decision. 24f-tracking has a better suppression of false positive alarms (3.4%) but lower detection rates.

With all the investigated settings the boat could be detected at the maximum distance from the buoy (5 km).

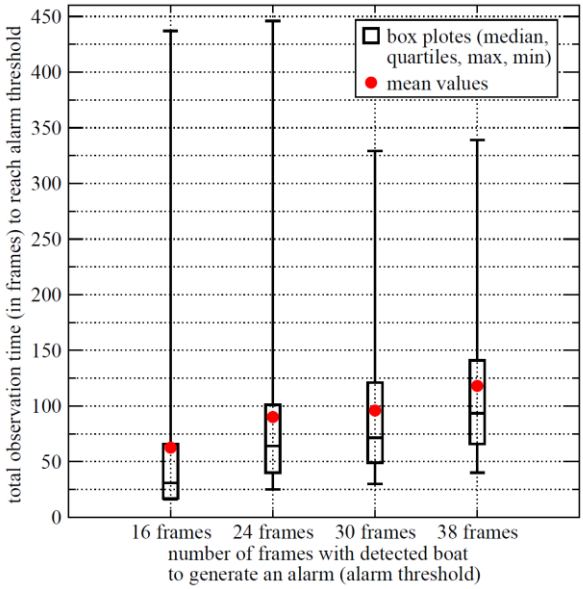


**Figure 8: Distributions (box plots) of total observation times to generate an alarm.**

The behaviour of the total observation time of the detected boat, in order to trigger an alarm, is shown in Figure 8 depending on alarm threshold. This is a box plot diagram with the median, 25% quantile, 75% quantile, minimum and maximum values. Red points represent mean values of the presented distribution. We see that the total observation time of the detected boat increases (between 0.25 s and 1 s, frame rate is 25 fps) with the increasing alarm threshold. On the other hand, we sometimes see very long total observation times (maximum values). For example, 450 frames mean 18 s at the frame rate of 25 fps.

16f-Tracking can be used in conjunction with classification [2] for this type of video material in order to reach a high alarm rate and a small false positive alarm rate as well as a good observation time performance (see Figure 8).

## 6 CONCLUSION

Fraunhofer-IOSB has developed the image exploitation subsystem for maritime surveillance and it was evaluated during the deployment of the buoy in Melenara Bay, Gran Canaria. Evaluation was carried out

up to 5 km distance with a rubber boot (5.8 m long with 4 persons).

Task of the image exploitation subsystem is to detect and track small distant vessels with a thermal imager (uncooled) with long focal length on a moderately stabilized autonomous platform. Therefore, IOSB developed a robust multi-algorithm solution, exploiting complementary image cues and integrated them in flexible multi-layer software architecture.

It turned out that the movements of the optronics platform (buoy) were large. This resulted in challenging working conditions for the image exploitation subsystem. Despite of rough conditions, the image exploitation provided respectable results.

For detection and tracking-based alarm generation without classification encouraging results could be obtained:

- Evaluation of 119 video clips with a length of 60 seconds each resulted in a 90.6% true positive detection rate with a 14.3% rate of false positive alarms. This includes the results of a detection range of 5 km, where also high detection rates could be shown.
- These results were obtained with a mean observation time of the vessels of less than 3 seconds. The median observation time is approx. 1 second. In the field of main application of such a system much longer observation times than 3 seconds can be expected. This has the benefit that detection rates can be tuned towards even higher detection rates and lower false alarm rates in exchange for acceptable longer observation times (time until an alarm is generated).

The classification described in [2] showed further improvement of results, by sorting out false alarms. The number of false alarms (false positives) could be reduced by the classification step by a factor of 5.

## ACKNOWLEDGMENT

This work has been done in cooperation with Carl Zeiss Optronics GmbH and Instituto Canario de Ciencias Marinas and was supported with funds from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. SP1-Cooperation-218290.

## REFERENCES

[1] W. Krüger and Z.Orlov, Robust Layer-based Boat Detection and Multi-target-tracking in Maritime Environments, Proc. of the 2010 NURC 2nd International Waterside Security Conference (WSS2010), Marina di Carrara, Italy, November 2010.

[2] M. Teutsch and W. Krüger, Classification of small Boats in Infrared Images for maritime Surveillance, Proc. of the 2010 NURC 2nd International Waterside Security Conference (WSS2010), Marina di Carrara, Italy, November 2010.

[3] S. Fefilatyev et al., Autonomous Buoy Platform for Low-Cost Visual Maritime Surveillance: Design and Initial Deployment, Proc. SPIE, Vol. 7317, Ocean Sensing and Monitoring, 2009.

[4] S. Fefilatyev et al., Towards detection of marine vehicles on horizon from buoy camera, Proc. SPIE, Vol. 6736, Unmanned/Unattended Sensors and Sensor Networks IV, 2007.

[5] S. Fefilatyev and D. B. Goldgof, Detection and tracking of marine vehicles in video, 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1-4, 2008.