2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)

# An Algorithm for the Detection of Vessels in Aerial Images

Jorge S. Marques, Alexandre Bernardino
ISR/IST Lisboa, Portugal
{jsm,alex}@isr.ist.utl.pt

Gonçalo Cruz, Maria Bento
CIAFA, Portugal
{gccruz,mfbento}@academiafa.edu.pt

## Abstract

*In this paper we present a sea vessel detection algorithm in aerial image sequences acquired by an unmanned aerial vehicle. The proposed method is robust to variable background lighting, highlights due to sun reflections, vehicle self motion and scale changes. By relying in simple blob analysis rules, based on both spatial and temporal constraints, the algorithm is capable of real-time operation onboard the vehicle, even with non optimized code. We evaluate our method on three sequences labeled with ground truth vessel position, with more that 2900 frames. Overall we are able to achieve very low false positive rates even in heavy sun reflection conditions.*

Figure 1. Vessel detection difficulties: moving camera, non stationary and dynamic background, sun reflections and sky.

## 1. Introduction

The scientific world uses unmanned aerial vehicles for ocean and climate research, environment and Earth research, magnetic, radiological, gravimetric mapping, and geophysical monitoring of natural processes, among others. Unmanned Aerial Vehicles are also used as platforms to develop new technologies and operation concepts. In this research we have used them for acquiring aerial image sequences, to answer the challenge of detecting vessel using airborne images in an affordable way. Other sensing technologies, such as air- or space-borne synthetic Aperture radar systems, are available, but just for medium and large UAVs. These technologies are still an unrealistic option for small UAVs due to the heavy weight and excessive power requirements of the equipment. However, since most UAVs carry an onboard camera, it makes sense to use these small and lightweight sensors to acquire information about the sea surface. However, the big problem, and also the big challenge, is to create a non-complex, efficient, robust and real time computer vision algorithm capable of detecting a vessel from the images of a moving camera, onboard a UAV flying over the ocean.

This paper addresses the detection of vessels in the ocean, using unmanned air vehicles (UAVs) equipped with color cameras. This is a challenging problem since the camera is moving and the ocean surface presents non stationary and dynamic behaviors, due to the presence of waves, wind and illumination changes. These properties prevent the use of straightforward techniques such as background subtraction [3, 5]. In this problem, the background image moves and cannot be represented by a stationary model. In addition, airborne images often contain sun reflections and highly illuminated regions of sky (see Fig. 1) that turn vessel detection into a challenging problem.

The presence of a vessel in the ocean modifies the image texture and color. This suggests that texture and color features may provide useful cues to discriminate the vessel from the ocean surface. However, the visual difference between the vessel and the sun reflections or the sky is often small, meaning that a single feature (*e.g.*, pixel color) is not enough to detect vessels and we must rely on additional information.

The main contributions of this work are the following:

- it proposes a novel algorithm for vessel detection and verification in ocean surveillance applications, with low computational cost capable of running in the hardware embedded in small UAVs.

- it proposes a set of rules based on both spatial and temporal constraints to discriminate vessel blobs from image artefacts such as reflections and sky.

The rest of the paper is organized as follows. Section 2 presents a brief overview of related work. Section 3 describes the proposed system. Section 4 presents experimental results with real data and section 5 concludes the paper.

## 2. Related Work

The problem of vessel detection from airborne images has not been very deeply analysed in the literature. Most systems use air- or space- borne Synthetic Aperture Radar systems (SAR) [15, 1] or optical Satellite images [2, 16]. Current SAR systems are very expensive and cannot be carried by lightweight UAVs. The use of UAVs with automatic vessel detection capabilities would be of major importance for the surveillance of large maritime areas for the control of economic activities, illegal operations and environmental monitoring. Due to the constrained price, dimensions and payload of these systems, video cameras are the sensors of choice for remote detection.

Maybe due to the lack of aerial datasets on maritime scenarios, as well as the difficulties of operation of UAVs over the sea, very few works on vessel detection from UAVs have been reported. Some works tackle the vessel detection problem from cameras installed in the coast or on other ships [6, 7, 9], focusing on foreground/background separation, detection of objects above the horizon and saliency methods. Anyway, the majority of systems that operate visible cameras for target detection from aerial images concentrate on land surveillance [11, 13], for instance to detect targets like buildings, cars and people. Typical model free techniques include background modeling via image registration to detect independently moving objects [12]. In model based approaches, the objects to detect are typically characterized by their shape, color and texture content, or trained from examples using features like Haar [4] or histograms of oriented gradients [10].

One of the few works considering ship detection in UAV imagery is [8]. They use the assumption that the appearance of the ship does not change significantly among two frames taken a few seconds apart, whereas the sea surface changes enough to become almost uncorrelated. Then, using image correlation, they can detect the parts of the image that did not change, expectedly the ships and boats. In fact, using time consistency is one of the main contribution of our paper but instead of using only two images at a relatively large time separation, we use temporal windows with consecutive images at high frame rate to "decorrelate" the sea surface time varying texture from the vessel more stable image characteristics. Beyond temporal cues, we also use spatial cues to remove image regions not matching vessel's shape characteristics.

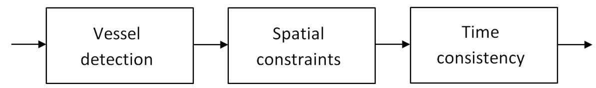


Figure 2. Block diagram of the vessel detection system.

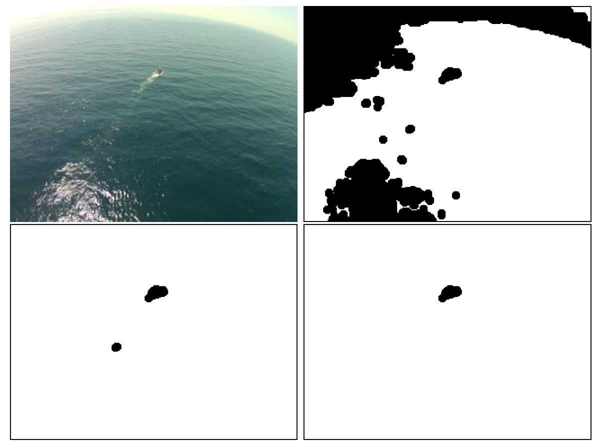

Figure 3. Vessel detection: original image (top-left), vessel detection mask (top-right), reflection & sky elimination using spatial constraints (bottom-left) and time consistency (bottom-right).

## 3. Proposed System

We wish to discriminate vessels from the ocean surface, discarding sun reflections and sky. This is not an easy task since sky and reflection pixels are characterized by high RGB values and the same happens with many vessel pixels. So, there is no way to distinguish the three types of pixels using color features only. In addition we wish to find fast algorithms that are able to operate in real-time inside a UAV. The strategy used to solve this problem is shown in Fig. 2.

First, we apply a vessel detection algorithm to obtain a set of tentative regions (blobs) which may be classified as vessels. Then, we estimate the sky & reflection blobs using spatial constraints and eliminate those blobs. Finally, we use additional frames acquired at different instants of time, in order to check for the time consistency of vessel estimates.

Figure 3 shows typical results obtained with this system: the top-left block shows the original image; the top-right block displays the output of the vessel detection algorithm; the bottom-left block exhibits the tentative vessel blobs after reflection and sky elimination and bottom-right block shows the final results combining multiple frames. A lot of spurious clutter is detected by the vessel detection algorithm. However, the clutter is efficiently eliminated by the other processing blocks that will be described in the sequel.

### 3.1. Vessel detection

Ideally, the vessel detection block should detect all the vessels in the image, keeping the number of false alarms as small as possible. This is a difficult task that was addressed by two alternative methods. The first method is a simple thresholding algorithm applied to the RGB components of each pixel. We assume that vessels have bright pixels with high RGB values. A pixel $\mathbf{x}$ is marked as active if the two following conditions are met

$$\max_c I_c(\mathbf{x}) > T_m \ , \quad \max_c I_c(\mathbf{x}) - \min_c I_c(\mathbf{x}) < T_d \ ,$$

where $I_c(\mathbf{x})$ denotes the amplitude of the $c-th$ color component of the input image $I$ at position $\mathbf{x}$, and $T_g, T_d$ are two thresholds. The output of this task is a binary image.

The second method follows a block based approach. The image is divided into non-overlapping blocks. Each block is characterized by a set of color features (color histograms, entropy, average color) and classified as vessel or background. The output of this method is also a binary mask as before, but all the pixels inside each block receive the same label.

Since computation time is an important issue and block based approaches presented some additional difficulties (the choice of the block size is difficult since it depends on the size of vessels in the image), we chose the pixel based approach in this paper. The results obtained with the thresholding algorithm are shown in Fig. 3 (top-right). The boat is detected as expected but a lot of clutter is detected as well.

### 3.2. Spatial validation

As we saw, sky and sun reflections blobs are detected by the thresholding operation and we need further processing to separate vessel blobs from the artifacts. In order to remove the sky and reflections, we use the binary mask $B$, computed by the vessel detection block. First, we dilate $B$, in order to link neighboring reflection regions and then apply a connected component algorithm. Each connected region $\mathcal{R}_i$ (blob) is characterized by $\text{Blob}_i = (\mathbf{x}_i, R_i, A_i)$, where $\mathbf{x}_i$ is the blob center and $R_i$ is the radius and $A_i$ the blob area.

We classify each blob as sky and reflection if it meets at least one of the following conditions: i) the blob area is larger than a threshold; ii) the blob touches the image boundary or iii) the blob is not isolated. The first condition selects regions larger than the maximum vessel area. The second condition selects regions touching the image boundary, typically sky. These two conditions eliminate many artefacts but they are not enough to eliminate the sun reflections since they often spread into large set of small regions. Therefore we add an additional criterion to eliminate such reflections by assuming that the true target is isolated in the ocean and no other region should be detected in a neighborhood of radius, $R$.

The first and third conditions require the use of thresholds that are empirically chosen. The proposed algorithm is summarized in Table 1.

**Spatial Validation Algorithm**

*Input:* binary mask $B$

*Parameters:* $A_{max}, R$

Dilate $B$ and determine the blob parameters $\text{Blob}_i = (\mathbf{x}_i, R_i, A_i)$.

For each blob $Blob_i$,

- 1st test: check if $A_i < A_{max}$;
- 2nd test: check if does not touch image boundary;
- 3rd test: check if it is isolated i.e., the distance to the nearest blob is larger than $R$.
- accept blob if all conditions are valid;

The new binary mask $B'$ is the union of the accepted blobs.

Table 1. Spatial validation algorithm

**Time Validation Algorithm**

*Input:* $\text{Blob}_i(t) = (\mathbf{x}(t)_i, R(t)_i, A(t)_i)$.

*Parameters:* $D, D^*$

For each blob $Blob_i(t)$

- check if it persists in time, i.e., count the number of times a region was detected with approximately the same area and at the same location, in $D$ past frames (time horizon). The number of occurrences, $O_i$, is compared with a threshold $D^*$. Accept $Blob_i(t)$ if $O_i \geq D^*$.

Table 2. Time consistency

### 3.3. Time consistency

The active regions detected in each frame still contain many false alarms that are not acceptable in a routine operation of the system. Some of these regions are located at random positions and they may be easily eliminated by comparing the results obtained at consecutive frames. This will be done by using a buffer of D past frames and count how many times the active region is detected in this set of frames. Only the persistent blobs that are detected multiple

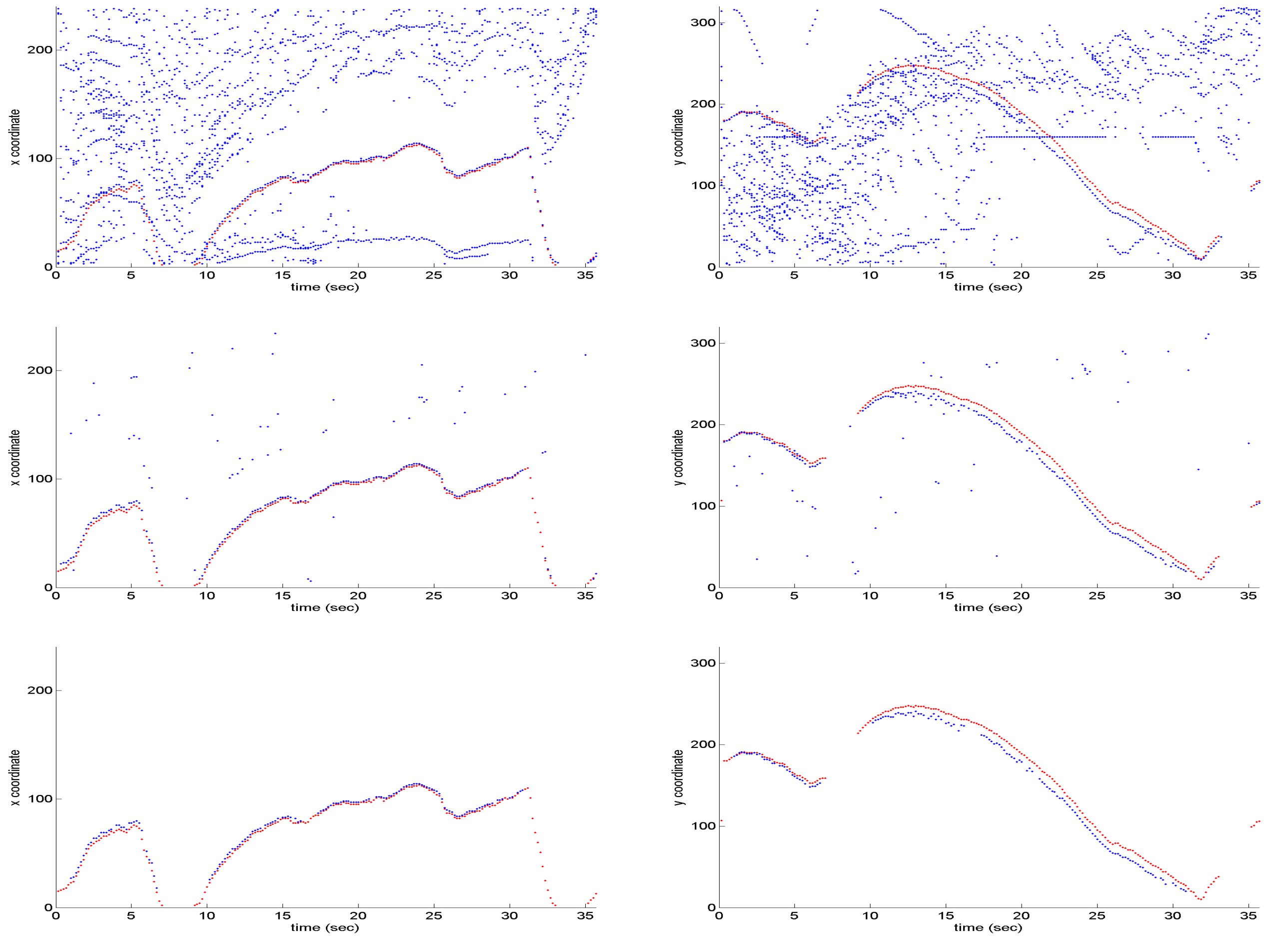


Figure 4. Center of the Detected blobs: original (top), with spatial constraints (middle) and after time consistency (bottom)). Ground truth blobs are displayed as red points and detected blobs as blue points. Data relative to sequence 2.

times are considered as valid candidates. This operation involves the ability to match corresponding blobs detected at different times. Several criteria can be used (e.g., see [14]).

The time consistency algorithm based on a buffer of D frames is summarized in Table 2.

Figure 4 shows the xy coordinates of all blobs detected in a video sequence and the output of the two validation steps. The spatial constraints remove most of the outliers and the time validation eliminates the rest. We conclude that a reliable detection of the vessel is therefore possible. It should be emphasized that most applications don't require vessel tracking along multiple frames but only a reliable detection of the vessel in one of the frames. The number of false alarms should of course be kept as small as possible.

## 4. Experimental Results

The proposed algorithm was tested using three video sequences, acquired by a UAV, flying at an altitude of 50 meters, equipped with a $1/2.3''$ format color camera and a wide angle lens. All of these sequences display a vessel with a length of 27 meters and a maximum width of 6 meters, and two of them show sun reflections and sky (see Table 3).

The images were acquired at a frame rate of 30 fps and have a spatial resolution of $480 \times 640$ pixels. Since the vessels move slowly in the image, we process every 5 frames, discarding the others. In addition, the images are subsampled with a factor $2:1$ to reduce the computation time. The proposed algorithm performs in real time in a standard PC, despite the fact that it was programmed in Matlab.

The following values were adopted in all the experiments: maximum vessel area, $A_{max} = 6000$ pixels, smallest distance between blobs, $R = 10$ pixels, time horizon, $D = 5$ subsampled frames, and threshold, $D^* = 4$ subsampled frames. These values were chosen by trial and error and were not changed during the experiments.

Figure 5 shows images extracted from three video sequences (left) and the output of the detection system (right),

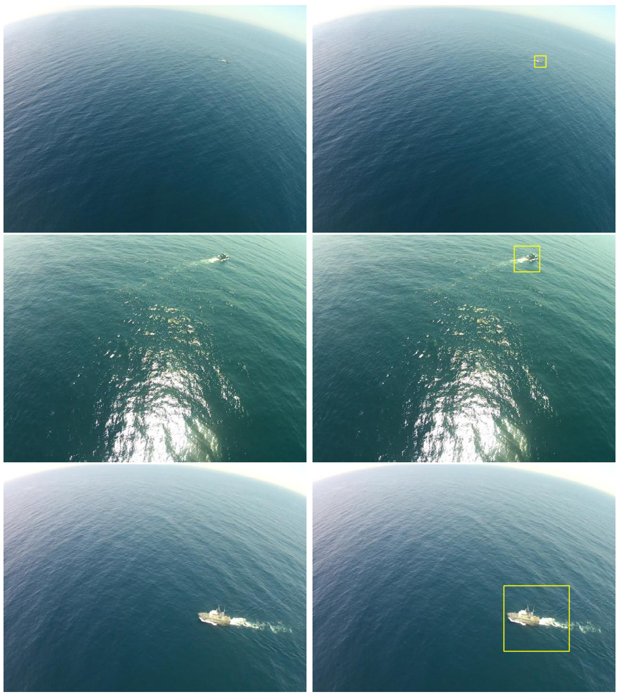

Figure 5. Vessel detection: images extracted from three test sequences (left) and the output of the detection algorithm (right).

| | No frames | Target | Sky | Refl. |
|---|---|---|---|---|
| Seq. 1 | 505 | Small | Yes | No |
| Seq. 2 | 1070 | Medium | Yes | Yes |
| Seq. 3 | 1400 | Large | Yes | Yes |

Table 3. Video data

showing the ability of the system to cope with vessels of different sizes and large amounts of reflections and sky.

To assess the algorithm, the target position and size (bounding box) was manually annotated for all the test images. Then we compute recall and precision

$$\text{recall} = \frac{TP}{TP + FN} \quad , \tag{1}$$

$$\text{precision} = \frac{TP}{TP + FP} \quad , \tag{2}$$

where TP (true positives) denotes the number of detected vessels, FN (false negatives) the number of non-detected vessels and FP (false positives) the number of detected non-vessels, aka false alarms. We consider the boat as detected if the overlap between the detected bounding box and the bounding box of the true vessel together with its trail is larger than 30%.

Table 4 shows the recall and precision achieved by the system after each of the main processing blocks. The first block detects most of the vessel in most of the frames (95.6%). However, the precision is very low (11.8%) *i.e.,* there is a very high number of false alarms that makes this

| Algorithm | Recall | Precision |
|---|---|---|
| Initial vessel detection | 95.6 | 11.8 |
| Spatial constraints | 88.7 | 77.5 |
| Time consistency | 74.6 | 99.7 |

Table 4. Performance of artifact elimination algorithms (statistics over all sequences).

| Sequence | Recall | Precision |
|---|---|---|
| Seq. 1 | 69.3 | 100.0 |
| Seq. 2 | 74.7 | 100.0 |
| Seq. 3 | 76.7 | 99.4 |

Table 5. Performance of the overall system.

a useless output. Fortunately, most of the false alarms are removed in the two following steps using spatial constrains and time consistency. Both steps play a major role in outlier elimination, especially the spatial constraints.

Table 5 shows the performance (recall vs. precision) of the proposed system for each of the three video sequences showing that the vessel is detected in most of the frames with a low false alarm probability. The final output can be considered as very robust since we do not need to detect the vessels in all the frames, we need to detect the vessels in all video sequences without false alarms and this can be easily achieved without errors by performing a majority voting along multiple frames.

## 5. Conclusion

This paper proposes a simple and efficient algorithm for the detection of vessels in the sea, using color images acquired by an unmanned aerial vehicle. Vessel detectors typically produce a high number of false alarms caused by sun reflections, sky and other dynamical artefacts. These false alarms cannot be avoided at the pixel level since the statistical properties of vessel pixels are often identical to the statistical properties of the reflections and sky.

To overcome these difficulties, false alarms are eliminated by using spatial and temporal constraints based on simple heuristic rules. Geometrical properties of the detected blobs are used to discard artifacts as well as time consistency. The proposed algorithm exhibits a remarkable robustness in the tests performed with real data annotated by an expert. Furthermore, it is computational efficient and runs in real-time in the embedded UAV hardware.

## Acknowledgement

This work was partially supported by projects SEAGULL (QREN SI IDT 34063) and FCT PEst-OE/EEI/LA0009/2013. The images were kindly provided by the Força Aérea Portuguesa. The authors would like to thank project PITVANT funded by Portuguese Ministry of National Defence for providing the UAVs, and Miguel

Griné from Instituto Superior Técnico for the ground truth annotations.

## References

[1] A. Arnold-Bos, A. Khenchaf, and A. Martin. An evaluation of current ship wake detection algorithms in sar images. In *Marine Environment Characterization, SeaTechWeek '06*, 2006.

[2] F. Bi, B. Zhu, L. Gao, and M. Bian. A visual search inspired computational model for ship detection in optical satellite images. *IEEE Geoscience and Remote Sensing Letters*, 9(4):749–753, 2012.

[3] T. Bouwmans. Recent advanced statistical background modeling for foreground detection-a systematic survey. *Recent Patents on Computer Science*, 4(3):147–176, 2011.

[4] T. Breckon, S. Barnes, M. Eichner, and K. Wahren. Autonomous real-time vehicle detection from a medium-level uav. In *International Conference on Unmanned Air Vehicle Systems*, 2009.

[5] M. Dikmen and T. S. Huang. Robust estimation of foreground in surveillance videos by sparse error estimation. In *IEEE International Conference on Pattern Recognition*, 2008.

[6] S. Fefilatyev and D. Goldgof. Detection and tracking of marine vehicles in video. In *International Conference on Pattern Recognition (ICPR)*, 2008.

[7] K. Gupta, D. Aha, R. Hartley, and P. Moore. Adaptive maritime video surveillance. In *SPIE-09 Visual Analytics for homeland security*, 2009.

[8] A. Kadyrov, H. Yu, and H. Liu. Ship detection and segmentation using image correlation. In *IEEE Intl. Conf. on Systems, Man and Cybernetics*, pages 2866–2871, 2013.

[9] W. Kruger and Z. Orlov. Robust layer-based boat detection and multi-target-tracking in maritime environments. In *Waterside Security Conference*, 2010.

[10] O. Oreifej, R. Mehran, and M. Shah. Human identity recognition in aerial images. In *Computer Vision and Pattern Recognition (CVPR)*, 2010.

[11] V. Ramakrishnan, A. Prabhavathy, and J. Devishree. A survey on vehicle detection techniques in aerial surveillance. *International Journal of Computer Applications*, 55(18):43–47, 2012.

[12] V. Reilly, H. Idrees, and M. Shah. Detection and tracking of large number of targets in wide area surveillance. In *European Computer Vision Conference (ECCV)*, pages 186–199, 2010.

[13] P. Robertson and J. Brady. Adaptive image analysis for aerial surveillance. *IEEE Intelligent Systems and their Applications*, 14(3):30–36, 1999.

[14] C. J. Veenman, M. J. T. Reinders, and E. Backer. Resolving motion correspondence for densely moving points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:54–72, 2001.

[15] W. XiaoLong and C. CuiXia. An automatic ship detection method based on local gray-level gathering characteristics in sar imagery. *Electronic Letters on Computer Vision and Image Analysis*, 12(1):33–41, 2013.

[16] C. Zhu, H. Zhou, R. Wang, and J. Guo. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 48(9):3446–3456, 2010.