Detectability of Objects at the Sea Surface in Visible Light and Thermal Camera Images

Christopher Dahlin Rodin
Maritime Robotics A/S &
Centre for Autonomous Marine Operations and Systems
Department of Engineering Cybernetics
Norwegian University of Science and Technology
Trondheim, Norway
christopher.rodin@ntnu.no

Tor Arne Johansen
Centre for Autonomous Marine Operations and Systems
Department of Engineering Cybernetics
Norwegian University of Science and Technology
Trondheim, Norway
tor.arne.johansen@ntnu.no

***Abstract*—In a number of ocean surveillance and remote sensing applications, visible light and thermal cameras are used to detect and identify objects at the sea surface. Knowing beforehand what the camera can detect or not can be important, yet highly difficult to determine. Optical models such as Modulation Transfer Functions can help in evaluating a camera system, but requires a deeper knowledge in optics, and detailed specifications of each component. The models also does not handle noise coming from the scene background, which in many cases is the major limiting factor of detectability. In this paper, we evaluate the results of an edge detection algorithm on images from two commercial off-the-shelf camera system – one visual light and one thermal. We then draw conclusions on the detectability of objects which commonly needs to be detected at the sea surface.**

I. Introduction

In ocean surveillance and remote sensing applications, the first step in various computer vision algorithms, is to detect the object in the camera image. Practical examples include sea ice detection and monitoring for situational awareness and environmental research [1], and detecting marine vessels and people in search and rescue missions [2] [3].

Two common sensors used in maritime surveillance and remote sensing missions are visible light and thermal cameras. Visible light cameras commonly provide a high spatial and temporal resolution image sequence of the scene in the visible light spectrum, which makes it suitable to detect very small objects, and objects slightly below the sea surface since the visible light penetrates water relatively well [4]. Thermal cameras commonly provide a lower spatial and temporal resolution image sequence in the thermal spectrum. This makes thermal cameras suitable for detecting objects with a different thermal footprint than the sea surface, in particular during low visible light conditions. Since the sea surface radiance is more homogeneous in the thermal spectrum than the visible light spectrum, the segmentation of objects at the sea surface is often also simpler in thermal camera images than visible light camera images.

Determining beforehand what a camera system can detect can prove difficult. Johnson's criteria [5] is still commonly used as the theoretical basis for object detectability based on the number of line pairs (or pixels) that an object represents in a camera image. Borghgraef et al. [6] evaluated different methods for detecting objects at the sea surface, where a high detection accuracy can be obtained if the background is uniform. This can also be seen in experimental results in e.g. [2] and [7]. Some difficulties mentioned, however, are highly dynamic backgrounds and a large camera to object angle, which causes objects to partly disappear behind ocean waves. It also makes it more difficult to detect flat objects, as the area of the object projection on the image sensor decreases with an increased angle. The work by Borghgraef et al. uses a stationary camera at a low altitude, which poses different challenges than a moving airborne camera. Bloisi et al. [8] similarly uses a stationary Pan-Tilt-Zoom visual light camera in order to evaluate the detectability. The experimental data contain camera images during different lighting conditions and camera angles, however these are not evaluated in regards to the detectability of the objects.

In this paper we aim to evaluate the detectability of common objects at the sea surface as they appear in images captured with two commercial off-the-shelf cameras – one visual light and one thermal – in order to evaluate how well objects can be detected compared to the theoretical geometric limit. Special emphasis is put on how the detectability varies with the distance and angle between the camera and objects. The detectability metric is based on the performance of an edge detection algorithm, which commonly forms the basis for more advanced computer vision algorithms [9]. Although a number of specialized edge detection algorithms have successfully been used to detect objects at the sea surface, e.g. by Can et al. [10] and Zhang and Skjetne [11], this paper uses a Sobel filter for edge detection due to its general and common usage.

II. Methodology

The aim of the methodology is to find how well objects can be detected in the camera images. Edge detection accuracy is chosen as the detectability metric due to that many higher level algorithms depend on the accuracy of the edge detection, such as object recognition and image segmentation [9]. This section first describes three general initial steps in edge detection algorithms – converting the image to grayscale, reducing the noise, and calculating the gradient of the image. These steps are performed in e.g. the Canny edge detection algorithm, which remains one of the most commonly used algorithms for edge detection [12]. Finally the edge detection algorithm, the method for determining the position of the objects in the camera images, the camera to object distance and angle

calculations, the detectability metric, and the edge detection algorithm parameter tuning are explained.

A. Grayscale Conversion

The images from the visible light camera are in RGB (Red-Green-Blue) format, i.e. each pixel contains intensity data for the red, green, and blue spectra separately. Since the subsequent methods work with one intensity value per pixel, the images first need to be converted to grayscale. Four different methods are considered for grayscale conversion: an average of the three color spectra, only using the red spectrum, only using the blue spectrum, and only using the green spectrum respectively. Due to the sea surface generally containing higher intensities in the green and blue spectra than the red, the choice of method can affect the result of the edge detection algorithm. The different methods are considered in the parameter tuning for the visual light camera dataset.

The thermal camera images contain only one intensity value per pixel, hence no grayscale conversion is required.

B. Noise Reduction

Noise can come from different sources: the image acquisition process (e.g. sensor noise), or from the scene background. At the sea, the water surface is a major source of noise, and can make it difficult to detect small objects. A common method for reducing noise is to apply a low-pass filter on the data. For images, a common low-pass filtering method is to convolve the image with a Gaussian kernel, which is a discrete approximation of a two-dimensional Gaussian function [13]. The two-dimensional Gaussian function is described in equation (1), and an example of a Gaussian kernel of size 3x3 is shown in equation (2).

$$g(x, y, \sigma) = \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

$$g = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \tag{2}$$

The parameters used in Gaussian smoothing are σ, which determines the width and height of the peak in the Gaussian function, and the size of the Gaussian kernel, s. A larger σ and size will smoothen the image more, resulting in less noise, but also less defined edges.

C. Edge Detection

In order to find the amplitude of the edges in an image, the gradient of the image is calculated. The method commonly used in e.g. the Canny edge detection algorithm is the Sobel operator [12]. Other methods exists, such as Robert's cross and the Prewitt operator, but they are generally outperformed by the Sobel operator [12]. Thus only the Sobel operator is considered in this paper.

The Sobel operator consists of two kernels, one for the horizontal edges, s_x, and one for vertical edges, s_y. An s_x of size 3x3 is shown in equation (3), and s_y is the transpose of s_x, see equation (4). Larger Sobel kernels, commonly 5x5 and 7x7, can also be used.

$$s_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{3}$$

$$s_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{4}$$

The kernels are convolved with the image in order to calculate an approximation of the image gradient in each direction, and finally the complete image gradient is calculated according to

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{5}$$

where $G_x(x, y)$ and $G_y(x, y)$ are the results of convolving s_x and s_y with the image, respectively.

The parameter used in edge detection is the size of the Sobel kernel, s. Three different sizes are considered: 3x3, 5x5, and 7x7.

D. Locating Objects in the Images

In order to find the location of the objects in the images, their positions were semi-automatically determined. The procedure varies between visual light images and thermal images due to the different noise characteristics – the visual light camera images are taken at an angle and at a high resolution resulting in a complex scene, making robust object detection difficult. In the method used, the location of the boundaries of the objects were manually selected by a human camera operator in every nth camera image, while the position of the boundaries in the intermediate images were found via linear interpolation. n was chosen to be 10, which gave a good compromise between an accurate position estimate and a low manual workload.

The thermal camera images are taken close to vertically at a lower resolution resulting in a homogeneous background, which is a good precondition for robust automatic object detection. A Gaussian Mixture Model [14] (GMM) was first used to find the position of the objects in the images. The GMM segments the image into background and foreground, and is suitable for thermal images at the sea since there are two distributions present in the incoming radiance: the radiance reflected from the sky and the heat emitted from the water. The GMM algorithm used was implemented as part of the Background Subtraction Library [15]. The algorithm detected most objects in the thermal image dataset, but the detections were manually corrected by a human camera operator to improve the detection quality.

E. Distance and Angle

In order to evaluate the detectability of the objects in relation to the distance and angle between the camera and the objects, the distance and angle need to be calculated. The distance refers to the three-dimensional Euclidean distance between the camera and the object at a given time. The angle refers to the angle between the vector pointing from the camera towards the Earth and the vector pointing from the camera towards the object. See figure 1 for an illustration of the angle.

In order to calculate the distance and the angle, the world position of the camera and objects need to be determined. The world position of the camera and boat and pallet are obtained from their onboard GNSS. In order to find the world position of the human and buoys, the attitude and position data of the camera are used together with the assumption that the object altitude is zero to project the object image position onto the world coordinate system using the pinhole camera model [16]. In order to improve the camera attitude estimate, the roll and pitch angles are corrected by solving Wahba's problem [17], where the position of the boat and pallet in the image coordinate system are aligned with their world positions projected onto the image coordinate system.

F. Geometric Limits

In order to find the geometric limit of whether an object is detectable or not, its apparent area (perpendicular to the camera sensor) is projected onto the camera sensor. The geometric limit for detecting an object is considered to be when an objects smallest apparent dimension (width or height) is represented by one pixel in the camera image. The following approximations are made regarding the dimensions of the objects:

- The boat has a height of 1.2 meters. Due to the complex geometry, this is a rough approximation.
- The human is 0.5 meters wide, and sticks out 0.25 meters from the water surface.
- The surface of the pallet is at the water surface.
- The buoys are floating on top of the water surface.

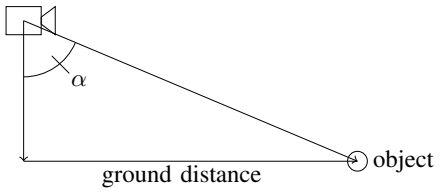

Fig. 1. The angle between the camera and object, α.

G. Error Metric and Evaluation

In order to quantitatively evaluate the detectability of objects, a metric needs to be chosen for how well the objects were detected in the camera images. The goal of the metric is to give a higher score when the difference between the object edges and background edges are larger.

The proposed method first generates a number of windows at random locations in the image. The windows are generated so to not contain any object, i.e. they represent the background. For the visual light camera images, the background windows also does not contain any part of the sky or land areas. An object is then considered detectable if its average edge value (the intensity of the image gradient) is larger than the average edge in a predefined fraction of the background windows.

H. Parameter Tuning

In order to obtain results which are not affected by poorly chosen parameters, the parameters of the edge detection algorithm are tuned using grid search. A random sample consisting of 20% of the images from the visual light camera and thermal camera datasets are selected to tune the parameters for each of the datasets. As the optimal parameters will differ between the visual light and thermal camera images due to their difference in resolution, scene, and noise, they are tuned independently.

A summary of the parameters and their ranges are shown in table I for the visual light camera, and table II for the thermal camera.

TABLE I. ALGORITHM PARAMETERS FOR THE VISUAL LIGHT CAMERA IMAGES.

Grayscale conversion	
Type	average, only red, only green, only blue
Blurring	
Gaussian kernel size	1x1 – 15x15
Gaussian σ	0.5 – 7.5
Edge detection	
Sobel kernel size	3x3 – 7x7

TABLE II. ALGORITHM PARAMETERS FOR THE THERMAL CAMERA IMAGES.

Blurring	
Gaussian kernel size	1x1 – 7x7
Gaussian σ	0.5 – 7.5
Edge detection	
Sobel kernel size	3x3 – 7x7

Each combination of the parameters in table I and II are tested on the visual light and thermal camera datasets, and the parameters producing the highest detectability rate for the worst case object type (boat, pallet, human, buoy) are chosen as the best parameters, which will be used in the final evaluation of detectability.

III. EXPERIMENTAL DATA

This section describes the datasets used in this paper. The dataset consists of images captured by a visible light and a thermal camera, as well as the position and orientation of the cameras. The cameras were attached to UAVs and flown at varying positions relative to the objects. The subsections describe the objects and information about the visible light and thermal cameras.

A. Objects

The objects placed at the water surface was a 26 foot boat, an EUR-pallet, a human wearing an immersion suit, and two red buoys with a diameter of 60 cm, which were chosen as common objects to detect in maritime missions such as search and rescue and seismic operations. The objects are shown in figure 2.

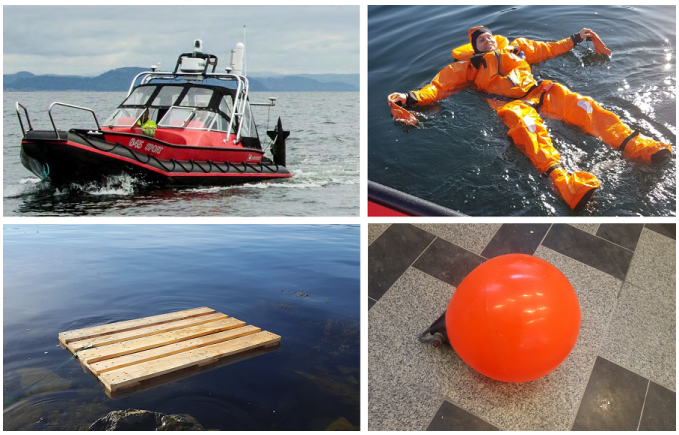

Fig. 2. The objects present in the scene. Upper left: boat, upper right: human in immersion suit, lower left: EUR-pallet, lower right: buoy.

The boat and pallet were equipped with GNSS (Global Navigation Satellite Systems), which means that their world position data is available.

B. Visible Light Camera Data

The visible light camera is a DJI FC350, which provides images at a 3840 px x 2160 px resolution at 25 Hz in 24 bit H.264 format. The lens has a focal length of 3.61 mm, providing an angle of view of 83.6° x 50.2°. A subset of the whole dataset is used for the detectability evaluation – 2000 images for when the UAV is moving towards the objects, and the objects transitions from not detectable to detectable. See figure 3 for a sample image from the dataset. It can be seen that the lighting conditions varies throughout the image – the upper left part is brighter than the lower right part. Different lighting conditions can result in under- or overexposure of the objects, which might decrease the detectability.

Fig. 3. A sample image from the experiment dataset captured by the visible light camera.

The position and orientation of the camera is available as metadata. This is used to calculate the distance and angle between the camera and object for each image, as well as determining the world position of some of the objects (see section II-E). Figure 4 shows the distances and angles between the camera and the objects. The camera altitude is kept steady at 63–65 meters throughout the dataset.

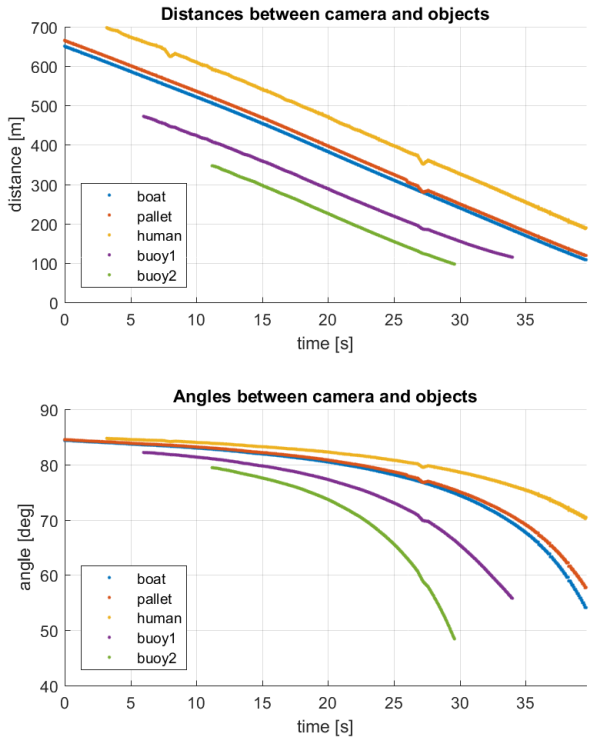

Fig. 4. The distances and angles between the visual light camera and the objects at the sea surface.

C. Thermal Camera Data

The thermal camera is a FLIR Tau2, which provides images at a 640 x 512 pixels resolution at 9 Hz in 16 bit raw format. The lens has a focal length of 19 mm, providing an angle of view of 32° x 26°. See figure 5 for a sample image from the dataset. A total of 1974 thermal camera images, containing at least one object each, are used to evaluate the detectability of the objects.

Position and orientation data of the camera is also available as metadata. The thermal images are taken close to vertically, i.e. with the camera pointing straight down. Because of this, the total number of detections are evaluated rather than the distance and angle between the camera and objects. The altitude of the camera varies between 235 and 250 meters during the flight.

IV. Experimental Results

The datasets were processed according to the methods described in section II. At first the parameters were tuned for the visual light and thermal camera datasets, and the best

Fig. 5. A sample image from the experiment dataset captured by the thermal camera. The boat is visible in the upper left corner, and a human or buoy is visible below the boat.

performing parameters were evaluated. The best performing parameters were then applied on the full data sets, and the results were evaluated in relation to the camera to object distance and angle.

A. Parameter Tuning – Visual Light Camera

The parameter were tuned using 100 background windows with a width and height of 200 pixels for each image. An object is considered detectable when it has an average intensity larger than 90% of the background windows.

The detection rate for each type of object for different parameters can be seen in figures 6–9.

The grayscale conversion parameter influence on detectability can be seen in figure 6. The method of converting the RGB image to grayscale has a noticeable effect on the detectability of humans and buoys. For detecting the human, using only the green channel produces the best results, while this yields the worst detectability of the buoys. Only using the red channel produces the best detectability for buoys, while it produces among the worst detectability for the human. Since both objects are closest to red, while the sea surface is closest to blue, it would be expected that only using the red channel would produce the best results. The poor detectability for the human when only using the red spectrum might be due to image compression artifacts, rendering the colors of the human closer to its surrounding.

Since the human has the lowest detectability of all objects, this will be the deciding factor, thus only using the green channel is set as the best parameter for grayscale conversion.

The noise reduction parameter influence on detectability can be seen in figure 7 and 8 for σ and size respectively. The noise reduction algorithm uses two parameters – σ and kernel size. An increase of either parameter reduces the noise more, but produces less defined edges. It can be seen that the detectability is greatly improved for all objects when applying minor blurring. For the larger objects, in particular the boat, it

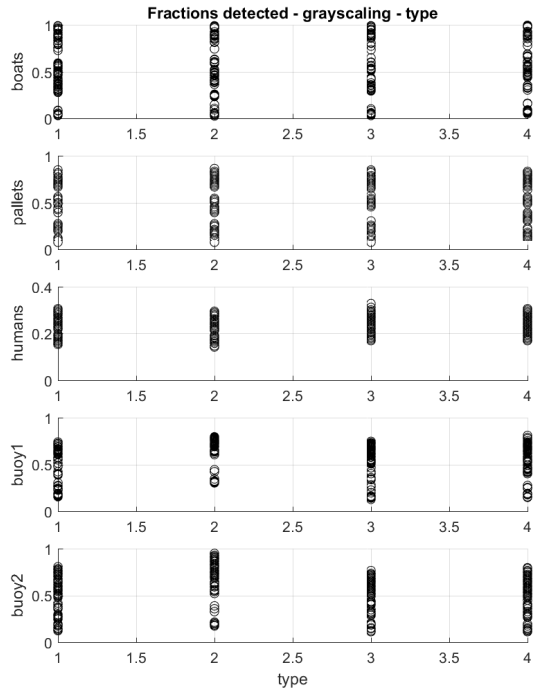

Fig. 6. The fraction of objects detected for different methods for converting the image to grayscale in the visual light camera image test dataset. 1: RGB average, 2: only red, 3: only green, 4: only blue.

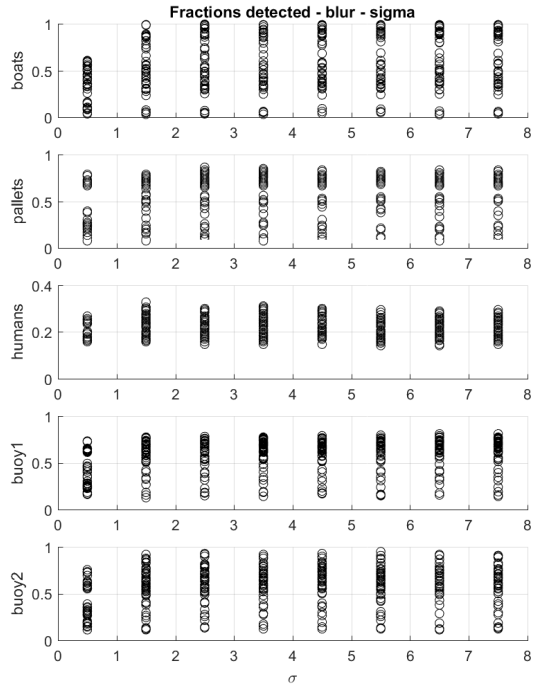

Fig. 7. The fraction of objects detected for different noise reduction σ in the visual light camera image test dataset.

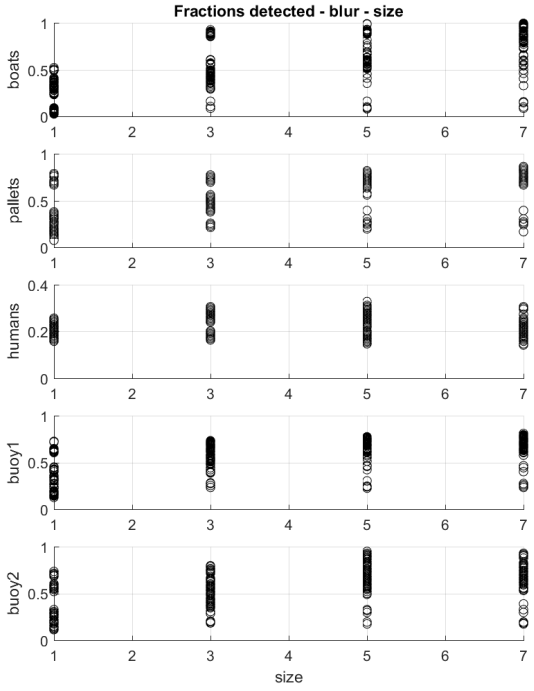

Fig. 8. The fraction of objects detected for different noise reduction kernel sizes in the visual light camera image test dataset.

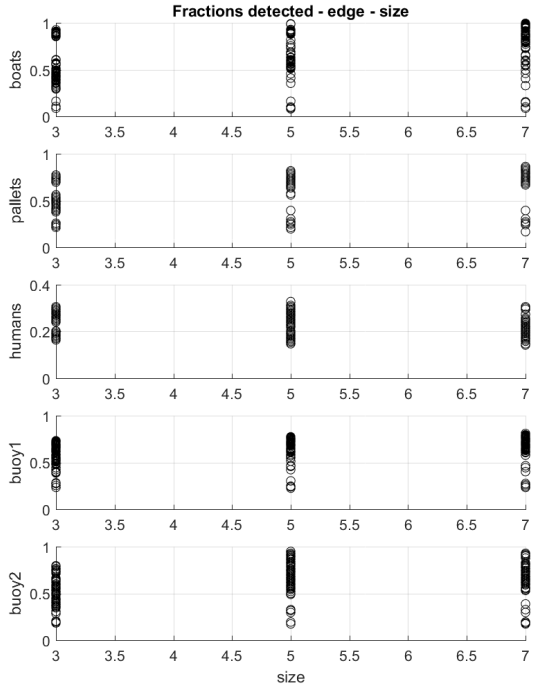

Fig. 9. The fraction of objects detected for different edge detection kernel sizes in the visual light camera image test dataset.

keeps increasing, while for the smaller objects the detectability decreases after a Gaussian kernel size of 5x5 px, and σ of 1.5 for the human and around 5.5 for the buoys.

Since the human has the lowest detectability of all objects, this will be the deciding factor, thus a σ of 1.5 and kernel size of 5x5 is chosen as the best parameters for noise reduction.

The edge detection parameter influence on detectability can be seen in figure 9. It can be seen that for all objects except the pallet, using a Sobel kernel size of 5x5 produces a better detectability. A kernel size of 7x7 seems to be able to increase the detectability of certain objects, but in general the 5x5 Sobel kernel produces the best results.

Since the human has the lowest detectability of all objects, this will be the deciding factor. The parameter set which produces the best results had a Sobel kernel size of 3x3, and is thus chosen as the best parameters for edge detection.

The parameter set which provided the best detectability for the visual light camera dataset is shown in table III.

A sample image before and after applying the edge detection algorithm using the best and worst parameter set is shown in figure 10. It can be seen that an edge detector reducing noise rather than emphasizing the edges produces better results.

TABLE III. THE PARAMETERS PRODUCING THE BEST RESULTS IN THE THERMAL TEST DATASET.

Grayscale conversion	
Type	Only green
Blurring	
Gaussian kernel size	5x5
Gaussian σ	1.5
Edge detection	
Sobel kernel size	3x3

B. Detectability – Visual Light Camera

The optimal parameters were used to run the algorithms on the full visual light camera dataset. The results are visualized in figures 11 – 15 as its detectability status (detectable by the algorithm, detectable by the operator, or not detectable) in graphs showing the distance between the camera and the objects, and the number of pixels in the smallest dimension when projected onto the camera image plane. The status ”detectable by the operator” means that the human camera operator was able to locate the object in the camera image (see section II-D). The status ”not detectable” is thus only available for the boat and the pallet since their world position is available from their GNSS.

It can be seen that the boat is detectable by both the operator and edge detection algorithm throughout the image dataset. The algorithm misses the boat in two images, which can be considered outliers. This is as expected, as the number of pixels representing the smallest dimension of the boat in the image barely dips below 4.

The pallet can be detected by the operator until its smallest dimension reaches around 1.5 pixels, while the algorithm can detect the pallet until it reaches 2.5 pixels.

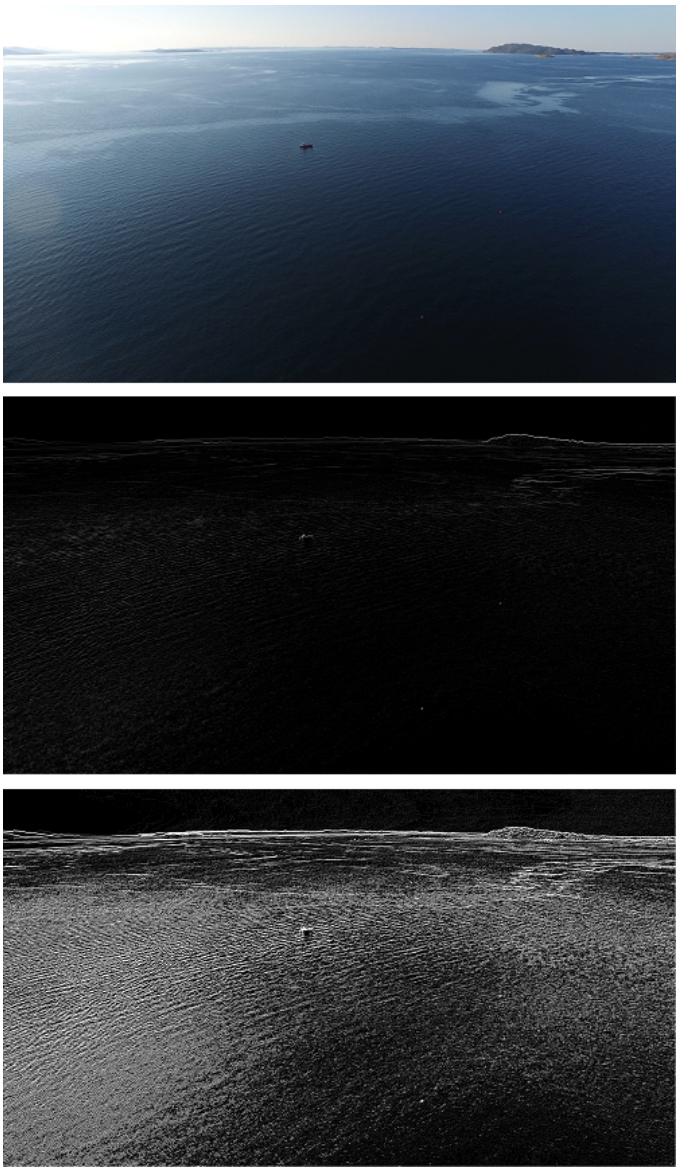

Fig. 10. A visual light camera image (top), and its edges using the best parameters (middle) and worst parameters (bottom).

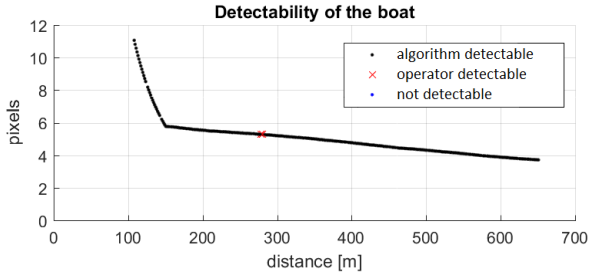

Fig. 11. The detectability of the boat compared to the distance between the camera and the number of pixels representing the smallest dimension of the boat.

The human remains very detectable by both the algorithm and operator until a camera to object distance of around 300 meters. At that distance, the human is represented by slightly above 1 pixel. The greater detectability of the human than the

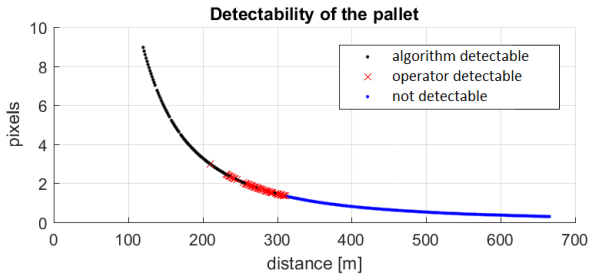

Fig. 12. The detectability of the pallet compared to the distance between the camera and the number of pixels representing the smallest dimension of the pallet.

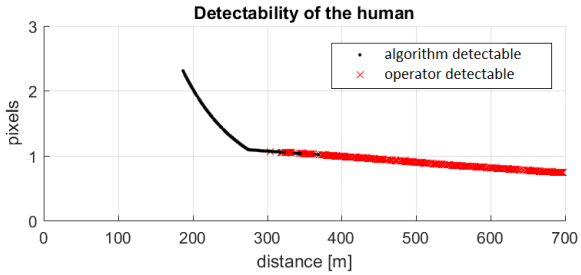

Fig. 13. The detectability of the human compared to the distance between the camera and the number of pixels representing the smallest dimension of the human.

pallet can be explained by that the human is at a brighter part of the sea surface, creating a larger contrast between the object and the background and thus a larger edge intensity.

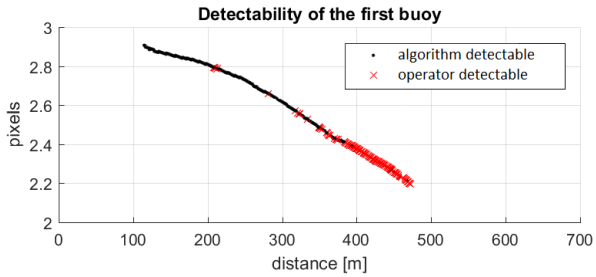

Fig. 14. The detectability of the first buoy compared to the distance between the camera and the number of pixels representing the smallest dimension of the first buoy.

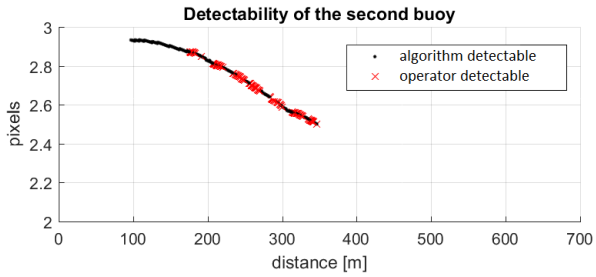

Fig. 15. The detectability of the second buoy compared to the distance between the camera and the number of pixels representing the smallest dimension of the second buoy.

The first buoy is more detectable then the second buoy, both for the algorithm and the operator. Looking the the images,

this is mainly caused by the lighting – the first buoy has less sun reflections, and thus has a deeper red color in the camera images. The more red color separates it better from the sunlight reflections in the sea surface. See figure 16. The buoys are sporadically not detectable by the algorithm, but when their size is close to 3 pixels, there are no missed detections.

Fig. 16. The first buoy (left) versus the second buoy (right), as they appear in a visual light camera image.

C. Parameter Tuning – Thermal Camera

The parameter were tuned using 100 background windows with a width and height of 30 pixels for each image. An object is considered detectable when it has an average intensity larger than 100% of the background windows.

The detection rate for each type of object for different parameters can be seen in figure 17–19.

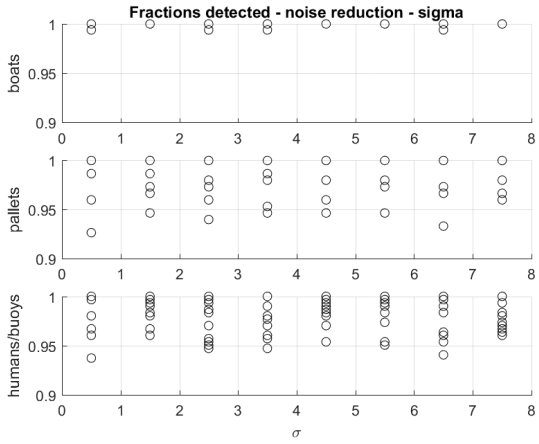

Fig. 17. The fraction of objects detected for different noise reduction σ.

It can be seen that the Sobel kernel size has the largest effect on the detectability, with a larger kernel resulting in more objects detected. The effect on boats, which are larger, is minor, while the detectability of humans/buoys and pallets are greatly increased, and often reach 100% using a Sobel kernel size of 7x7.

The parameter set which provided the best detectability for the thermal camera dataset is shown in table IV.

A sample image before and after applying the edge detection algorithm using the best and worst parameter set is shown in figure 20. It can be seen that an edge detector emphasizing edge detection over noise reduction produces better results.

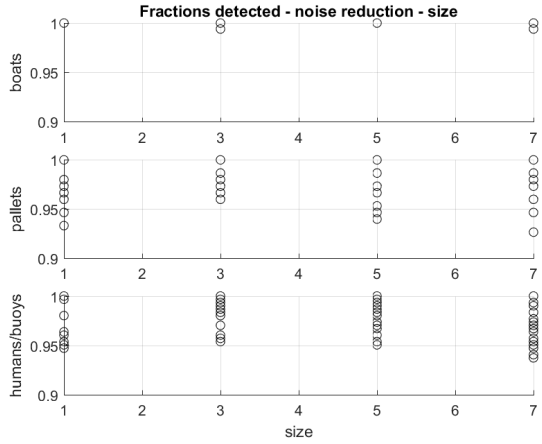

Fig. 18. The fraction of objects detected for different noise reduction kernel sizes.

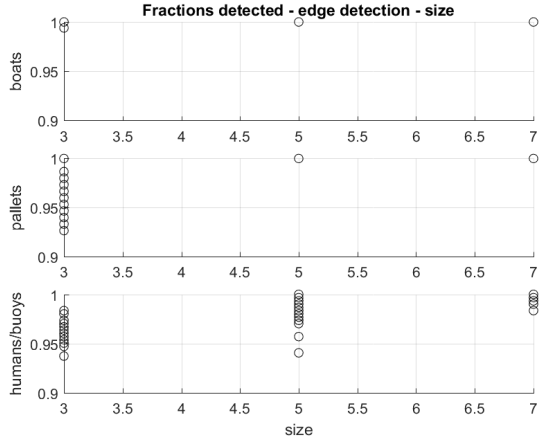

Fig. 19. The fraction of objects detected for different Sobel kernel sizes.

TABLE IV. THE PARAMETERS PRODUCING THE BEST RESULTS IN THE THERMAL TEST DATASET.

Blurring	
Gaussian kernel size	1x1
Gaussian σ	1
Edge detection	
Sobel kernel size	5x5

D. Detectability – Thermal Camera

Using the best parameter set from the parameter tuning, the algorithm is applied for the full thermal camera dataset. The number of detections are summarized in tableV. In order to estimate how well the objects can be detected geometrically, the pixels representing the objects in the camera images are shown in table VI.

It can be seen that all boats and pallets in the dataset were detected. 6 humans/buoys were not detected, representing 0.4% of all objects of the same class. As can be seen in table VI, most objects are above 2 pixels in their smallest dimension, which is the general limit for detectability according to Johnson's criteria [5]. An exception is the human, which can fall

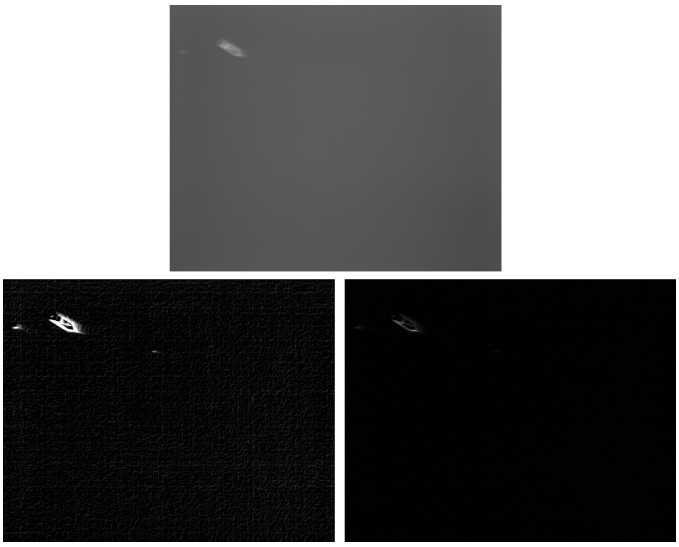

Fig. 20. A thermal image (top), and its edges using the best parameters (bottom left) and worst parameters (bottom right).

TABLE V. THE NUMBER OF OBJECTS CONSIDERED DETECTED IN THE THERMAL CAMERA IMAGE DATASET.

Object	Detected	Total	Fraction
Boats	869	869	100%
Pallets	787	787	100%
Humans/buoys	1504	1510	99.6%

TABLE VI. THE SIZE OF THE OBJECTS IN PIXELS, ASSUMING A VERTICAL PHOTO AT AN ALTITUDE OF 240 METERS.

Object	Size [px]
Boat	38 x 12
Pallet	5.7 x 3.8
Human	2.4 x 1.4 – 2.4 x 8.1
Buoy	2.9 x 2.9

below 2 pixels when standing upright (only the head and shoulders are above the water surface).

V. CONCLUSIONS

In this paper we have evaluated the detectability of objects that commonly needs to be detected in maritime surveillance missions. The detectability metric was the average edge intensity of an object compared to 100 randomly located background windows in the same image, after applying grayscale conversion, noise reduction, and edge detection algorithms on the images. In order to avoid poorly chosen algorithm parameters, parameter tuning was performed using 20% of the full datasets. The detectability was then evaluated in relation to the number of pixels representing the smallest dimension of the objects when projected onto the camera image plane for the visual light camera images, and the total number of detections for the thermal camera images.

The results show that objects can be reliably detected when the smallest dimension is around 3 pixels or larger. At lower pixel counts, the noise from the sea surface, e.g. waves, is too large compared to the contrast between the sea surface and the objects. The lighting is also shown to have a certain affect on the detectability, where two buoys of the same size and color have a different detectability due to their different lighting conditions.

Further work should be done to generalize detectability for different algorithms further, e.g. evaluating other noise reduction and edge detection algorithms, as well as other object detection algorithms. Due to their good object detection and classification performance, the results of Convolutional Neural Networks could be used to determine detectability of the objects. In order to develop a more sophisticated model of object detectability at the ocean surface, more advanced camera sensor and optics models, as well as water surface models, could be used. E.g. could the modulation transfer function (MTF) be used to model how well the camera can detect the objects. The lighting conditions for the different objects should also be further investigated, since it was shown to be an important factor of detectability when the objects were represented by a very small number of pixels in the camera images.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 642153. The information presented in this paper reflects only the author's view, and the Agency is not responsible for any use that may be made of the information it contains. The work was also supported by the Research Council of Norway through the Centre for Autonomous Marine Operations and Systems (NTNU-AMOS), grant number 223254. The authors would like to express their gratitude to the people from Maritime Robotics AS and NTNU UAV Lab involved in the experiment performed for this paper.

REFERENCES

[1] C. D. Rodin and T. A. Johansen, "Accuracy of sea ice floe size observation from an aerial camera at slant angles," in *2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)*, Oct 2017.

[2] F. S. Leira, T. A. Johansen, and T. I. Fossen, "Automatic detection, classification and tracking of objects in the ocean surface from uavs using a thermal camera," in *IEEE Aerospace Conference*, 2015.

[3] T. Sumimoto, K. Kuramoto, S. Okada, H. Miyauchi, M. Imade, H. Yamamoto, and T. Kunishi, "Machine vision for detection of the rescue target in the marine casualty," in *Industrial Electronics, Control and Instrumentation, IECON '94., 20th International Conference on*, vol. 2, 1994, pp. 723–726.

[4] H. Egna and C. Boyd, *Dynamics of Pond Aquaculture*. Taylor & Francis, 1997.

[5] F. A. Rosell and R. H. Willson, "Basics of detection, recognition and identification in electro-optical formed imagery," in *Proceedings of SPIE*, vol. 33, 1973.

[6] A. Borghgraef, O. Barnich, F. Lapierre, M. Van Droogenbroeck, W. Philips, and M. Acheroy, "An evaluation of pixel-based methods for the detection of floating objects on the sea surface," *EURASIP Journal on Advances in Signal Processing*, no. 1, Feb 2010.

[7] J. Yong-Xin, Y. Qun-Zhe, S. Cheng-Yong, Y. Chang-Qin, and Y. Xi-Yong, "Study on detection method of small object on sea based on kernel-mrf foreground segmentation model," in *International Conference on Computer Science and Service System*, 2012, pp. 1869–1872.

[8] D. Bloisi, L. Iocchi, M. Fiorini, and G. Graziano, "Automatic maritime surveillance with visual target detection," in *Proceedings of the International Defense and Homeland Security Simulation Workshop (DHSS)*, 2011, pp. 141–145.

[9] K. Plataniotis and A. Venetsanopoulos, *Color Image Processing and Applications*, ser. Digital Signal Processing. Springer Berlin Heidelberg, 2000.

[10] T. Can, A. O. Karali, and T. Aytaç, "Detection and tracking of sea-surface targets in infrared and visual band videos using the bag-of-features technique with scale-invariant feature transform," *Applied Optics*, vol. 50, no. 33, pp. 6302–6312, Nov 2011.

[11] Q. Zhang and R. Skjetne, "Image processing for identification of sea-ice floes and the floe size distributions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2913–2924, May 2015.

[12] M. Nixon and A. S. Aguado, *Feature Extraction & Image Processing for Computer Vision*, 3rd ed. Academic Press, 2012.

[13] R. Bourne, *Fundamentals of Digital Imaging in Medicine*. Springer London, 2010.

[14] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.

[15] A. Sobral, "BGSLibrary: An opencv c++ background subtraction library," in *IX Workshop de Visão Computacional (WVC'2013)*, Rio de Janeiro, Brazil, Jun 2013. [Online]. Available: https://github.com/andrewssobral/bgslibrary

[16] S. J. D. Prince, *Computer Vision: Models, Learning, and Inference*, 1st ed. New York, NY, USA: Cambridge University Press, 2012.

[17] F. L. Markley, "Equivalence of Two Solutions of Wahba's Problem," *Journal of the Astronautical Sciences*, vol. 60, pp. 303–312, Dec. 2013.