

Object detection with semantic segmentation of LWIR images in maritime environment

Shailesh Nirgudkar¹ and Paul Robinette²

Abstract—This paper describes early work on object detection in maritime environment using long wave infrared (LWIR) images. The focus is on objects floating on water surface or are visible from the water surface. The detection is achieved by the semantic segmentation that divides the scene into three categories - object, water and sky. The LWIR images are less susceptible to water dynamics and lighting conditions. The results obtained indicate that this direction is promising for further research.

I. INTRODUCTION

The unmanned and autonomous surface vehicles (USVs/ASVs) have wide applications in the maritime environment as they reduce human efforts and cost in periodic and long haul tasks on the seas such as surveillance, water quality and environment monitoring, and transportation. This rise of surface vehicles (SVs) has highlighted the significance of improved maritime object detection and tracking mechanisms. This is because to navigate the water autonomously, the SVs must first detect objects in the path and work around the objects without collision. The object detection is done by employing various sensors. Historically, radar was used to detect objects. Then the community started experimenting with optical cameras. During this time, objects were detected by using classical image processing algorithms. However, for past few years, the detection and tracking with the aid of optical cameras has gained significant traction. This is because of the excellent performance of

deep learning algorithms in image analysis [1] coupled with availability of ultra-fast hardware namely graphics processing unit (GPU). However, this environment poses a different set of challenges as against the land, namely - dynamic water background, visibility impact due to conditions such as rain, fog or darkness, minimal availability of static cues, camera that's always subjected to wave motion, reflections and glitter caused in water due to sunlight, detection of objects at horizon etc. As a result, the existing algorithms need to be tailored and new approaches considered for the maritime environment. A lot of the maritime research [2], [3], [4], [5], [6] thus far has tried to adapt the algorithms that are typically used in the vehicle or face detection domain. These algorithms typically make use of electro-optical images and some of these approaches have given remarkable results. However, research focused on thermal images has received relatively less attention. Although thermal images, more specifically, long wave infrared (LWIR) camera images (wavelength range: 8-15 μm) contain relatively less information than color images, they come with their own set of advantages as well. Infrared (IR) images fair much better than optical images in poor lighting conditions when the lack of visible light naturally decreases image contrast. Reflections caused due to water and sun glitter are problems native to the maritime environment for which IR images do a better job than the traditional optical ones because IR images are less susceptible to water dynamics. Another major obstacle that is peculiar to the maritime environment is the sea-sky boundary. Determination of this boundary reduces the search space significantly. However, it is difficult to determine its precise location under various weather

¹Shailesh Nirgudkar is a graduate student in Department of Electrical and Computer Engineering, University of Massachusetts, Lowell, MA 01854, USA shailesh_nirgudkar@student.uml.edu

²Paul Robinette is with the Department of Electrical and Computer Engineering, University of Massachusetts, Lowell, MA 01854, USA paul_robinette@uml.edu

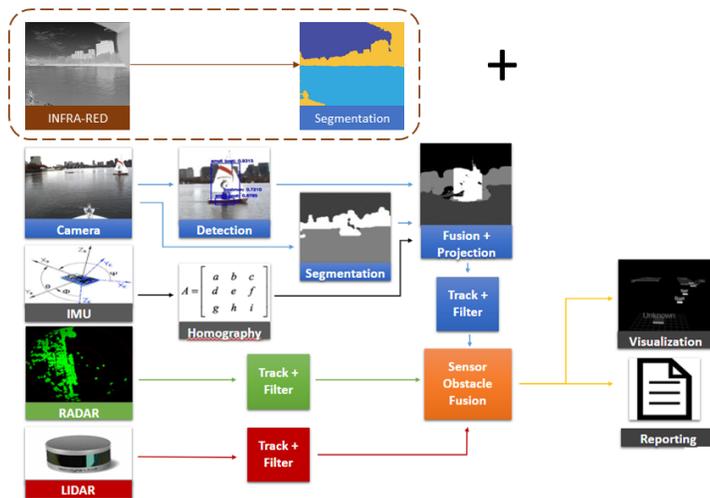


Fig. 1: Multi-sensor object tracking system, image based on [8]

and lighting conditions. It is a complex problem and thermal images may assist in providing more alternatives to tackle the same. Recently, Schöller et al. [7] has shown that deep learning algorithms can be applied to thermal images as well. Their work was related to predicting bounding boxes around the objects in the water. In this paper, we present our recent work on using deep learning algorithms on thermal images in the maritime environment. We use semantic segmentation for training and predicting the navigable surface. Our team has been working on fusing data from multiple sensors, Clunie et al. [8] has demonstrated an architecture combining radar, lidar and optical cameras to detect and track obstacles. Addition of an IR sensor will bring new capability to the architecture as shown in Fig. 1.

This paper has been organized as follows. Section II describes the hardware and image setup mechanism. Section III describes various experiments done on IR images. Section IV describes the early results and comparison with the corresponding equivalent optical images. We

conclude by identifying problems and the future steps.

II. SYSTEM SETUP

The system consists of a small autonomous boat, R/V Philos as shown in Fig. 2. It has 3 electro-optical (EO) cameras and 2 forward looking infrared automotive development kit (FLIR ADKTM) cameras. The FLIR cameras have resolution 640 x 512 pixels and have 75 degrees field of view (FOV). The output of the FLIR can be saved in the form of compressed 8 bit PNG image or 16 bit TIFF format. The videos were recorded in the Charles river near Cambridge, Massachusetts, USA in the months from July-October 2020 under various weather and traffic conditions.

III. EXPERIMENTS

A representative sample of 100 IR images were selected from MIT dataset [10]. These images were annotated manually using a semantic segmentation tool. Recently ‘water-obstacle separation and refinement’ (WaSR) [2] had produced excellent inference results on the

‘Marine Obstacle Detection Dataset’ (MODD-2) [3]. This dataset consisted of challenging optical images in marine environment. WaSR was trained on ‘Marine Semantic Segmentation Training Dataset’ (MaSTr1325) [11] image dataset. WaSR [2] divided the pixels into 3 categories - water, sky and obstacles. Because of its superior performance, we chose WaSR [2] architecture to train and to run inference on the IR images. The inference produces an image with a mask. In following segmented images, yellow indicates object, turquoise indicates water and dark blue indicates sky.

A. Inference without any retraining

a) WaSR [2] was trained on optical images in maritime environment so we decided to run the inference on our IR images by using WaSR [2] model as-is. Since the IR images consist of a single channel and the WaSR [2] model requires three channels corresponding to a color image, during pre-processing step, we converted the single channel gray-scale images to equivalent color images by copying the value in remaining two channels. The inference on sample IR images is shown in Table. I, second row.

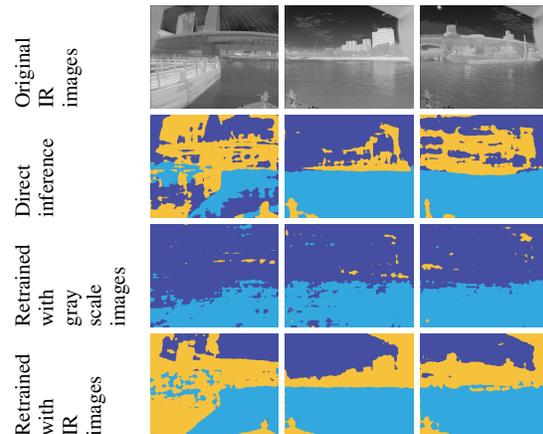
b) For comparison, we are also showing the results obtained by the same model on some colored images. As can be seen in Table II, second row, the segmentation results obtained on colored images are better than obtained on IR images.

c) Next, we converted these color images to gray scale images and again to color images. By doing this conversion, all the color information is removed, and they become similar to IR images.



Fig. 2: R/V Philos System [9]

TABLE I: Inference on IR images using WaSR



Since WaSR [2] model operates on color images, it requires three channels and hence gray scale images need to be converted to color images. The output segmentation masks obtained on these images are similar to b), still better than a). The inference is shown in Table. II, fourth row.

B. Retraining WaSR [2] with its own dataset converted to gray scale

We decided to retrain WaSR [2] on gray scale images because the IR images are gray

TABLE II: Inference on optical images using WaSR [2]

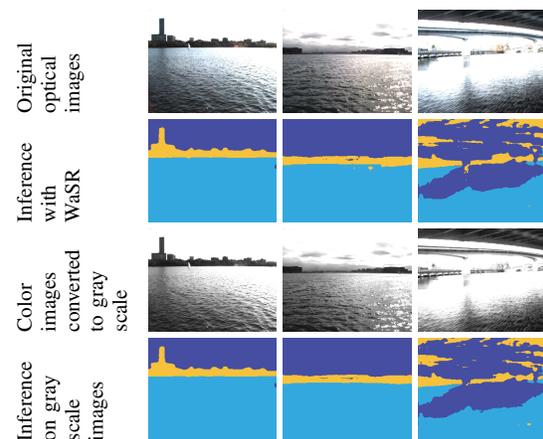


TABLE III: Configuration of experiments

Name	Number of images	Batch size	Number of iterations	Image resolution
WaSR [2] + training on MaSTR1325 [11] grayscale images	1325	3	10050	256 (w) x 192 (h)
WaSR [2] + training on IR images [10]	88	2	10050	320 (w) x 256 (h)

scale images. For this purpose, we converted the MaSTR1325 colored images to gray scale images and back to color images thereby removing all the color information. Then WaSR [2] was trained on these images and then inference was run on the IR images. The results of this experiment are as shown in Table. I, third row. Surprisingly, the results were not better than Table. I, second row and in fact, a little worse. Since the mean pixel value of IR image was 121 which was much lower than the mean pixel value of gray scaled MaSTR1325 images, 166, we hypothesized that it could be a contributing factor. However, increasing the mean pixel value of the IR images (brightening operation) as a pre-processing step in the inference stage did not change the outcome.

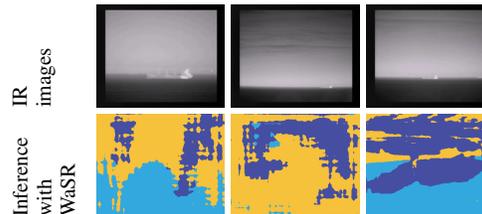
C. Retraining WaSR [2] with IR images

In this experiment, we trained WaSR [2] on our IR images. For this purpose, we created our own ground truth masks of 103 IR images. We split this dataset into 88 images for training and 15 images for validation. We used the model weights obtained from this training to run inference on some of the images from validation. The training time required was about 6 hours. As shown in Table I, fourth row, the results are much better than any of the previous experiments.

However, when the inference is run on the IR images from some of the publicly available maritime datasets [12], the results are not good, refer Table IV. Since the number of training images used is quite low, the trained model cannot be generalized.

Table III summarizes various parameters used during these experiments.

TABLE IV: WaSR [2] retrained on our IR images and inference on IR images from public dataset [12] using these WasR weights



IV. RESULTS

With 88 grayscale images used as training dataset, we were able to retrain the WaSR [2] model such that it produced fair results on the images from the test set. The recall obtained on the test images was 0.917 and the precision obtained was 0.458. The low precision indicates higher percentage of false positives. The clouds were incorrectly identified as obstacles. Also it did not produce any meaningful results on the IR images from some of the publicly available datasets. It indicates that further work on the model is necessary.

V. CONCLUSION

We plan to refine the model further to improve the precision. During this process, we will also create our own dataset of semantically segmented IR images covering various weather, location, and lighting conditions. The dataset will help in making the model robust. We also plan to release this dataset to public for further research. The initial results obtained during this work indicate potential of IR sensors as complement to existing sensors to

improve the overall quality of object detection in maritime environment.

ACKNOWLEDGMENT

We kindly acknowledge the support of MIT AUV (Autonomous Underwater Vehicle) lab and especially Michael DeFilippo for sharing the IR dataset. This work was completed in part with resources provided by the University of Massachusetts' Green High Performance Computing Cluster (GHPCC).

REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [2] B. Bovcon and M. Kristan, "A water-obstacle separation and refinement network for unmanned surface vehicles," in *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31 - August 31, 2020*, IEEE, 2020, pp. 9470–9476. DOI: 10.1109/ICRA40945.2020.9197194. [Online]. Available: <https://doi.org/10.1109/ICRA40945.2020.9197194>.
- [3] M. Kristan, V. S. Kenk, S.Kovačić, and J. Perš, "Fast image-based obstacle detection from unmanned surface vehicles," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 641–654, 2016. DOI: 10.1109/TCYB.2015.2412251.
- [4] T. Cane and J. Ferryman, "Saliency based detection for maritime object tracking," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016. DOI: 10.1109/CVPRW.2016.159.
- [5] T. Cane and J. Ferryman, "Evaluating deep semantic segmentation networks for object detection in maritime surveillance," *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2018. DOI: 10.1109/AVSS.2018.8639077.
- [6] D. D. Bloisi, F. Previtali, A. Pennisi, D. Nardi, and M. Fiorini, "Enhancing automatic maritime surveillance systems with visual information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 824–833, 2017. DOI: 10.1109/TITS.2016.2591321.
- [7] F. E. T. Schöller, M. K. Plenge-Feidenhans'l, J. D. Stets, and M. Blanke, "Assessing deep-learning methods for object detection at sea from lwir images," *International federation of automatic control workshop*, vol. 52, no. 21, pp. 64–71, 2019. DOI: 10.1109/OCEANSKOBE.2018.8559310.
- [8] T. Clunie, M. DeFilippo, M. Sacarny, and P. Robinette, "Development of a perception system for an autonomous surface vehicle using monocular camera, lidar, and marine radar," *International Conference on Robotics and Automation*, 2021.
- [9] M. DeFilippo, "MIT autonomous underwater vehicle (AUV) lab," 2021.
- [10] "MIT autonomous underwater vehicle (AUV) lab dataset," 2021. [Online]. Available: <https://seagrant.mit.edu/auvlab-datasets-marine-perception-2/>.
- [11] B. Bovcon, J. Muhovič, J. Perš, and M. Kristan, "The MaSTr1325 dataset for training deep usv obstacle detection models," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.
- [12] D. D. Bloisi, L. Iocchi, A. Pennisi, and L. Tombolini, "Argos-venice boat classification," *12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, 2015. DOI: 10.1109/AVSS.2015.7301727.