**Ship Detection and Tracking from Airborne Patrol Videos for Maritime Security**

JaeHyun Kim1 , Bui Ngoc An1 , YeonGon Oh1 and Impyeog Lee \*2

1Graduate Student, Department of Geoinfomatics, University of Seoul

163, Seoulsiripdae-ro, Dongdaemun-gu, Seoul, Republic of Korea

Email: {hourguide98, buingocan1989, ohrgon}@uos.ac.kr

2Professor, Department of Geoinfomatics, University of Seoul

163, Seoulsiripdae-ro, Dongdaemun-gu, Seoul, Republic of Korea

Email: iplee820@gmail.com

**KEY WORDS:** Ship detection, Ship Tracking, Illegal ships, Calculating the number of ships,

**ABSTRACT:** Marine resources are limited, and legal users must pay to satisfy demand. However, illegal vessels involved in maritime crimes such as marine pollution continue to occur every year. Detecting and blocking these illegal vessels is crucial. In Korea, vessels are identified and zoomed in when taking maritime images by manually pointing a camera at them from a manned aircraft, and only location information and image data are stored in the images, causing delays in sharing the situation. To solve these challenges, we propose an integrated process of ship detection and tracking using manned aircraft. Our approach is aimed at automatically identifying the location and number of ships in a video to overcome the limitations of current maritime surveillance. For the vessel detection process, we use the VarifocalNet (VFNet) model to train a system capable of detecting vessels. We obtain a non-overlapping dataset for vessel detection and measure the evaluation in terms of mean accuracy (mAP). For ship tracking, we use the DeepSort tracking model and measure the true ship detection rate and Mean Object Tracking Precision (MOTP). It succeeds in identifying 80.2 AP of the ships in the ship images. The tracking model successfully tracked 94% of the real ships, but the MOTP value was a relatively low 0.232. Future improvements aim to increase the detection rate by addressing false ship detections and recognition issues due to changes in the coverage area.

# Introduction

In order to collect marine resources, they must be collected legally and within a designated space. However, there are vessels that collect marine resources illegally to save time and money. Illegal vessels cause negative impacts such as marine pollution and ecosystem destruction, which can be prevented by using vessel monitoring to detect illegal vessels.

Vessel monitoring is primarily accomplished through the use of Automatic Identification Systems (AIS) and video surveillance from manned aircraft. However, to avoid these monitoring methods, illegal vessels tamper with the vessel's information on the AIS terminal or cut off power to conceal the vessel's location. This results in a mismatch between the vessel's AIS identification and the information obtained from the camera's thermal imaging sensor. Once the discrepancy is confirmed, the vessel is checked for illegal activities and the camera is zoomed in to take detailed images. During the imaging process, the vessel is constantly moving and changing direction. This results in a constantly changing view of the vessel from the manned aircraft. This dynamic environment makes ship monitoring difficult. Therefore, it is difficult to identify all vessels identified during ship monitoring or to determine whether a vessel has already been identified. Recently, object detection algorithms have been studied to detect vehicles and people on land, and computer vision has been applied to detect obstacles and ships at sea. However, in previous computer vision, misrecognition often occurs due to noise caused by reflections. Therefore, we want to utilize computer vision for ship detection by applying deep learning. The purpose of this research is to automatically monitor ships by utilizing deep learning and computer vision technology. In the proposed method, a model is proposed to detect and analyze ships on the screen in real time, and an algorithm is applied to recognize the unique features of the ship based on the captured data to determine the similarity with previously captured ships. It is used as a unique identification code and accurately determines whether duplicate films are taken.

In recent years, research has been conducted on detecting ships using object detection algorithms. However, existing studies utilize SAR image data to train detection models, which is not suitable for optical images from manned aircraft. In addition, shortcomings in learning bias have been reported in SAR images (Liu *et al*., 2023). This study aims to solve the problem by eliminating the bias by using optical images as training data. Various types of learning models for ship detection and classification using optical remote sensing images were developed and compared (BoLI *et al*., 2021). However, many differences in detection performance occurred depending on environmental variables such as the weather at the time of the ship acquisition. In this study, we aim to solve the problem of deviations in detection performance by including various weather conditions in the dataset. A study combining the YOLOv5 network model and the DeepSort algorithm to achieve target detection and tracking in ship monitoring was conducted (Jiang *et al*., 2023). In this study, they identified the phenomenon of missed detection when a part of the vessel is occluded. They also identified the problem that most of the vessels can be detected, but the tracking model cannot be tracked when the vessel is far away. The ship target tracking and self-correction coefficient algorithm based on YOLOv3 was proposed and successfully detected 90% of ships(Chun *et al*., 2021). However, there is a problem that the learned ship images are mostly composed of side view data from close to the ground, and the generated model stops tracking when the hull is obscured by the ship in front of it when photographing multiple ships. To solve the problem, this study addresses the problem of degraded tracking of distant ships by addressing the dataset imbalance of the camera's zoom level and environmental variables.

# Methodology

This research was conducted to develop a detection and tracking algorithm to improve the efficiency of vessel monitoring by determining location and number. The manned aircraft compares the visual information seen by the naked eye with the information displayed on the AIS. If any discrepancies or illegal behavior are identified, optical sensors are used to monitor the vessel. Figure 1 shows the process of real-time vessel monitoring.

Video data from a manned aircraft is used as input data. The video data was captured by optical/infrared sensors. To train a deep learning model, we need labeled image data. Therefore, the real-time surveillance video is segmented and converted into images. Images were created by extracting one frame from every second of the recorded video and consist of optical and infrared images. As for the training image, we found that the size of the ship is very small or unidentifiable in the infrared image as shown in Figure 2, so it is not suitable as training data.

Therefore, after removing the thermal image from the optical/thermal image set, the optical image is used to generate training data. To separate the images, we used OCR to recognize the information of the camera lens contained in the image. The lens type and information in the image are shown in Table 1.

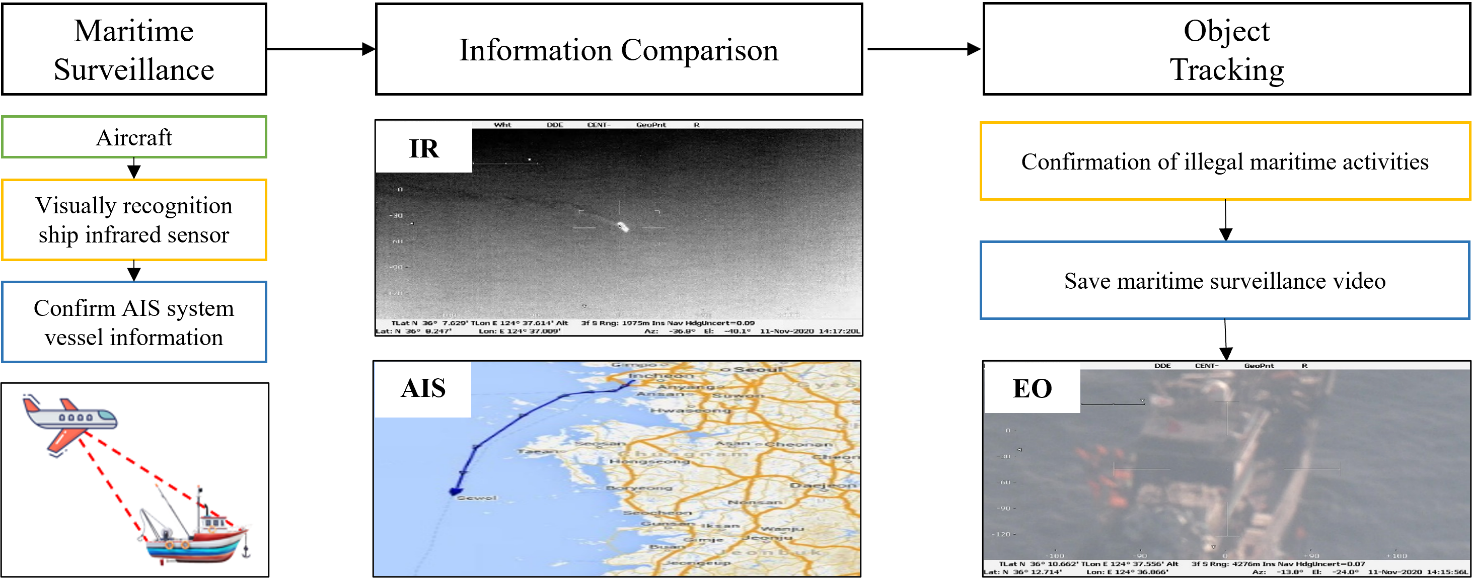


Figure 1. Real-time ship monitoring process

텍스트, 흑백, 스크린샷, 모노크롬이(가) 표시된 사진

자동 생성된 설명

Figure 2. Example of IR image

Table 1. In-screen information based on sensor type

|  |  |  |
| --- | --- | --- |
| Sensor Type | IR | EO |
| Display in the image | HDIR | HDTV |

To create a deep learning model, the ships in the optical image are labeled using VGG (Visual Geometry Group) Image Annotation, and a COCO Json result file is generated to create training data. The training data contains the location information of the ships.

## Ship Detection

In this work, we use video data from manned aircraft equipped with EO/IR sensors to detect and track ships. For deep learning training for ship detection, we use VFNet researched and developed by Zhang, Haoyang, et al. VFNet is based on Res2Net-101-DCN and achieves performance on the COCO dataset. The model uses bounding boxes and utilizes non-maximum suppression (NMS) to handle nested bounding boxes. For training the ship detection model, we augmented the training material by flipping 50% of the data, and a ship was judged as a ship if its confidence score exceeded 0.50. After creating the learning model, we check the ship detection rate through test data to improve the ship detection rate by supplementing the Zoom-level training data with insufficient detection rate. The pickle file extracted by the detection process allows you to check whether a ship is detected and its size.

## Ship Tracking

In the vessel tracking phase, we continuously track the vessel based on the vessel detection results. This process aims to accurately identify real-time vessel movements and track the vessel's path of travel. We utilize a tracking algorithm called DEEPSORT. DEEPSORT uses bounding box-based prediction, setting up multiple anchor boxes with nested bounding boxes and using non-maximum suppression (NMS) to remove overlaps. DEEPSORT improves the performance of Simple Online and Realtime Tracking (SORT) by incorporating appearance information for object tracking. It is also characterized by storing the visual characteristics and appearance of the object so that it can be continuously tracked in a long video even if the object is occluded for a long time. To generate the tracking model, the previously generated detection model is used. To solve the phenomenon that tracking is lost due to obstacles, we supplement the training data with simultaneous images of multiple ships. In addition, by checking and storing the detection results of the ships in the video by checking their features, the tracking path of the ships is derived by repeating the process of checking the identity of the previously detected ships and grouping the same ships into one group.

# result

## DATASET

The dataset used in this study was obtained from FLIR's StarSafire HD sensor mounted on the CN-235, a medium-range multirole airplane operated by the Coast Guard. The sensor is used to capture EO/IR (electro-optical and thermal) video, with a resolution of 1280\*720. The video is captured at a refresh rate of 30 frames per second and supports thermal and optical imagery, allowing the EO/IR sensor to detect and track vessels in a variety of environments.

스크린샷이(가) 표시된 사진

자동 생성된 설명

Figure 3. Example of each Zoom-Level

To perform vessel detection, we selected 28 videos in which vessels were present and used them to train the VFNet deep learning model. The training dataset consists of images from 19 videos, generating a total of 2,071 images. The validation dataset consists of 518 images from 9 videos, and the test dataset consists of 410 images from 4 videos. Each dataset is constructed using non-overlapping optical images. All image labeling was done manually, and the train, validation, and test datasets have different ground coverage at different zoom levels. There are five zoom levels: Wide, Medium, Narrow, Ultra-Narrow, and Super-Narrow, and the images are magnified differently for each zoom level, as shown in Figure 3. The dataset composition for each zoom level is organized as shown in Table 2.

Table 2. Train-Val-Test dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| Zoom level | Train | Validation | Test |
| Wide | 473 | 200 | 80 |
| Med | 213 | 201 | 78 |
| Nar | 597 | 66 | 50 |
| Utnar | 177 | 49 | 96 |
| Spnar | 611 | 2 | 106 |
| Total | 2071 | 518 | 410 |

## DETECTION

|  |  |  |
| --- | --- | --- |
| Zoom level | Number of images | Detection performance: mAP(%) |
| Wide | 80 | 25.8 |
| Med | 78 | 77.4 |
| Nar | 50 | 70 |
| Utnar | 96 | 66.8 |
| Spnar | 106 | 84.1 |
| Total | 410 | 80.1 |

To detect ships, we train the VFnet model with 28 videos containing many ships among the maritime videos provided. When training the ship detection model, the learning rate is set to 0.01 and the momentum is set to 0.9 for 20 iterations using the stochastic gradient descent method. Then, a flip process is applied to the test data for data augmentation to improve the performance of the model. The test results are shown in Table 3.

스크린샷, 물, 야외, 교통이(가) 표시된 사진

자동 생성된 설명

Figure 4. Example of ship detection

In the detection results for each zoom level, except for the Wide zoom level, the average ship detection rate is 75mAP in the other four zoom levels. The low detection rate at the Wide zoom level indicates the difficulty of detection based on visual characteristics due to the low resolution and small objects. The overall detection rate for the video is 80.1% when the threshold is set to 0.5. Figure 4 shows an example of ship detection results.

Table 3. Configuration by Zoom-level dataset.

## Tracking

The ship tracking was performed by combining the ship detection model with the DeepSort model. For the specific performance test, we select a video of 2 minutes and 39 seconds with 1280\*720 resolution to evaluate the performance. The performance evaluation is shown in Table 4. Our results show that we can detect 5,154 vessels out of a total of 5,446 labeled vessels in the video, which is about 94% or more of the vessels.

물, 야외, 교통, 선박이(가) 표시된 사진

자동 생성된 설명

Figure 5. Example of ID assignment

To verify the tracking model, the total number of ships present in the video is checked. We used the 2 minutes and 39 seconds of video [a] and the 4 minutes and 13 seconds of video [b] with a resolution of 1280\*720 to determine the total number of ships in the video by identifying each ship. Table 5 shows the actual number of vessels in the two videos and the number of vessels counted by the tracking algorithm. Figure 5 is an example of vessel tracking.

Table 4. The number of the image in test dataset Table 5. Predicting the number of ships in video

|  |  |
| --- | --- |
|  | The number of the images |
| Ground Truth | 5446 |
| Our Result | 5154 |
| Recall | 0.529 |
| MOTP | 0.232 |

|  |  |  |
| --- | --- | --- |
|  | Predict the number of ships | The number of ships in the video |
| [a]video | 390 | 8 |
| [b]video | 137 | 16 |

# conclusion

This research combines deep learning models and tracking techniques to develop an algorithm for detecting and tracking ships based on EO/IR sensor images taken by manned aircraft. In the detection phase, a VFNet deep learning model is developed to automatically detect and track ships by utilizing video data from manned aircraft equipped with EO/IR sensors in various environments. The VFNet model applies a varifocal loss function to effectively overcome the class imbalance problem and reaches a performance that can detect most of the ship's location and features. In particular, the detection rate is high at various zoom levels except for the WIDE zoom level. More than 80% of all ships are detected, confirming that the model in this study is effective in detecting ships quickly. In the tracking phase, the DeepSort tracking model is applied to successfully track more than 94% of the labeled vessels, which confirms that our model has sufficient tracking performance even for vessels that are difficult to identify with the naked eye or for vessels that are only partially captured in the video. However, it also recognizes the same vessel as a new vessel after the zoom level of the screen or creates false IDs when the screen is shaky. This tracking ability also works well in multi-vessel situations, which can improve real-time monitoring and response to maritime situations. This could have many applications, including maritime crime prevention. In our next research, we will solve the problem of misrecognition when the zoom level changes during tracking, and we cannot confirm that the tracked object is the same object.

# Acknowledgement

This study was supported by "Development of satellite based system on monitoring and predicting ship distribution in the contiguous zone(20200495)" of the Korea Coast Guard.

# reference

Bo, L. I., Xiaoyang, X. I. E., Xingxing, W. E. I., & Wenting, T. A. N. G. 2021. Ship detection and classification from optical remote sensing images: A survey. Volume 34, Issue 3, pp. 145-163.

Jiang, Q., & Li, H. 2023. Silicon energy bulk material cargo ship detection and tracking method combining YOLOv5 and DeepSort. *Energy Reports*, 9, pp. 151-158.

Liu, C., & Li, J. 2021. Self-correction ship tracking and counting with variable time window based on YOLOv3. Complexity, pp. 1-9

Liu, S., Wang, Z., Wang, R., Chen, Y., Fan, X., & Li, W. 2023. SAR Ship Detection Based on Deep Domain Adaptation with Limited Samples. *Procedia Computer Science*, pp. 378-385.

Wojke, Nicolai, Alex Bewley, and Dietrich Paulus. 2017. Simple online and realtime tracking with a deep association metric. *IEEE international conference on image processing*, pp. 3645-3649.

Zhang, H., Wang, Y., Dayoub, F., & Sunderhauf, N. 2021. VarifocalNet: An IoU-aware Dense Object Detector. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8514-8523.