3D Mapping for Improving the Safety of Autonomous Driving in Container Terminals

Ngo Quang Vinh* · Ji-Hoon Park** · Hyeon-Soo Shin*** · † Hwan-Seong Kim

*M.S. Course, Graduate School of Korea Maritime and Ocean University, Busan, 49112, KOREA
**PhD. Course, Department of Smart Port Logistics, Korea Maritime and Ocean University, Busan, 49112, KOREA
***KMI-KMOU Cooperation Program, Graduate School of Korea Maritime and Ocean University, Busan, 49112, KOREA
† Professor, Dept. of Logistics, Korea Maritime and Ocean University, Busan, 49112, KOREA

Abstract : *Automated container terminals (ACTs) contribute many benefits to operation shipments, such as productivity, management cost, and real-time freight tracking, in which ensuring a high level of safety in container terminals is extremely important to prevent accidents and optimize operations. This study proposes a method for increasing safety levels in container terminals through the application of object detection with state-of-the-art EfficientDet model support. A distance estimation method is employed to accurately measure the distance to objects and provide proximity alerts when predefined safety thresholds are exceeded. 3D Mapping technology is utilized to create a detailed representation of the container terminal in a virtual environment. This enables a comprehensive visualization of the surroundings, including structures, equipment, and objects. By combining object detection and distance estimation results with the 3D map, potential safety issues can be identified with greater precision. A realistic container terminal scenario was used to evaluate the robustness of the proposed method.*

Key words : *object detection, distance estimation, 3D mapping, deep learning, autonomous container truck, container terminal*

1. Introduction

Maritime transportation serves as the fundamental pillar of global trade, encompassing approximately 80% of all international trade. The maritime containerization market reached over 165 million Twenty-foot Equivalent Units (TEUs) in 2021(UNCTAD, 2022). Therefore, Automated Container Terminals (ACTs) or smart ports are paid more attention by mainstream scientists around the world with the purpose of increasing productivity, cost-effective management, lean operations, and cutting off idle time. Moreover, the safety level in container handling is considered as an essential factor when managing container terminals. The conversion of a port into a smart port has emerged as a highly pragmatic approach to establishing an advanced and intelligent port ecosystem. Implementing the advanced technologies from the Fourth Industrial Revolution changed how these ports operate and manage their daily activities. This modernization journey encompasses the integration of several cutting-edge technologies, such as Big Data Analytics, Artificial Intelligence (AI), Internet of Things (IoT), Block Chains, Robotics, etc.(Heikkilä et al., 2022). The integration of automation technologies automates various tasks within the port, such as container handling, stacking, and transportation. Autonomous Guided Vehicles (AGVs), remote handling, and Automated Guided Cranes (AGCs) streamline operations, minimize human error, and increase productivity while ensuring a safer working environment. Especially, Autonomous Yard Trucks (AYTs) are vital in revolutionizing container handling operations within terminals since, compared to AGVs or AGCs, the AYTs contribute a broader versatility, flexibility, last-mile connectivity, and significant safety improvements. One of the key technologies employed is sensor technology. These trucks are equipped with a diversity of sensors such as vibration, Light Detection and Ranging (LiDAR)(Ogawa and Takagi, 2006), Global Positioning System (GPS), radar, and cameras. These sensors enable the trucks to perceive their surroundings, detect obstacles, and navigate through complex port environments with precision and accuracy.

In this paper, to ensure the safety of container handling operation by AYTs, the application of object detection in the container terminals with the support EfficientDet is proposed. Estimating distance from AYTs to detected objects method is considered to avoid the obstacles, and reduce collisions. In addition, this paper proposes a 3D

Mapping procedure between the AYTs and a built virtual terminal map in order to remote management assistance and operation surveillance.

The rest of this study is presented as follows. Section 2 discusses the related research on object detection and distance calculation in container ports. Section 3 contributes the detailing of proposed methodologies about the application in AYTs operation. Finally, Sections 4 and 5 describe the experiment results and conclusion of the paper, respectively.

2. Literature Review

Over the past few years, object detection and distance estimation are critical components of AYTs in the container terminals. Many studies have proposed innovative solutions by leveraging advanced techniques to enhance safety, optimize operations, and improve overall efficiency within port environments. One notable study proposed a framework of deep learning approach for object detection, object moving displacement mapping, and speed estimation (Xu et al., 2022). The method was implemented in both AGVs and people for port surveillance with the support of You Only Look Once (YOLO). However, in fact, object occlusion and clutter at ports can be crowded and chaotic environments with numerous objects, equipment, and people, the paper was just experienced with the simple case.

Another relevant study utilized LiDAR to estimate distance in container terminals, aim to provide real-time proximity alerts to prevent collisions and promote safe working conditions. Zhao et al.(2019) conducted a study investigating the utilization of LiDAR sensors for the detection and tracking of pedestrians and vehicles at intersections. LiDAR sensors offer portability and a compact size, making them suitable for deployment on AGVs in real-world AYT operations, thereby providing high-resolution localization information(Zou et al., 2021). However, these two papers face the restriction of Field of View (FoV) and challenging environmental conditions (rain, snow, fog, or dust). Besides, radar data can be employed for accurate obstacle detection and identification to support navigation tasks(Patole et al., 2017). The paper's limitations arise from the availability of data and performance metrics associated with the reviewed signal processing techniques. Additionally, several studies have focused on leveraging

spatial-temporal information to enhance traffic safety, highlighting its potential for improving the safety of AGVs(Bayouh et al., 2021). However, The limitations include the availability and size of the dataset used for training and evaluating the proposed transfer learning-based hybrid Convolutional Neural Networks (CNNs). Small or imbalanced datasets can impact the generalizability and performance of the model, potentially leading to limitations in the reported results.

Apart from the utilization of these sensors, cameras can be a low-cost approach for enriching the Driving Assistance System (DAS). By employing a monocular vision for 3D detection, Zhe et al.(2020) proposed an end-to-end inter-vehicle distance estimation methodology to improve the accuracy and robustness even in complex traffic scenarios. Furthermore, CNNs have shown their outstanding performance in object detection. Thus, Liu et al.(2021) improved CNNs to detect ships, and vessels under different weather conditions. Nonetheless, the paper's limitations include a lack of comprehensive comparative analysis with other ship detection methods or benchmarking against existing approaches.

From the restriction analysis, this paper implements the EfficientDet model integrated with distance estimation to improve the safety operation. In addition, there is an innovation by connecting 3D maps to reflect real-time container terminal environments for visualization and management purposes.

3. Proposed Methodology

3.1 Object Detection with EfficientDet in container terminal

The EfficientDet model, introduced by Google Research in 2019, encompasses a family of object detection models known for their exceptional accuracy and efficiency(Tan et al., 2020). To capture more contextual information, EfficientDet models utilize a Bi-directional Feature Pyramid Network (BiFPN) and enhance the accuracy of small objects through a weighted bi-ocular loss function.

In the context of container terminals, the EfficientDet model can be trained on extensive datasets containing container images in diverse positions and orientations. This enables real-time detection of containers, providing information on their presence and location. The

effectiveness of this approach relies on the quality of the training data and the accuracy of the model. Moreover, employing the model for container detection enhances safety by enabling the automated detection of hazardous materials or objects requiring special handling within the container terminal environment.

The EfficientDet model builds upon the EfficientNet as its backbone, incorporating an enhanced architecture known as BiFPN. BiFPN is designed to facilitate multi-scale feature fusion, which involves the aggregation of features at various resolutions. In detail, given a list of multi-scale features:

$$\overrightarrow{P^{in}} = (P_{l_1}^{in} + P_{l_2}^{in}, \cdots, P_{l_n}^{in}) \tag{1}$$

where $P_{l_i}^{in}$ represents the feature at level l_i, the goal is to find a transformation f that can effectively aggregate different features and output a list of new features:

$$\overrightarrow{P^{out}} = f(\overrightarrow{P^{in}}) \tag{2}$$

The conventional top-down FPN(Lin et al., 2017) is illustrated in Fig. 1. In detail, it will take 4 levels from 3 to 7 input features $\overrightarrow{P^{in}} = (P_3^{in}, \cdots, P_7^{in})$, in which P_7^{in} indicates a feature level with resolution of $1/2^i$ the input images.
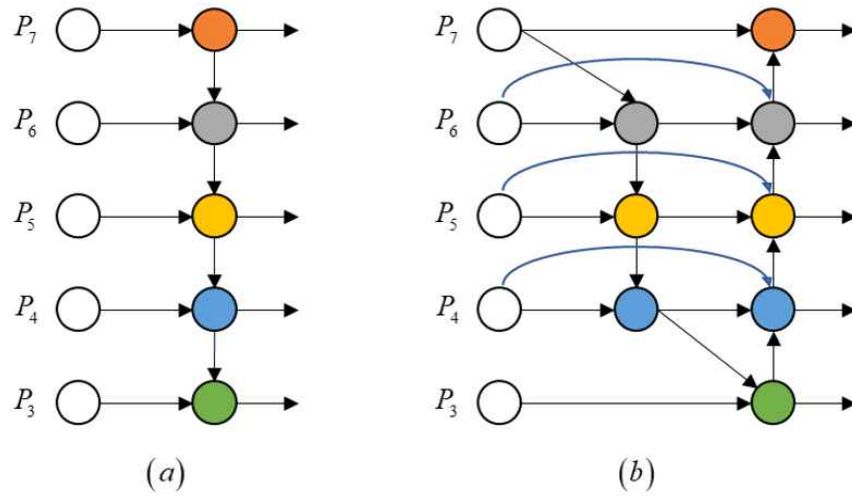

Fig. 1 Feature framework: (a) Feature Pyramid Network(FPN) (b) Bi-directional Feature Pyramid Network(BiFPN)
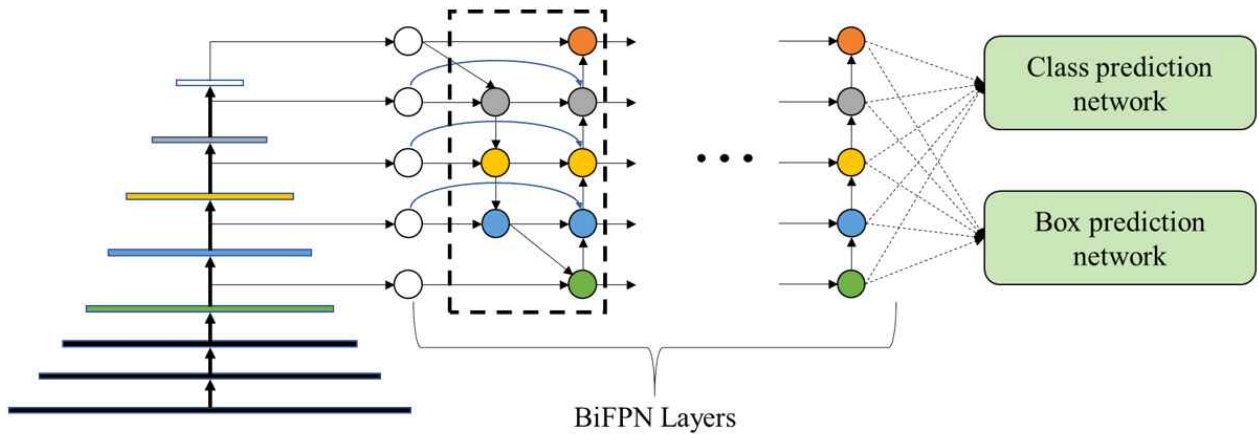

Fig. 2 The architecture of the EfficientDet

3.2 Distance Estimation by Pin Hole Image Rule

The proposed framework utilizes the principles of imaging to map the displacement of objects, specifically containers, from the container terminal to the real world. The pinhole camera rule is employed to reveal the movement of detected containers in the physical world. This involves transforming the object's movement from the two-dimensional imaging plane of the camera to the three-dimensional plane in the real world using geometric perspective transformation, also known as the pinhole imaging procedure. Fig. 3 illustrates the mapping rule for object movement, where the object's imaging size is larger when the distance between the object and the camera is smaller. The specific formulas for displacement mapping are provided in Equation (3).

$$D = \frac{f^* H}{H'} \qquad (3)$$

3.3 3D Mapping Program in container terminal

3D mapping refers to the process of creating a three-dimensional representation of objects or areas in the real world. Various methods can be employed for 3D profiling, including the use of stereo cameras or measuring the depth of objects or features based on focus. One significant advantage of 3D mapping is its ability to employ advanced techniques for visualization and data gathering. By generating a realistic view of a location, a 3D map offers valuable insights for port authorities and decision makers, enabling them to make informed decisions and effectively utilize the information provided. For container terminal context, by profiling the port scenario and updating real-time to a 3D map in Fig. 4, significant safety surveillance can occur in the container handling operation.

The workflow of the proposed system can be summarized as follows: real-time data on position and camera streaming are collected and updated into a database. The proposed method then retrieves this data to map realistic features onto a 3D map.

In detail, an algorithm was built to map the container terminal. Let m denote a 3D map:

$$m_t = \left\{ \left(o_t^1, p_t^1 \right); \ldots; \left(o_t^n, p_t^n \right) \right\} \qquad (4)$$

where o_t^1 indicates the objects in the terminal at the time t. p_t^1 denotes their positions. Thus, the map formulation can be rewritten:

$$m_t = \left\{ \left(o_t^i, p_t^i \right) \right\} \qquad , i := 0 \rightarrow n \qquad (5)$$

The objective function is to find the best appropriate map, and can be formulated as follow:

$$\underset{m}{\arg\max}\, E\!\left(m\, vertd_t \right) \qquad (6)$$

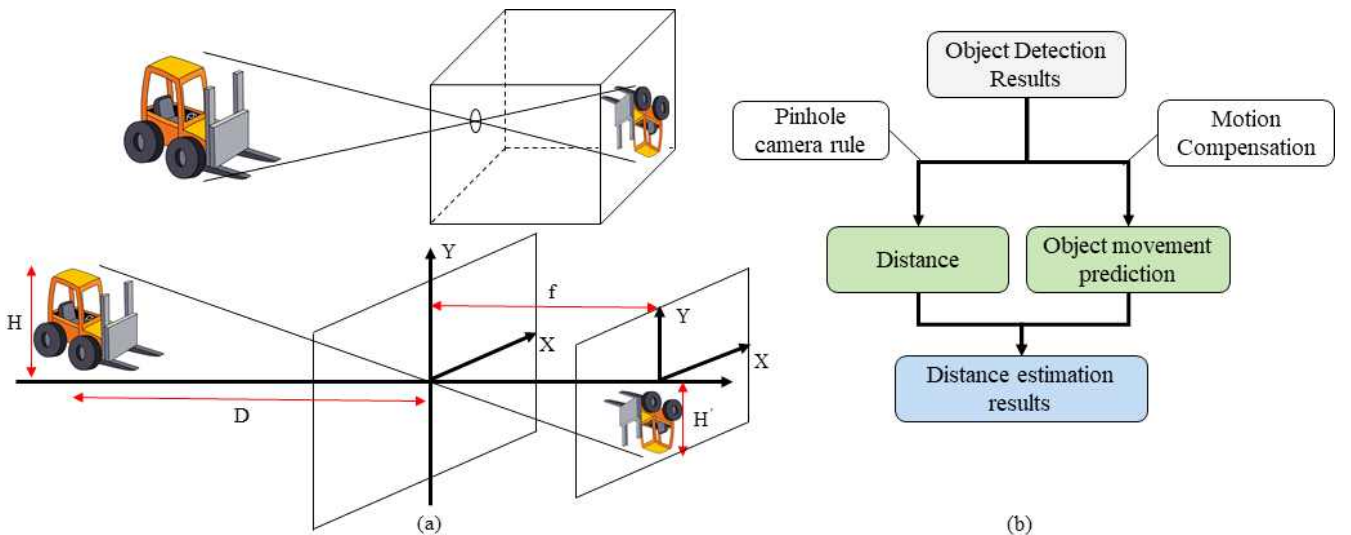where d_t indicates the sequence of actual data collected from sensors in the terminal.

Fig. 3 The distance estimation theory
(a) The geometry of a pinhole camera (b) The diagram of distance estimation theory

As followed in [Thrun et al., 1998] the map function can be transformed:

$$E(m|d_t) = E(m) \int ... \int \prod_{i=0}^{n} E(m'|d_t,m) \cdot \prod_{i=0}^{n-1} E(d_t^{i+1}|d_t) dm_t \qquad (7)$$

4. Experiment Results and Analysis

4.1 Training Preparation

Container terminal data was manually collected from videos depicting container handling activities in a seaport. These videos captured various elements such as containers, gantry cranes, quayside container cranes, yard tractors, workers, and ships. Specifically, five videos were recorded at a frame rate of 30 frames per second (fps) and separated into 2000 images with a resolution of 1920x1080 pixels, representing typical port traffic scenarios. For experimentation, a testing program was developed using the Python programming language and the Tensorflow library, running on a computer equipped with an AMD Ryzen 5 5600G processor (12 cores) operating at a base clock speed of 3.9 GHz, 16 GB of RAM, and supported by an NVIDIA GeForce RTX 3060 graphics card. For the 3D map, the virtual terminal map was built in Unity (2021.3.25.f1).
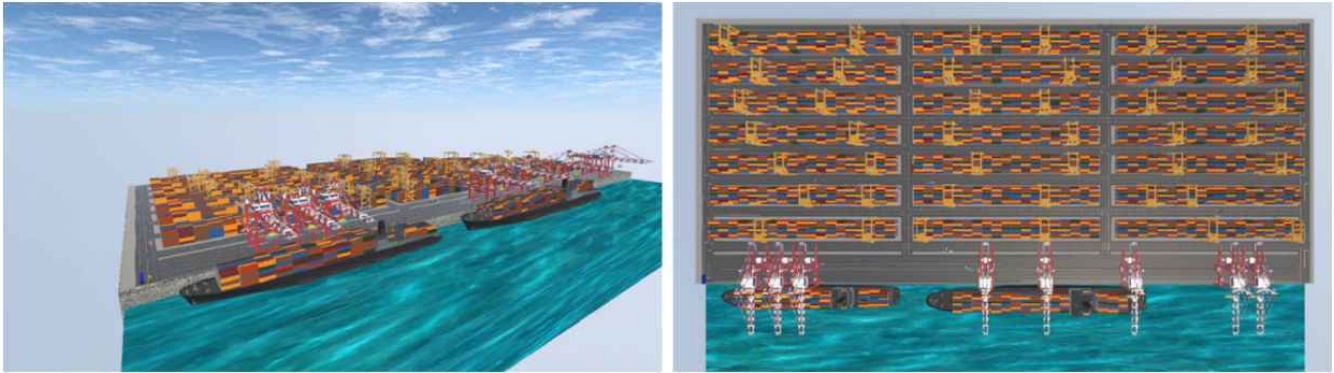

Fig. 4 The 3D map of container terminal

Fig. 5 The results of object detect and distance estimation

4.2 Object Detection and Distance Estimation Results

The object detection outcomes for each video were depicted in Fig. 5, demonstrating the successful detection of various target objects such as containers, cranes, yard tractors, and humans using the proposed framework. Notably, the bounding boxes accurately aligned with the objects, irrespective of variations in object imaging sizes. Additionally, the framework calculated the distances to the detected objects and visualized them at the center of the respective bounding boxes.

4.3 3D Mapping Simulation

The results of the 3D mapping experiment demonstrated promising outcomes in accurately mapping the objects within the container terminal environment, as in Fig. 6. The proposed method effectively captured the three-dimensional profiles of various objects, including containers, cranes, yard tractors, and humans. The generated 3D maps provided a realistic representation of the location, enabling enhanced visualization and information-gathering capabilities. The precise alignment of objects in the maps, regardless of variations in object imaging sizes, showcased the robustness of the approach. Moreover, the incorporation of distance estimation further enhanced the utility of 3D maps by providing valuable spatial information. Overall, the results highlight the potential of the proposed 3D mapping technique for improving situational awareness, planning, and safety measures within container terminal operations.

Table 1 The analysis of object detection results

Objects	No. Testing images	Failed	Detected successfully proportion (%)	Average processing time (s)
Crane	641	17	97.3	0.102
Container	2000	93	95.4	0.118
Yard Tractor	568	36	93.7	0.101
Human	90	5	94.4	0.111
Average			95.2	0.108

Fig. 6 The 3D mapping results: (a) The realistic container terminal scenario (b) The visualization of the port in 3D map

5. Conclusions

This paper aims to enhance safety and efficiency in the operation of container terminals with the application of object detection, distance estimation, and 3D mapping technologies in container terminals offer significant benefits and advancements. By employing object detection algorithms, such as the EfficientDet model, the presence and location of containers can be accurately detected in real-time, reducing the need for manual inspection and improving overall operational efficiency. Moreover, the built 3D mapping in Unity provides a realistic representation of the container terminal by the connection with the IoT modules. The 3D map reflects the real-time environment at container ports aiding in visualization, planning. Terminal operators are able to give seasonable aid or decision when remoting container handling surveillance. These technologies collectively contribute to increased safety levels, reduced risks of accidents and delays, and improved management of container handling operations. In terms of future research, employing these advancements into more objects in container terminal operations opens up new possibilities for automation, optimization, and enhanced situational awareness, ultimately driving towards more efficient and secure port operations in the future.

References

[1] Bayoudh, K. et al.(2021), "Transfer Learning Based Hybrid 2D-3D CNN for Traffic Sign Recognition and Semantic Road Detection Applied in Advanced Driver Assistance Systems", Applied Intelligence, Vol. 51, No. 1, pp. 124-142.

[2] Heikkilä, M. et al.(2022), "Innovation in Smart Ports: Future Directions of Digitalization in Container Ports", Journal of Marine Science Engineering, Vol. 10, No. 12, 1925.

[3] Liu, R. W. et al.(2021), "An Enhanced CNN-enabled Learning Method for Promoting Ship Detection in Maritime Surveillance System", Ocean Engineering, Vol. 235, 109435.

[4] Lin, T. Y. et al.(2017), "Feature Pyramid Networks for Object Detection", 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 936-944.

[5] Ogawa, T. and Takagi, K.(2006), "Lane Recognition using On-vehicle LIDAR", 2006 IEEE Intelligent Vehicles Symposium, pp. 540-545.

[6] Patole, S. M. et al.(2017), "Automotive Radars: A Review of Signal Processing Techniques", IEEE Signal Processing Magazine, Vol. 34, No. 2, pp. 22-35.

[7] Tan, M. et al.(2020), "EfficientDet: Scalable and Efficient Object Detection", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10778-10787.

[8] Thrun, S. et al.(1998), "A probabilistic approach to concurrent mapping and localization for mobile robots, Autonomous Robots, 5, pp. 253-271.

[9] UNCTAD(2022), "Review of Maritime Report 2022" https://unctad.org/system/files/official-document/rmt2022_en.pdf (accessed on May 26th, 2023)

[10] Xu, X. et al.(2022), "Exploiting High-fidelity Kinematic Information from Port Surveillance Videos via a YOLO-based Framework", Ocean & Coastal Management, Vol. 222, 106117.

[11] Zhao, J. et al.(2019), "Detection and Tracking of Pedestrians and Vehicles using Roadside LiDAR Sensors", Transportation Research Part C: Emerging Technologies, Vol. 100, pp. 68-87.

[12] Zou, Q. et al.(2021), "A Comparative Analysis of LiDAR SLAM-based Indoor Navigation for Autonomous Vehicles", IEEE Transactions on Intelligent Transportation Systems, Vol. 23, No. 7, pp. 6907-6921.

[13] Zhe, T. et al.(2020), "Inter-Vehicle Distance Estimation Method Based on Monocular Vision Using 3D Detection", IEEE Transactions on Vehicular Technology, Vol. 69, No. 5, pp. 4907-4919.