대한조선학회논문집 Research Paper
Journal of the Society of Naval Architects of Korea
eISSN:2287-7355, Vol. 63, No. 3, pp. 130-140, June 2026
https://doi.org/10.3744/SNAK.2026.63.3.130

무인수상선의 자율 접안을 위한 카메라 기반 상대 위치 추정에 관한 연구

임예은 · 표춘선 · 김기훈 · 박한솔†
한국해양과학기술원 부설 선박해양플랜트연구소

A Study on Relative Position Estimation Using a Camera-based Localization Method for Autonomous Berthing of Unmanned Surface Vehicles

YeEun Lim · Chunseon Pyo · Kihun Kim · Hansol Park†
Korea Research Institute of Ships and Ocean Engineering

This study proposes a camera-based relative position estimation technique for autonomous berthing of unmanned surface vehicles (USVs). The system estimates the relative distance and bearing angle from the USV to the berth using only a single onboard camera. Existing camera-based approaches for berth recognition have primarily relied on marker detection, which is vulnerable to light reflection and requires costly installation and maintenance. To overcome these limitations, this study adopts a 3D detection approach based on deep learning. A key challenge in applying deep learning is the need for large amounts of training data, which is often expensive and time-consuming to collect. To address this issue, we employed a 3D scene reconstruction technique to generate training data. First, images of the berth were captured using a camera mounted on a drone. Then, a 3D reconstruction of the environment was performed using Neural Radiance Fields (NeRF), enabling the generation of multi-view images of the berth as seen from various positions and angles from the perspective of the USV. These synthetic images were then used to train a Real-Time Monocular 3D Detection (RTM3D) model, which performs keypoint-based 3D object detection to estimate relative distance and bearing. To validate the proposed method, we conducted experiments using a camera mounted on the USV Aragon Cat 40-02 at the berth of Jebu Marina in Hwaseong-si, Gyeonggi-do, South Korea. The results confirm that the proposed method can successfully estimate the relative distance and bearing angle from the USV to the berth, demonstrating its potential for reliable, marker-free autonomous berthing.

Keywords : Unmanned surface vehicle(무인수상선), Autonomous berthing(자율 접안), 3D object detection(3차원 객체 탐지), 3D reconstruction(3차원 형상 복원), Localization(위치 추정), Markerless(비마커)

1. 서 론

최근 무인수상선의 자율운항 기술이 많은 관심을 받고 있으며 그에 따라 다양한 연구가 활발하게 진행되고 있다. 무인수상선의 자율운항을 위해서는 상황인식, 자율제어 등 여러 기술들의 고도화가 필요하다. 그 중에서도 자율 접안 기술은 무인수상선의 운항을 위해 반드시 필요한 기술로, 계류장에서 출발하여 운항을 마친 후 다시 계류장에 접안하기 위한 기술을 의미한다. 기존의 자율 접안 연구의 경우 접안 위치에 마커를 설치하여 센서를 통해 해당 마커를 인식하는 연구가 대다수이다. 무인수상선의 자율 접안 기술에 주로 활용되는 마커 형식으로는 ArUco 마커, 광원 마커 등이 있다. 이러한 마커와의 기하학적 관계를 이용하여 상대 위치를 추정할 수 있다. 그러나 마커의 일부분이 훼손되거나 조도에 의해 특정 부분이 센서에 제대로 인식되지 못하면 예측 정확도가 떨어지거나 예측 실패로 이어질 가능성이 있다. 이는 무인수상선의 접안 실패 및 정박되어 있는 타선 또는 계류장과의

Received : October 1 2025 | Revised : 2 February 2026 | Accepted : 19 March 2026
† Corresponding author : Hansol Park, hspark@kriso.re.kr
It is noted that this paper is a revised edition based on the proceedings of KAOSTS 2025 in Busan

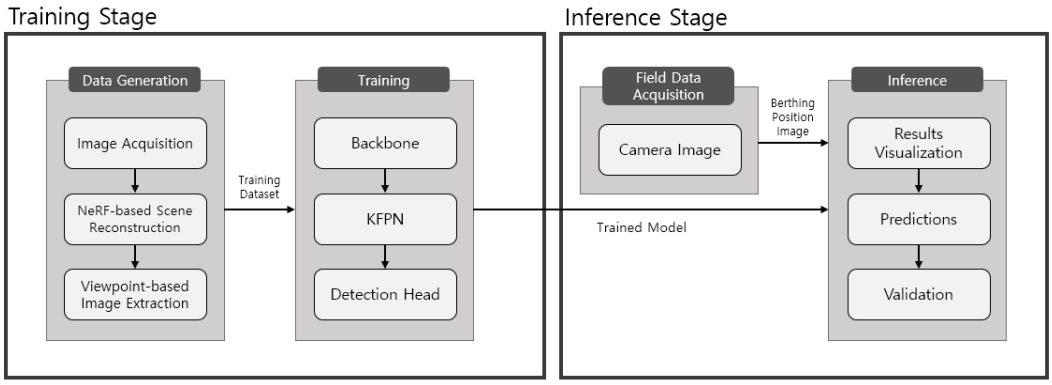

Fig. 1 Flow diagram

충돌을 일으킬 수 있다. 따라서 본 연구에서는 마커 기반 방법에서 발생할 수 있는 위와 같은 문제들을 극복함과 동시에 운용 편의성을 함께 확보하고자 하였다.

3차원 객체 탐지 기술은 2차원 탐지와 다르게 탐지 대상의 위치, 크기, 방향 추정이 가능하다. 따라서 본 연구에서는 비마커 방식으로 3차원 객체 탐지 기술을 접안 위치에 적용하여 기존 연구의 문제를 보완하였다. 제안하는 기법은 특징점 기반 탐지 기법인 RTM3D(Real-Time Monocular 3D Detection)(Li et al., 2020) 알고리즘을 바탕으로 한다. 이 방법을 통해 카메라 이미지만을 활용하여 계류장을 3차원으로 탐지한 뒤 해당 정보를 바탕으로 무인수상선으로부터 계류장까지의 상대 거리 및 상대 방위각을 추정한다. 아울러 무인수상선의 자율 접안을 위한 접안 위치 학습 데이터는 일반적으로 선박에 탑재되어 탐지에 사용될 카메라를 통해 취득한다. 그 이유는 선박의 접안 과정과 동일한 시야와 조건에서 데이터를 취득하기 위함이다. 그러나 이러한 데이터 취득 방식은 많은 시간과 인력, 비용이 요구된다. 또한 기상이 좋지 않을 경우 데이터 취득에 더 많은 시간이 소요된다. 따라서 이러한 문제를 해결하기 위해 본 연구에서는 3차원 형상 복원 기술을 활용하였다. 먼저 선박에 탑재되어 탐지 시 활용되는 카메라가 아닌, 활용이 용이한 드론과 드론에 탑재된 카메라를 통해 접안 위치인 계류장을 다각도에서 촬영하여 영상 이미지 데이터를 취득하였다. 이후 3차원 형상 복원 기술인 NeRF(Neural Radiance Fields)(Mildenhall et al., 2021)를 활용하여 계류장 환경을 3차원으로 복원하였다. 복원된 계류장 환경에서 다양한 접안 위치 데이터를 취득하여 학습 데이터로 활용하였다.

인지 센서를 활용한 무인수상선의 자율 접안 기술로 카메라 기반의 인지 기술이 다수 연구되어 오고 있다. Liu et al. (2018)의 연구에서는 단안 카메라를 사용하여 마커 기반의 도킹 연구를 수행하였다. 이때 학습 데이터의 날씨 및 색상 변화에도 강건한 합성곱 신경망을 개발하여 접안 위치 인식을 수행하였다. Woo et al. (2019)의 연구에서는 단안 카메라와 라이다를 사용하여 마커 기반의 접안 연구를 수행하였다. 그 결과 VGG-19 모델 기반으로 다양한 환경 변화에도 강인한 상대 위치 추정 기술을 개발하였다. Figueiredo et al. (2020)의 연구에서는 단안 카메라를 사용하여 마커 기반의 도킹 연구를 수행하였다. 그 결과 물의 탁도에 따라 광원의 색상을 자동으로 조정하고 마커가 일부 가려졌을 때 전체 부분을 인식할 수 있도록 하는 필터링 기법을 개발하였다. Volden et al. (2022)의 연구에서는 단안 카메라, 스테레오 카메라 및 3차원 라이다를 사용하여 마커 기반 접안 연구를 수행하면서 단안 카메라와 스테레오 카메라, 라이다 데이터를 각각 비교하여 실험 후 결과를 고찰하였다. Xu et al. (2023)은 스테레오 카메라를 사용하여 마커 기반의 도킹 연구를 수행하였으며 이를 통해 스테레오 카메라 데이터로 3차원 정보를 취득한 후 수중 시계 제한으로 인한 왜곡을 보정하여 탐지 정확도를 높이는 알고리즘을 개발하였다. 이상의 선행 연구 분석 결과로부터 기존 연구들은 대부분 마커를 활용하여 계류장을 인식하는 방법을 활용하였음을 알 수 있다. 또한 카메라 외에도 다른 센서를 같이 활용하였다. 그러나 이러한 접근은 빛 반사에 의한 인식 오류 가능성이 있으며 마커 설치 및 유지보수에 자원이 지속적으로 소모된다는 단점이 있다. 이에 본 연구에서는 별도의 마커나 추가 센서를 사용하지 않고 카메라만으로 계류장 환경 자체의 특징 정보를 활용하여 딥러닝 기반의 상대 위치 추정 기술을 제안한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 본 연구에서 제안하는 상대 위치 추정 방법에 대해 설명하고, 단안 카메라 기반의 계류장 인식을 위해 사용된 학습 네트워크인 RTM3D의 구조와 동작 원리에 대해 서술한다. 3장에서는 본 연구가 제안하는 학습 데이터 취득 방식인 3차원 형상 복원 과정에 대해 소개하고 생성한 학습 데이터를 활용한 학습 과정에 대해 다룬다. 4장에서는 실해역 데이터를 기반으로 제안 기법의 성능 검증을 진행한 내용을 다루고 있으며 5장에서는 본 연구의 결론을 맺고 있다.

2. 상대 위치 추정 방법

본 연구에서 제안하는 상대 위치 추정 방법은 Fig. 1에 제시되어 있는 순서를 따른다. 마커 없이 주변 환경 정보를 바탕으로 무인수상선으로부터 계류장까지의 상대 위치를 추정하기 위하여 특징점 탐지 기반의 RTM3D 네트워크를 활용하여 신경망 구조를 구축하였다. 이 학습 과정은 단안 카메라로 촬영된 RGB 이미지를 입력으로 받아 3차원 바운딩 박스의 중앙점 1개와 각 꼭짓점 8개를 추정하고 이를 카메라 좌표계 기준 좌표로 출력하는 것이

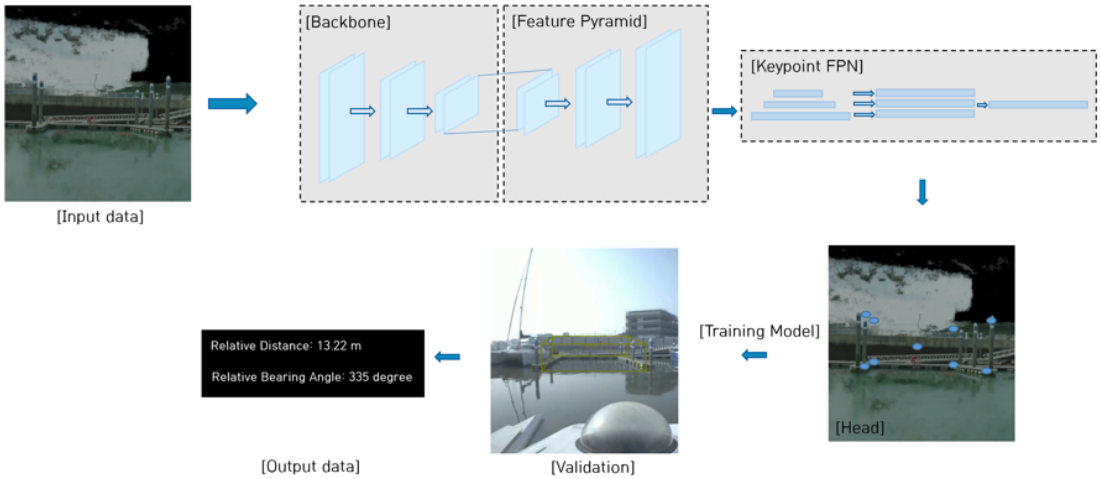

Fig. 2 Overview of keypoint detection network

다. 이 네트워크는 Fig. 2와 같이 backbone, keypoint feature pyramid, detection head로 구성되어 있다. 학습 데이터는 가상의 데이터를 사용하였으며 성능 검증 데이터는 실해역 데이터를 활용하였다. 관련 내용은 3장 및 4장에서 다룬다.

2.1 RTM3D 네트워크

본 연구에서 활용한 3차원 객체 탐지 네트워크인 RTM3D는 단일 RGB 이미지 입력만으로 실시간 3차원 객체 탐지를 수행하는 one-stage 기반의 신경망 구조이다. 이 네트워크는 다중 센서 없이도 이미지 내 객체의 특징점을 예측하고, 이를 통해 3D 바운딩 박스의 크기, 위치, 방향 등의 속성을 동시에 추정할 수 있다. 특히 RTM3D는 경량화된 구조와 빠른 연산 속도 덕분에 자율운항 시스템과 같은 실시간성이 요구되는 환경에서 많이 활용된다. 본 연구의 목적은 상대 거리 및 상대 방위각 추정이다. RTM3D의 특징점 기반 학습은 3차원 위치 추정 및 방향 추정에 직관적이고 효과적인 구조이기 때문에 본 연구에서 해당 네트워크를 적극적으로 활용하였다. 복잡한 후처리 없이도 결과를 활용할 수 있어, 후속 연구나 실제 시스템 연동에도 적합하다고 판단하였다.

Backbone은 ResNet-18(He et al., 2016)을 활용하였다. ResNet-18은 residual connection을 활용하여 깊은 네트워크에서도 안정적인 학습이 가능하며, 적은 파라미터 수에도 불구하고 우수한 성능을 보이는 것이 특징이다(Lim et al., 2025a). 입력된 RGB 이미지에 대해 네트워크는 다운샘플링을 통해 공간 해상도를 줄이고 채널 수를 확장시켜 정보를 요약한다. 이후 세 번의 bilinear interpolation과 1×1 convolution을 통해 점진적으로 업샘플링이 이루어지며, 이로써 고해상도 feature map이 복원된다. 각 업샘플링 단계에서 채널 수는 순차적으로 256, 128, 64로 축소되며, 이렇게 얻어진 feature map은 초기의 저수준 feature와 결합되어 높은 수준의 의미 정보와 저수준의 세부 정보를 동시에 반영한다.

KFPN(Keypoint Feature Pyramid Network)은 동일한 크기의 특징점 탐지에 특화된 구조를 갖는다(Li et al., 2020). KFPN은 여러 스케일의 feature map을 동일한 해상도로 정규화한 뒤, softmax 기반의 soft weight를 계산하여 각 스케일의 중요도를 반영한다. 이후, 해당 soft weight와 feature map을 요소별 곱셈하고 선형 가중 합산하여 scale-space score를 얻으며 이는 식 (1)과 같다. 이러한 방식은 다양한 스케일에서 일관된 특징점을 탐지할 수 있도록 하여, 스케일 변화에 강건한 3차원 객체 탐지를 가능하게 한다.

$$S_{score} = \sum_{f} \hat{f} \cdot softmax(\hat{f}) \tag{1}$$

Detection head는 KFPN으로부터 얻은 feature map을 기반으로 실제 3차원 객체의 속성을 예측하는 단계이다. 본 구조는 CenterNet 방식을 기반으로 각 픽셀이 특징점일 확률을 예측하고 이를 통해 객체의 중심점을 탐지한다. 이후 중심점 주변의 정보를 활용하여 바운딩 박스의 크기, 방향, 중심으로부터의 거리와 같은 정보를 추정한다. 특히, 객체의 일부가 이미지 경계를 벗어나더라도 중심점만 이미지에 포함된다면 객체 전체를 탐지할 수 있다는 점에서 높은 실용성을 가진다. 본 연구에서도 계류장의 일부가 카메라 FOV(Field Of View)를 벗어나는 경우가 있는데 이러한 경우에도 탐지 가능함을 확인하였다.

2.2 네트워크 기반 3차원 계류장 탐지

본 연구에서는 앞서 설명한 RTM3D 네트워크 구조를 기반으로 계류장의 3차원 탐지를 위한 학습을 수행하였다. 학습된 모델을 통해 무인수상선의 시점에서 촬영된 이미지에 적용하여 무인수상선으로부터 계류장까지의 상대 위치 정보를 추출하고자 하였다. 학습 라벨링 생성 과정에서 계류장을 3차원 객체로 정의하였으며 해당 객체의 중심점, 크기, 방향 등을 포함하는 3차원 바운딩 박스를 추정 대상으로 설정하였다.

학습된 모델은 성능 검증 이미지에서 계류장을 탐지하여 무인

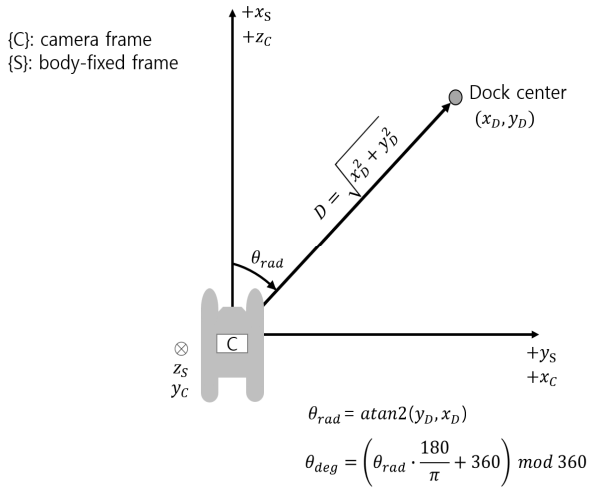

Fig. 3 Definitions of relative distance(D) and bearing(θ_{deg}) in camera frame and body fixed frame

수상선으로부터의 3차원 상대 위치를 추정할 수 있다. 이를 통해 무인수상선에서 계류장까지의 상대 거리(relative distance) 및 상대 방위각(relative bearing)을 계산하였다. 이 두 값은 선체의 위치를 원점으로 하는 선박 기준 좌표계에서 계류장의 위치를 추정하는 핵심 요소로 활용되며, Fig. 3과 같이 나타낼 수 있다.

RTM3D 네트워크를 계류장 탐지에 적용함으로써, 다중 센서 없이도 카메라 이미지로부터 무인수상선과 계류장 간의 상대 위치를 추정하도록 하였다.

3. 학습 데이터 취득 및 학습 과정

본 연구에서는 3차원 형상 복원 방법인 NeRF를 활용하여 3차원 객체 탐지 학습에 사용될 학습 이미지를 취득하였다. 먼저, 드론에 탑재된 카메라를 활용하여 계류장의 영상을 다각도에서 취득하였다. 이후 NeRF 계열 모델인 Instant-NGP(Müller et al., 2022)를 활용하여 계류장 환경을 3차원으로 복원하였다. 이후 무인수상선에 탑재된 카메라 위치를 기준으로 다양한 거리 및 각도에서의 계류장 이미지를 추출하였고, 이를 3차원 객체 탐지 학습 데이터로 활용하였다.

3.1 계류장 데이터 취득

무인수상선의 계류장 접안을 위한 3차원 객체 탐지 학습을 위해 실제 계류장 데이터를 취득하였다. 계류장 영상은 드론에 탑재된 카메라를 이용하여 촬영하였으며, 드론은 DJI Inspire 3 모델을, 카메라는 Zenmuse X7 모델을 사용하였으며 사양은 Table 1에 제시하였다. 촬영 시 드론은 다양한 고도와 시점에서 계류장 전반을 포함하도록 촬영 경로를 계획하였으며, 실제 계류장 형상을 다각도에서 확보하는 데 초점을 두었다. 촬영 대상은 경기도 화성시 제부마리나에 설치된 계류장으로, 본 연구에서 무인수상선 접안의 목표 지점으로 설정되었다. 계류장의 실제 구조는 Fig. 4에 나타내었다.

Table 1 Specifications of the camera(model: Zenmuse X7)

Parameter	Measurement
Sensor size	23.5 mm × 15.7 mm
Effective pixels	24 MP
Maximum video resolution	6016 × 3200
Maximum frame rate	Up to 60 FPS
Gimbal control range	Tilt: +40° to −125°, Pan: ±300°, Roll: ±20°

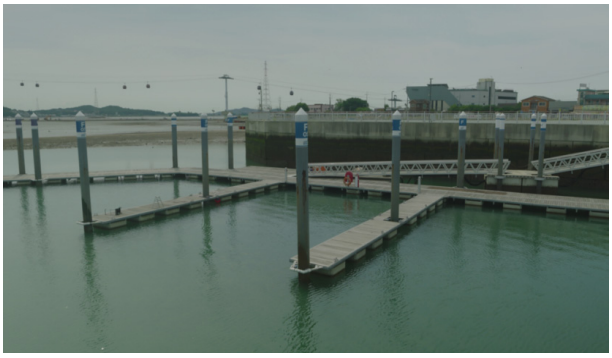

Fig. 4 Target berth

3.2 NeRF 모델 기반 계류장 형상 복원

NeRF는 불연속적인 시점에서 촬영한 이미지를 입력으로 사용하여 새로운 시점에서의 물체 이미지를 3차원으로 복원하는 view synthesis 모델이다. 본 연구에서 필요한 학습 데이터는 선박을 직접 운항하며 다양한 각도와 위치에서 취득해야 하기 때문에 시간적, 비용적 어려움이 존재한다. NeRF는 불연속적인 학습 데이터의 연속적인 장면 표현을 위해 이미지들 간의 관련 정보를 추출하여 공간에 대한 일반화를 수행하여 공간을 복원한다. 이를 통해 다양한 위치 및 각도에서 데이터를 취득할 수 있으므로, 기존의 데이터 취득 어려움을 보완할 수 있다.

본 연구에서는 Instant-NGP를 활용하여 NeRF 기반 파이프라인을 학습 및 렌더링하였다. Instant-NGP는 다중 해상도의 학습 가능한 해시 임베딩을 도입하여, NeRF의 메모리 효율성과 이미지 품질을 유지하면서도 복잡한 MLP 연산으로 인한 렌더링 속도 저하를 효과적으로 극복하였다(Chen et al., 2024). 이는 multi-resolution hash encoding 방식을 활용하여 기존 NeRF 방식의 학습 속도를 개선한다. 해당 방식은 식 (2)와 같으며 장면의 복잡도에 따라 서로 다른 해상도의 다중 격자를 구성하고, 각 격자의 꼭짓점에는 학습 가능한 F차원의 특징 벡터가 할당된다. 이때 F는 각 격자의 꼭짓점이 보유하는 임베딩 공간의 차원을 의미하며, 모델은 이러한 벡터를 학습함으로써 장면의 색상 정보를 효율적으로 인코딩한다. 특징 벡터들은 해시 테이블에 저장되며, 주어진 위치 좌표를 해싱하여 대응되는 특징 벡터를 효율적으로 탐색

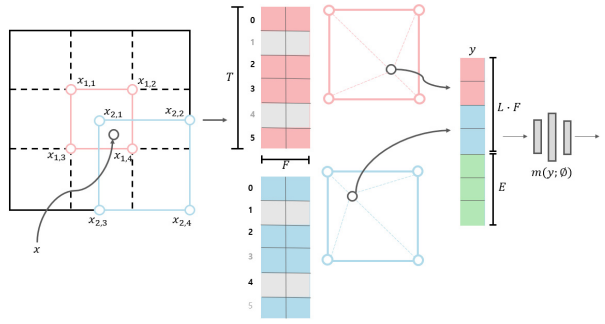

Fig. 5 Multi-resolution hash encoding(Lim et al., 2025b)

할 수 있다. 이후 인코딩 과정에서는 해당 위치가 포함된 격자들에서 인접한 꼭짓점들의 특징 벡터를 선형 보간하여 신경망에 입력 가능한 고차원 인코딩 벡터를 생성한다.

$$h(x) = \left(\bigoplus_{i=1}^{d} x_i \pi_i \right) \bmod T \tag{2}$$

이때, x_i는 d차원 좌표의 i번째 성분을 의미하며 π_i는 차원별 가중치를 나타낸다. 해시 함수 $h(x)$는 좌표를 해시 테이블 크기 T에 대하여 모듈로 연산한 후에 유효한 인덱스로 변환하는 과정을 거친다. 이를 통해 각 격자의 꼭짓점에 저장된 특징 벡터를 효율적으로 탐색할 수 있다.

또한, 카메라의 시선 방향 정보가 보조 입력으로 추가되어 최종 입력은 위치 정보와 시선 방향 정보로 구성된다. 모델의 출력은 주어진 위치와 방향에 대응하는 볼륨 밀도(density, 투명도)와 방사 복사량(radiance, 픽셀 색상 정보)이며, 이 과정을 통해 공간의 3차원 구조와 시각적 특성을 동시에 학습할 수 있다. Fig. 5는 이러한 multi-resolution hash encoding의 과정을 도식화한 것이다.

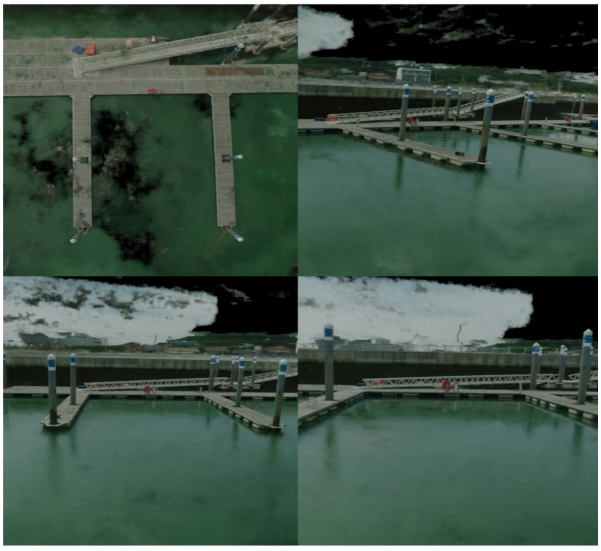

Fig. 6 Reconstructed 3D berth

Instant-NGP를 통해 계류장을 3차원으로 복원하였으며 복원한 계류장 형상을 Fig. 6에 나타내었다.

Fig. 7 The berth from the perspective of the USV: (a) distant view (left); (b) view from inside the berth (right)

3.3 라벨링

학습 라벨링 생성 과정은 모델이 정확하게 인식하고 예측할 수 있도록 예측이 필요한 정보를 부여하는 과정으로 객체 탐지 과정에 필수적이다. 본 연구는 KITTI 3D object detection benchmark (Geiger et al., 2012) 형식으로 라벨링 수행하였다. KITTI 데이터셋은 자율주행 연구를 위해 구축된 대표적인 공개 데이터셋으로 객체의 위치·크기·방향을 3차원 공간에서 정밀하게 기술할 수 있는 라벨 형식을 갖는다.

라벨링 과정은 카메라 파라미터 정보를 활용하여 자동 라벨링으로 수행되었다. COLMAP(Schönberger and Frahm, 2016; Schönberger et al., 2016) 알고리즘을 활용하여 카메라 파라미터를 취득하였다. COLMAP이란 이미지 간 특징점 정합에 기반한 알고리즘으로, 일반적인 정적인 환경에서는 높은 정합 정확도를 보인다. 본 연구에서도 정적 장면을 사용하였기 때문에 해당 알고리즘으로 카메라 파라미터를 취득하였다. 이후 실제로 얻은 데이터를 라벨링 툴로 확인하며 fine tuning 과정을 거쳤다.

장거리에서 바라본 계류장과 계류장 내부에 진입하여 바라본 계류장의 차이가 존재한다. 장거리에서 바라본 계류장의 경우 Fig. 7(a)와 같이 계류장 전체가 보이며, 무인수상선의 위치 및 각도에 따라 주변 환경이 다르게 보인다. 본 연구는 무인수상선의 카메라 위치에서 계류장의 중앙점으로부터 약 20m 초과 40m 미만인 구역을 장거리로 정의하였다. 또한, 계류장 내부에 진입하여 바라본 계류장의 경우 Fig. 7(b)와 같이 계류장의 일부분이 보이며, 무인수상선의 위치 및 각도에 따라 계류장의 일부분이 안보이기도 한다. 본 연구는 무인수상선의 카메라 위치에서 계류장의 중앙점으로부터 약 25m 미만인 구역을 단거리로 정의하였다. 이때 선박이 계류장의 중앙과 거리가 있는 좌·우 측면을 따라 진입하는 경우, 내부 진입 상태에서도 계류장 중앙점을 기준으로 계산된 상대 거리가 장거리 구간과 겹칠 수 있다. 본 연구는 장거리에서 바라본 계류장과 계류장 내부에 진입하여 바라본 계류장의 라벨링을 분할하였다. 두 라벨링의 경우 다른 GT(Ground Truth) 위경도를 갖는다.

Fig. 8 The USV used for verification (Aragon Cat 40-02)

3.4 학습 과정

모델 학습 과정은 목표 객체를 정확하게 탐지하고 관련 정보를 추정할 수 있도록 데이터를 기반으로 모델을 최적화하는 핵심 단계이다. 본 연구에서는 총 1,000장의 이미지를 사용하였으며, 학습 데이터(train) 700장, 검증 데이터(validation) 150장, 시험 데이터(test) 150장으로 구성하였다.

Train 및 validation 시 사용된 데이터들은 NeRF를 활용하여 재구성된 계류장에서 취득한 이미지를 활용하였으며, 양쪽에서 중복해서 사용되지 않았다. Test 데이터는 실해역 데이터만 사용하였으며, 이를 통해 모델 성능을 분석하였다. 일부 구간에서 탐지 성능 저하가 확인됨에 따라, 형상 복원된 계류장 환경에서 추가 데이터를 취득하고 이를 활용한 추가 학습을 통하여 성능 저하를 보완하였다. test 데이터는 고수위 및 중수위, 저수위에 대한 데이터 셋으로 이루어져 있으며, 성능 평가 시 각 수위에 대한 성능 검증에 활용되었다.

모델 학습은 Ubuntu 20.04 환경에서 NVIDIA GeForce RTX 3080 GPU를 기반으로 진행되었으며, PyTorch(Ver. 1.10.0) 프레임워크를 활용하였다. 학습 안정성과 정확도 향상을 위해 하이퍼파라미터를 조정하였으며, 최종적으로 학습률은 1.25×10^{-4}, 배치 크기는 4, 총 학습 epoch 수는 700으로 설정하였다. 또한 학습에는 ResNet-18 기반의 네트워크 구조를 적용하였고, 입력 해상도는 선박에 탑재된 카메라 촬영 기본 사이즈인 1000×1232로 통일하여 모델이 다양한 장면에 대해 안정적으로 수렴할 수 있도록 하였다.

4. 성능 검증

본 연구는 형상 복원 기술을 활용하여 학습 데이터를 취득하였고 3차원 객체 탐지 기술을 활용하여 계류장을 탐지하였으며 탐지 예측 결과를 통해 상대 위치값을 도출하였다. 성능 검증은 복원 데이터를 포함하지 않고 실해역에서 직접 취득한 영상으로만 수행하였다. 성능 검증용 계류장 영상은 무인수상선에 탑재된 카메라를 이용하여 촬영하였다. 무인수상선은 선박해양플랜트연구소의 아라곤 캣 40-02호를 활용하였으며, Fig. 8에 나타내었다. 카메라는 Ladybug5+(LD5P-U3-51S5C-B)의 전방 카메라 영상을 활용하였으며, 사양은 Table 2와 같다.

Table 2 Specifications of the camera (model: Ladybug5+)

Parameter	Measurement
Lens	6 × 4.4 mm focal length
Megapixels	30 MP (5 MP × 6 sensors)
Resolution	2448 × 2048
Frame rate	30 FPS (JPEG Compressed)
Pixel size	3.45 μm

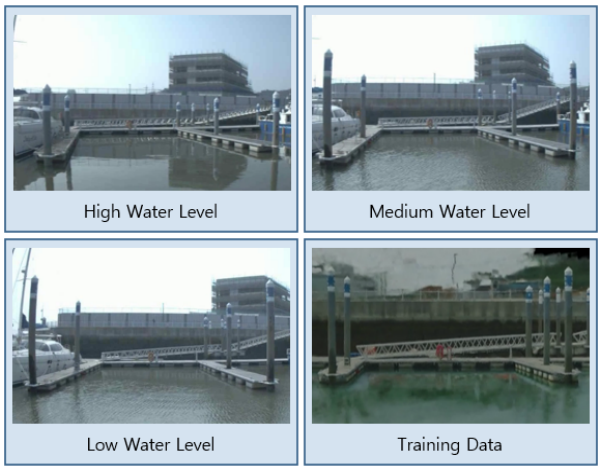

Fig. 9 The data used for verification at each water level and training

형상 복원 결과는 이미지의 유사도 평가에 주로 활용되는 성능 검증 지표인 PSNR(Peak Signal-to-Noise Ratio), SSIM(Structural Similarity Index Measure), LPIPS(Learned Perceptual Image Patch Similarity) 세 지표를 통해 성능 검증을 수행하였다. 또한 3차원 객체 탐지 결과는 위치 정확도를 비롯한 탐지 성능을 평가하는 AP(Average Precision)를 통해 성능 검증을 수행하였다. 이후 상대 위치 추정 결과는 추정값과 실제값의 위치 차이를 계산한 RMS(Root Mean Square)를 통해 최종 성능 검증을 수행하였다. 학습 모델의 성능 검증은 실제 무인수상선이 접안 시나리오로 운항하여 카메라로 취득한 데이터를 저장 후 활용하였다. 이때 조수간만의 차가 큰 편인 서해 특성상 시간에 따라 고수위, 중수위, 저수위 데이터로 성능 검증을 수행하였으며 각 성능 검증 데이터의 수위 및 학습 데이터의 수위는 Fig. 9에 나타내었다.

4.1 ground truth 위치 취득

무인수상선의 GT는 Hexagon NovAtel의 CPT7(SPAN GNSS + INS) 수신기를 이용하여 취득하였고, 계류장 GT 위치는 Trimble SPS855 RTK-GNSS 수신기를 이용하여 위경도 형태로 취득하였다. 무인수상선의 탑재 장비는 RTK 모드에서 1 cm + 1 ppm (RMS) 수준의 위치 정확도를 갖고, 계류장의 위치를 취득한 장비는 RTK 모드에서 수평 8 mm + 1 ppm (RMS), 수직 15 mm + 1 ppm (RMS) 수준의 위치 정확도를 갖는 고정밀 GNSS 수신기이다. 무인수상선은 접안 과정 동안 탑재 카메라를 통해 영상 데이

Table 3 Performance evaluation of 3D reconstruction

Metric	Result
PSNR ↑	27.17
SSIM ↑	0.83
LPIPS ↓	0.12

터를 취득함과 동시에 RTK-GNSS를 통해 선박의 위치 데이터를 기록하였다. 이때 이미지 데이터와 RTK-GNSS 데이터는 동일 ROS 환경에서 타임스탬프 기반으로 기록되었으며, 서로 다른 샘플링 주파수로 인해 발생하는 시간 불일치를 보정하기 위해 영상 프레임의 타임스탬프를 기준으로 RTK-GNSS 위치 데이터에 선형 보간을 적용하여 시간을 정렬하였다.

계류장의 GT는 계류장의 양 끝단 지점 및 중앙 기둥을 대상으로 RTK-GNSS를 이용하여 현장에서 직접 측정하였으며, 해당 끝단 및 중앙 기준 좌표를 활용하여 계류장의 GT를 정의하였다. 계류장에서 취득한 좌표는 동일 지점에 대해 반복 측정한 후 평균값을 사용하여 최종 GT를 산출하였다.

4.2 계류장 복원 성능 검증

원본과 유사한 형상 복원을 수행하는 것은 신뢰할 수 있는 학습 데이터를 확보하기 위해 매우 중요하다. 그러나 시각적 비교만으로 그 정확성을 충분히 판단하기엔 한계가 있기 때문에, 본 연구에서는 성능 검증 지표를 활용하여 학습 데이터의 품질을 정량적으로 평가하고자 하였다.

앞서 언급한 대로, 성능 검증 지표로는 PSNR, SSIM 및 LPIPS 세 가지 지표를 활용하였다. 이는 Mildenhall et al. (2021)에서 활용한 성능 검증 방식을 참고하였다. 세 가지 평가 지표는 원본 이미지의 장면과 복원된 이미지의 동일한 위치 장면을 비교하여 평가하는 방식이다. PSNR은 생성 혹은 압축된 영상의 화질에 대한 손실 정보를 평가하기 위해 개발되었으며 두 이미지의 화질 정보를 비교하여 결과값을 데시벨(dB) 단위로 표시한다. PSNR은 최소값 0에서 값이 클수록 정확도가 높은 것을 의미한다. 이때 값이 30 이상인 경우는 사람이 두 이미지의 차이를 인지하기 어려울 정도의 흡사한 경우를 의미한다. SSIM은 휘도, 대비, 구조 총 3가지 측면에 대해 고려하여 평가하는 방식이다. SSIM은 최소값 0에서 최대값 1인 범위를 가지며 값이 클수록 정확도가 높은 것을 의미한다. LPIPS은 인간의 시각 인식을 기반으로 개발되었으며 VGGNet을 통해 특징값을 구하고 특징값 간의 유사성을 측정하는 방식이다. LPIPS은 최소값 0에서 최대값 1인 범위를 가지며 값이 작을수록 정확도가 높은 것을 의미한다.

세 가지 지표로 평가한 결과는 Table 3와 같다. PSNR 평가는 27.17, SSIM 평가는 0.83, LPIPS 평가는 0.12로 확인되었다. PSNR 결과를 통해 사람이 인지할 수 있는 수준의 화질 저하를 확인하였으며, SSIM 결과를 통해 원본과의 높은 구조적 유사성을 확인하였다. 또한 LPIPS 결과를 통해 시각적으로 두드러진 차이가 없는 우수한 수준으로 확인되었다. 이는 형상 복원본이 구조적으로는 원본과 높은 유사성을 가지지만, 세부 화질에서는 일부 손실이 존재함을 의미한다. 그러나 이러한 수준 차이는 딥러닝 학습에 필요한 주요 시각적, 구조적 특징은 충분히 보존되는 결과이다. 따라서 본 연구의 계류장 복원 결과물을 학습 데이터로 활용하는 데 있어 큰 문제가 되지 않는 것으로 판단된다.

Fig. 10 Visualization of detection results

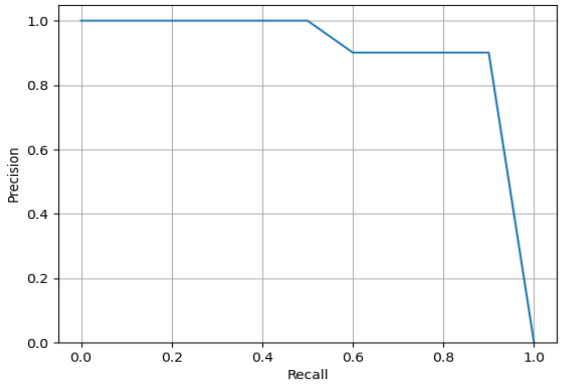

Fig. 11 Precision-Recall curve

4.3 계류장 탐지 성능 검증

정확한 형상 복원을 통해 학습 데이터를 확보하였더라도, 객체 탐지 단계에서 탐지 성능이 확보되지 않으면 실제 활용에 제약이 발생한다. 따라서 본 연구에서는 계류장을 대상으로 수행한 3차원 객체 탐지 결과의 신뢰성을 검증하기 위해 성능 평가 지표를 활용하여 정량적인 분석을 수행하였다.

성능 검증 지표로는 대표적인 객체 탐지 성능 검증 기법인 AP를 활용하였다. AP 기법은 정밀도(precision)와 재현율(recall)의 관계를 기반으로 계산되는 지표로, 객체 탐지 결과에서 얼마나 정확하게 목표 객체를 찾아냈는지를 종합적으로 평가한다. 정밀도는 모델이 탐지한 객체 중 실제로 올바른 탐지의 비율을 의미하며, 재현율은 실제 존재하는 객체 중에서 모델이 탐지해낸 비율을 의미한다. 이 두 지표는 상충 관계에 있으므로, 특정 임계값에서 하나의 값만으로는 탐지 성능을 온전히 평가하기 어렵다. AP는 다양한 임계값 구간에서 계산된 정밀도-재현율 곡선(precision-recall curve)의 면적을 적분하여 산출되며, 값이 클수록 탐지 정확도가 높은 것을 의미한다.

이러한 성능 검증은 형상 복원 평가 데이터셋을 탐지하여 탐지 결과를 통해 도출하였다. Fig. 10은 계류장 탐지 결과를 가시화

Table 4 RMS values for relative distance and bearing angle at each water level and distance

Distance	Water level		
	High	Intermediate	Low
Long-range relative distance	1.88m	1.95m	1.52m
Short-range relative distance	0.89m	0.89m	0.76m
Long-range relative bearing angle	4.58°	6.79°	4.15°
Short-range relative bearing angle	2.57°	1.67°	1.59°

한 모습을 나타내고, Fig. 11은 정밀도-재현율 곡선을 나타내며 AP 결과는 87.31%이다. 이는 객체 탐지 분야에서 일반적으로 우수한 성능으로 간주되는 85% 이상의 결과로, 안정적인 계류장 탐지 성능을 확인할 수 있다.

또한, 실시간성을 평가하기 위해 실시간 처리 성능을 확인하였다. 탐지 단계의 평균 처리 속도는 32FPS로 측정되었으며, 이는 프레임당 약 31.25ms/frame에 해당한다. 또한 저장된 라벨로부터 상대 거리 및 상대 방위각을 산출하는 후처리 단계의 평균 소요 시간은 0.28ms/frame으로 확인되었다. 따라서, 전체 인지 파이프라인 과정에서 결과값 출력에 소요되는 시간은 약 31.53ms/frame으로, 31.72Hz 정도가 된다. 이는 본 연구에서 사용된 무인수상선의 자율 제어 주기인 10Hz보다 충분히 빠르므로, 자율 접안 시 필요한 제어에 실시간으로 탐지 결과를 제공할 수 있을 것으로 판단된다.

4.4 추정 위치 성능 검증

형상 복원과 객체 탐지 단계가 성공적으로 이루어졌다면, 최종적으로는 추정된 상대 위치가 실제 위치와 얼마나 근접하는지에 대한 검증이 필요하다. 따라서 앞서 언급된 대로 본 연구에서는 3차원 객체 탐지를 통해 탐지된 계류장의 위치를 기반으로 상대 거리 및 상대 방위각을 추정하고, 추정된 위치와 실제 위치와의 차이로 RMS 값 계산을 통해 정량적으로 평가하여 성능을 검증하였다. 장거리와 단거리에 대해서 각 수위별로 상대 거리와 상대 방위각에 대한 RMS 값을 도출하였으며, 그 결과를 Fig. 12~Fig. 14 및 Table 4에 나타내었다. 저수위에서 단거리 상대 위치 정확도가 가장 높았으며, 이는 학습 데이터용 형상 복원 결과와의 장면 유사성이 가장 높았기 때문으로 판단된다.

아울러, 실제 운용 관점에서는 해당 오차가 접안 과정에서 계류장과의 충돌위험이 없는 수준인지에 대한 검토가 필요하다. Mentjes et al. (2022)의 연구에서는 실제 계류장 접안 지원 시 부두의 마커 표식을 기준으로 정지 지점을 설정하고, 정지 위치에 대한 목표 허용오차를 2m 미만으로 요구됨을 설명하였다. 이는 단거리 구간에서의 정확도가 운용적으로 중요한 평가 축임을 보여주며, 본 연구에서는 단거리로 정의한 구간의 예측 오차를 검토하여 운용 안정성을 평가하였다.

무인수상선의 제원은 길이(L) 11.9m, 폭(B) 6.0m이며, 계류장은 “ㄷ” 형태로 내부 치수는 가로(W) 13.1m, 세로(H) 20.0m이다. 계류장 내부 공간과 무인수상선을 직사각형으로 근사하고, 계류장을 무인수상선만 사용하며, 무인수상선이 계류장 내부 중앙으로 진입한다고 가정하면, 횡방향 여유 공간(C_T)과 종방향 여유 공간(C_L)은 각각 식 (3), (4)와 같이 계산된다.

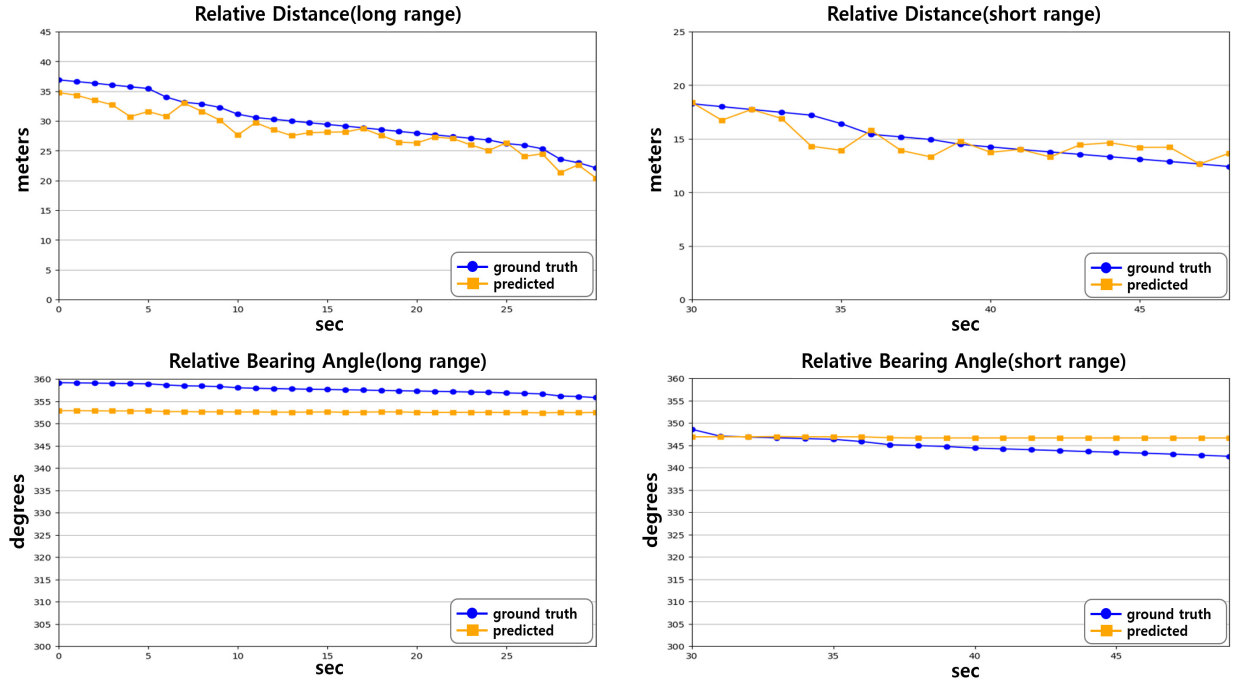

Fig. 12 Relative distance and relative bearing angle under high water level conditions

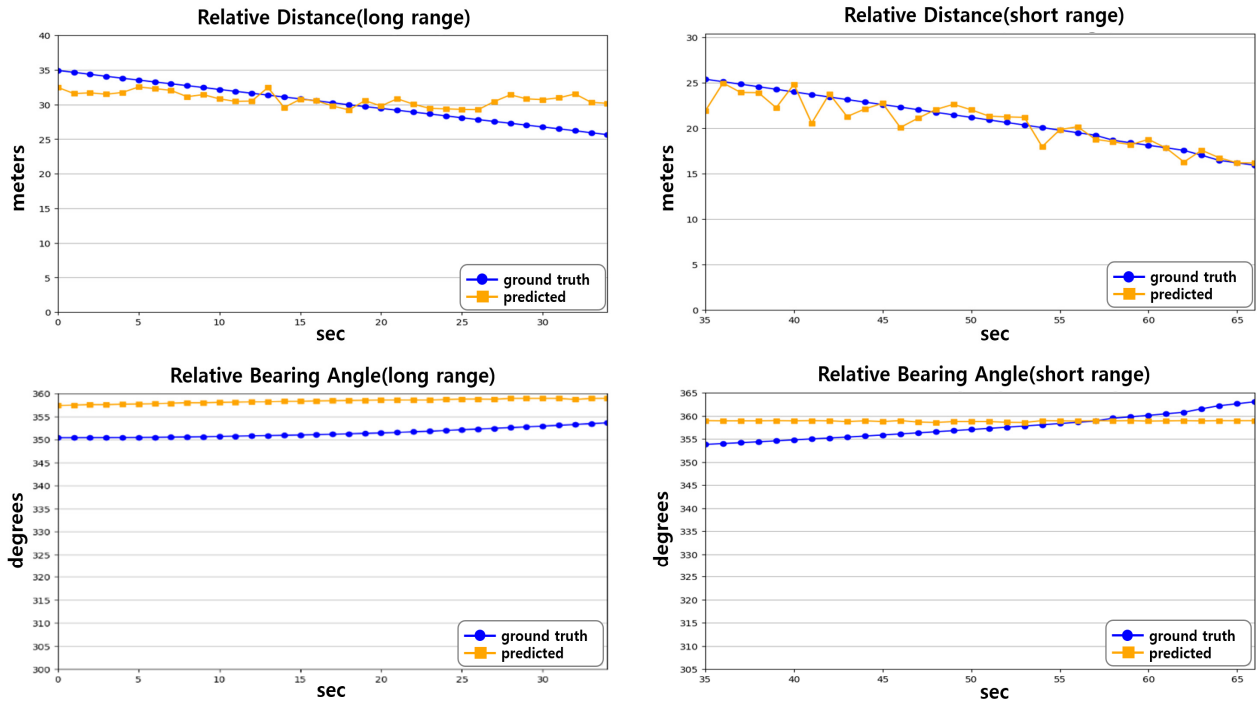

Fig. 13 Relative distance and relative bearing angle under medium water level conditions

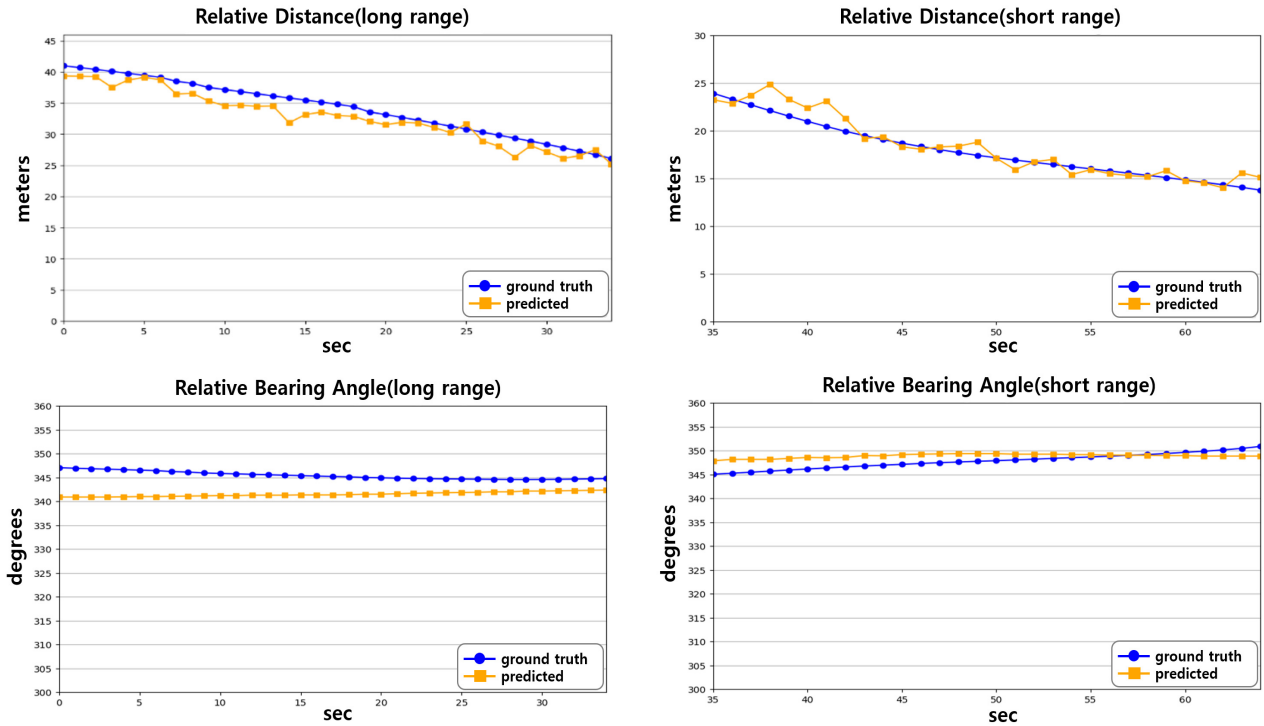

Fig. 14 Relative distance and relative bearing angle under low water level conditions

$$C_T = \frac{W-B}{2} = \frac{13.1-6.0}{2} = 3.55m \quad (3)$$

$$C_L = \frac{H-L}{2} = \frac{20.0-11.9}{2} = 4.05m \quad (4)$$

예측 오차가 접안 과정에서 계류장과 충돌 위험을 유발할 수 있는 수준인지 평가하기 위해 단거리 구간에서의 GT와 예측 결과를 선체 고정 좌표계에서 상대 위치로 표현한 뒤, 횡방향 및 종방향 위치 오차의 RMS 값을 계산하였다. 계산 결과를 Table 5에 나타내었다.

계산 결과를 C_T, C_L 값과 비교했을 때, 예측 오차의 RMS값이 충분히 작은 수준이라 판단된다. 따라서 단거리 구간에서는 본 연구의 예측 오차의 RMS 값이 계류장 내부 크기 대비 충분한 여유를 제공하는 것으로 생각되므로 무인수상선 접안 시 계류장과 충돌이 발생할 가능성이 낮은 것으로 판단된다.

5. 결 론

본 연구에서는 카메라 이미지 데이터만으로 무인수상선의 접안 과정 중 실시간 상대 위치 추정 기법을 제안하였다. 기존의 카메라 기반 접안 위치 추정 기술은 주로 다중 센서 또는 마커 기반 접근 방식에 의존해왔으나, 마커 기반 방식은 빛 반사에 취약하고 유지·보수에 많은 자원이 소요된다는 한계가 있다. 이에 본 연구에서는 카메라만을 활용한 비마커 방식으로 기존 연구의 단점을 보완하고, 딥러닝 기술을 활용하여 상대 위치 추정을 수행하였다. 특히, 드론 촬영 이미지와 3차원 형상 복원 기술을 활용하여 학습 데이터를 구축함으로써 학습 데이터 취득에 소요되는 시간과 비용을 절감하였으며, 학습 진행 과정에서 추가로 필요한 데이터를 유연하게 생성하여 학습에 활용할 수 있었다.

검증을 위해 선박해양플랜트연구소의 무인수상선인 아라곤 캣 40-02호를 활용하여 수위별로 3차원 계류장 탐지를 수행하였다. 이를 통해 무인수상선과 계류장 간의 상대 거리 및 상대 방위각을 추정하였다. 그 결과, 실제 학습 데이터와 수위가 가장 유사한 저수위 환경에서 상대 거리 RMS는 0.76m, 상대 방위각 RMS는 1.59°가 나왔으며, 이는 다른 수위 조건에 비해 높은 정확도를 보였다. 또한 예측 오차를 분석한 결과, 예측 오차의 RMS 값이 대상 무인수상선과 계류장 내부의 제원을 고려했을 때 충분한 여유를 제공함을 확인하였다. 따라서 제안된 방법이 탐지 관점에서 실용적 타당성이 있다고 판단된다.

본 연구의 한계로는 다양한 기상 조건과 장소에서 상대 위치 추정 성능을 검증하지 못했다는 점이 있으며, 후속 연구에서는 이를 보완하여 알고리즘의 강건성을 확보할 필요가 있다. 또한, 카메라 왜곡으로 인해 영상에 기둥의 휘어짐과 같은 왜곡이 발생할 경우 탐지 성능이 저하될 수 있다. 따라서, 향후 연구에서는 딥러닝 기반 왜곡 보정 알고리즘을 적용하여 실제 환경에서의 다양한 왜곡 조건을 학습하고 이를 실시간으로 보정하여 계류장 탐지 성능을 향상시킬 필요가 있다.

결론적으로, 본 연구에서 제안하는 상대 위치 추정 기법은 무인수상선과 계류장 간 상대 거리 및 상대 방위각을 추정하여 무인수상선이 계류장으로 자율 접안 시 필요한 정보를 제공할 수 있으며, 이는 향후 무인수상선의 자율제어 기술 발전에 기여할 수 있을 것으로 기대된다.

후 기

본 연구는 과학기술정보통신부의 재원으로 한국연구재단, 무인이동체원천기술개발사업단의 지원을 받아 무인이동체원천기술개발사업을 통해 수행되었음(2020M3C1C1A02086423)

References

Chen, Y., Wu, Q., Harandi, M. and Cai, J., 2024. How far can we compress instant-NGP-based NeRF?. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.20321-20330.

Figueiredo, A.B. and Matos, A.C., 2020. MViDO: A high performance monocular vision-based system for docking a hovering AUV. *Applied Sciences*, 10(9), 2991.

Geiger, A., Lenz, P. and Urtasun, R., 2012. Are we ready for autonomous driving? The KITTI Vision Benchmark Suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp.3354-3361.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770-778.

Li, P., Zhao, H., Liu, P. and Cao, F., 2020. RTM3D: Real-Time monocular 3D detection from object keypoints for autonomous driving. *European Conference on Computer Vision*. Springer International Publishing:Cham, pp.644-660.

Lim, Y.E., Kim, N. and Woo, J., 2025a. Real-time three-dimensional ship detection using a monocular camera. *Measurement and Control,* 00202940251355043.

Lim, Y.E., Kim, N. and Woo, J., 2025b. 3D reconstruction of underwater objects using NeRF. *Journal of Advanced Marine Engineering and Technology*, 49(3), pp.187-193.

Liu, S., Ozay, M., Okatani, T., Xu, H., Sun, K. and Lin, Y., 2018. Detection and pose estimation for short-range vision-based underwater docking. *IEEE Access*, 7, pp.2720-2749.

Mentjes, J., Wiards, H. and Feuerstack, S., 2022. Berthing assistant system using reference points. *Journal of Marine Science and Engineering*, 10(3), 385.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R., 2021. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), pp.99-106.

Müller, T., Evans, A., Schied, C. and Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4), 102.

Schönberger, J.L. and Frahm, J.M., 2016. Structure-from-motion revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

(CVPR), pp.4104-4113.

Schönberger, J.L., Zheng, E., Frahm, J.M. and Pollefeys, M., 2016. Pixelwise View Selection for Unstructured Multi-View Stereo. *European conference on computer vision*. Springer International Publishing:Cham, pp.501-518.

Volden, Ø., Stahl, A. and Fossen, T.I., 2022. Vision-based positioning system for auto-docking of unmanned surface vehicles (USVs). *International Journal of Intelligent Robotics and Applications*, 6, pp.86-103.

Woo, J., 2019. A convolutional neural network based localization method for docking of an unmanned surface vehicle. *Journal of Institute of Control, Robotics and Systems*, 25(7), pp.625-632.

Xu, S., Jiang, Y., Li, Y., Wang, B., Xie, T., Li, S., Qi, H., Li, A. and Cao, J., 2023. A stereo visual navigation method for docking autonomous underwater vehicles. *Journal of Field Robotics*, 41(2), pp.374-395.

Authorship Contribution Statement

YeEun Lim: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing; **Chunseon Pyo:** Investigation, Resources; **Kihun Kim:** Funding acquisition, Resources, Supervision, Writing - review & editing; **Hansol Park:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing - review & editing.


임 예 은


표 춘 선


김 기 훈


박 한 솔