

## Article

# Persistent Localization of Autonomous Underwater Vehicles Using Visual Perception of Artificial Landmarks

Jongdae Jung <sup>1,\*</sup> , Hyun-Taek Choi <sup>2</sup>  and Yeongjun Lee <sup>3</sup> <sup>1</sup> Department of Autonomous Vehicle System Engineering, Chungnam National University, Daejeon 34134, Republic of Korea<sup>2</sup> Advanced-Intelligent Ship Research Division, Korea Research Institute of Ships and Ocean Engineering, Daejeon 34103, Republic of Korea; htchoi@kriso.re.kr<sup>3</sup> Ocean and Maritime Digital Technology Research Division, Korea Research Institute of Ships and Ocean Engineering, Daejeon 34103, Republic of Korea; leeyeongjun@kriso.re.kr

\* Correspondence: jdjung@cnu.ac.kr

**Abstract:** Persistent localization is a critical requirement for autonomous underwater vehicles (AUVs) engaged in long-term missions. Conventional dead-reckoning (DR) methods for estimating the position and orientation of AUVs often suffer from drift, necessitating additional information to correct accumulating errors. In this paper, we propose a visual artificial landmarks-based simultaneous localization and mapping (SLAM) system for AUVs. This system utilizes two types of underwater artificial landmarks that are observed using forward and downward-looking cameras. The information obtained from these detected landmarks, along with incremental DR estimates, is integrated within a framework based on the extended Kalman filter (EKF) SLAM approach, allowing for the recursive estimation of both the robot and the landmark states. We implemented the proposed visual SLAM method using our yShark II AUV and conducted experiments in an engineering basin to validate its effectiveness. A ceiling vision-based reference pose acquisition system was installed, facilitating a comparison between the pose estimation results obtained from DR and those derived from the SLAM method.



Academic Editors: Weicheng Cui and Paolo Silvestri

Received: 10 February 2025

Revised: 16 April 2025

Accepted: 18 April 2025

Published: 22 April 2025

**Citation:** Jung, J.; Choi, H.-T.; Lee, Y. Persistent Localization of Autonomous Underwater Vehicles Using Visual Perception of Artificial Landmarks. *J. Mar. Sci. Eng.* **2025**, *13*, 828. <https://doi.org/10.3390/jmse13050828>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** autonomous underwater vehicles; simultaneous localization and mapping; artificial landmarks; optical cameras

## 1. Introduction

Precise localization of autonomous underwater vehicles (AUVs) in an underwater environment is challenging but essential for the successful accomplishment of various tasks assigned to AUVs. Typical sensor modalities used in AUV localization include the absolute heading reference system (AHRS), the Doppler velocity log (DVL), and depth sensors [1,2]. With these sensors, the AUV can estimate its pose state through the dead reckoning (DR); where vehicle kinematic or dynamic models are typically exploited [3]. However, localization with DR suffers from accumulating errors because it integrates derivative measurements that contain random and systematic noise [4]. In particular, heading estimation by integrating gyroscope measurements is inherently prone to drift errors, which become more severe in nonlinear motions, such as rotations with varying angular velocities. This accumulation of drift can lead to significant deviations between the estimated trajectory and the actual path. To ensure the successful long-term operation of AUVs, it is essential to mitigate this error.

To improve positioning accuracy, additional aids must be integrated with the vehicles. A common approach involves using exterior sensors such as ultrashort baseline (USBL) [5,6] or long baseline (LBL) [7,8] systems. However, the USBL method requires support from the mother ship for the USBL transceiver, and there is an acoustic delay in the position information received. In contrast, the LBL method requires the placement of acoustic beacons on the seafloor, which can limit coverage to fixed areas. Therefore, these approaches are not adequate for the long-term operations of AUVs in deep-sea environments.

AUVs are typically equipped with one or more optical cameras to obtain visual information about underwater environments. By integrating an active lighting system, visual observations can also be employed to determine the pose information of the AUV. In our previous studies [9–11], we validated that observing artificial landmarks can improve the accuracy of the AUV localization. However, the assumption that the location of the landmarks should be provided prior to the vehicle deployment limits the generality of the method. The AUV should be capable of creating a map from scratch and identifying the locations of various landmarks, while simultaneously using this information to correct its position. This process is known as simultaneous localization and mapping (SLAM), a well-defined topic extensively studied in robotics [12–14] and successfully applied in AUV navigation [15–17]. The extended Kalman filter (EKF), unscented Kalman filter (UKF) [18], particle filter [19], and graph optimization [20] are commonly used methods for addressing SLAM (Simultaneous Localization and Mapping) problems. Among these, the EKF has the advantage of being computationally less demanding, especially when dealing with a finite number of landmarks, as is the case in our study. In this paper, we present an EKF-based SLAM method that incorporates visual measurements from three types of artificial landmarks captured by forward- and downward-looking cameras. We designed our SLAM framework to incorporate the observations from each landmark independently, along with a kinematics-based DR estimate. The highlights of this paper are as follows.

- We proposed a visual navigation system that uses artificial landmarks detected by forward- and downward-looking cameras.
- Visual information was integrated into an EKF-based SLAM framework.
- Experiments conducted with our AUV in an engineering basin demonstrated improved pose estimation compared to dead reckoning methods.
- Ground truth pose data were obtained using a ceiling-mounted vision-based reference system.

This paper expands on the conference paper previously presented [21].

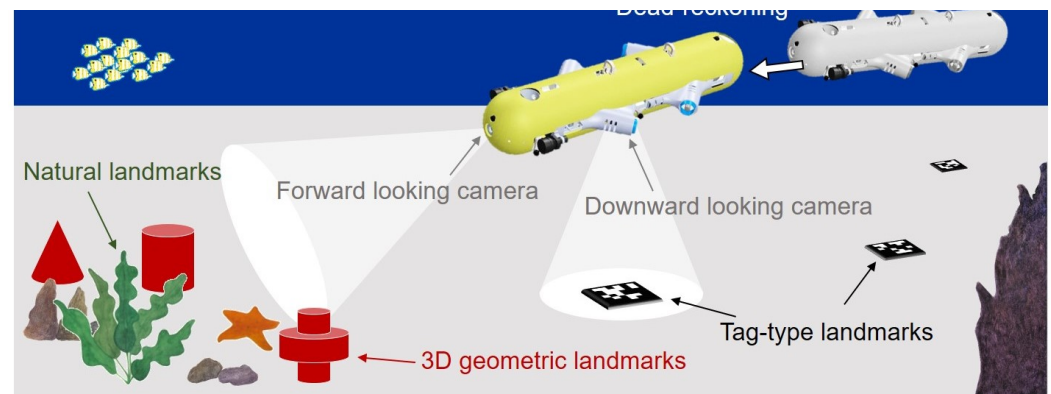
The remainder of this paper is organized as follows. The design of artificial landmarks is presented in Section 2. In Section 3, we present the formulation of the proposed visual SLAM, followed by the experimental setup and results in Section 4. Finally, the conclusions are provided in Section 5.

## 2. Design of Landmarks

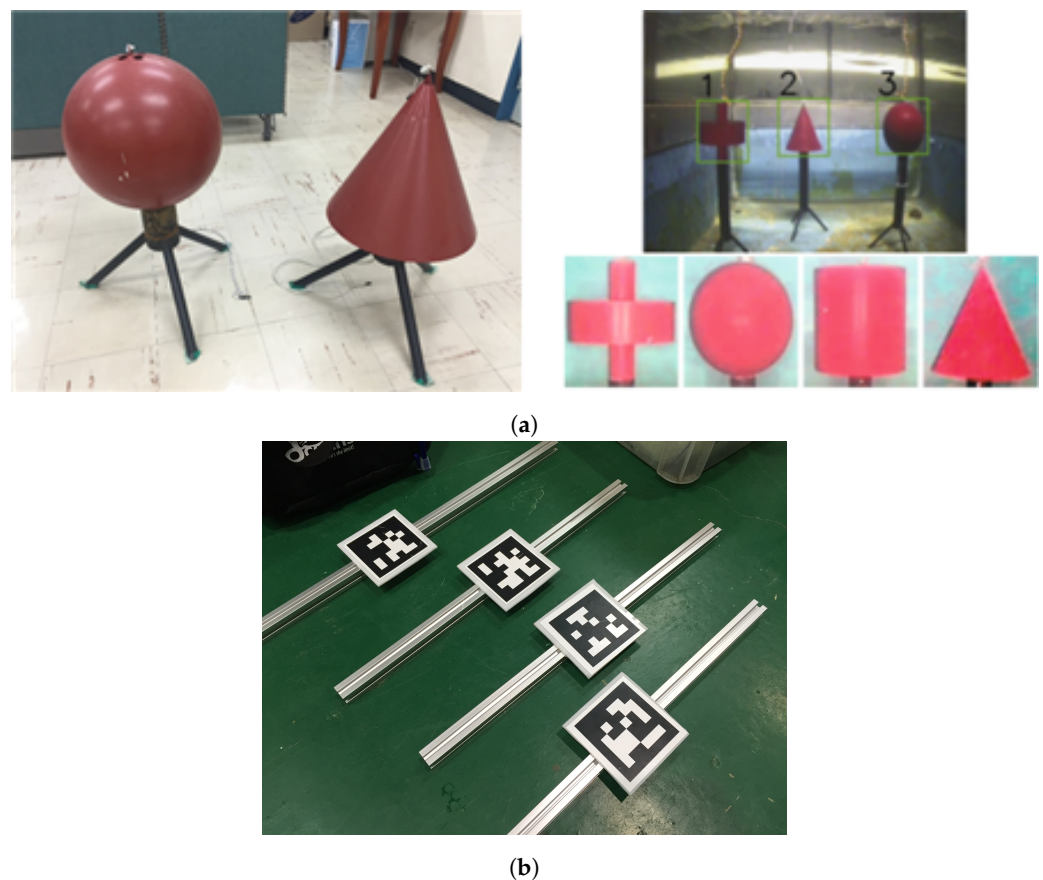
We designed three types of artificial landmarks, as shown in Figure 1. The first type is cylindrical landmarks. This one is designed for viewpoint invariant observations, as shown in Figure 2a. When the robot and landmarks are in a similar depth, basic geometric shapes (circle, triangle, etc.) can be seen in any direction from the forward-looking camera. Supervised learning with the random forest was used in the detection of the landmarks. Underwater images of these landmarks under various lighting conditions were used for the training of the ensemble of trees [9]. When the landmarks are detected in the current forward-looking image, conventional template matching with adaptive thresholding is performed to classify the shape of the landmarks, which outputs identification (ID) of the landmark.

The second one is tag-type landmarks, as shown in Figure 2b. This one is detected by the down-looking camera. A predefined 2D bar code is printed on these landmarks, allowing a unique ID for each tag. In this paper, AprilTag [22] was used. The advantage of this landmark is that we can obtain a 6 degree-of-freedom (DOF) pose of the camera from a single image. For underwater usage, the tags were coated with water-proof film.

As a third type of landmark, we installed steel panels cluttered with various objects such as rope, plastic pipe, small buoy, etc. In order to detect visual features of these objects, SURF [23] was used in this study. Any other feature extraction algorithms [24–29] can be used without the loss of generality.



**Figure 1.** Underwater navigation scenario.



**Figure 2.** Artificial landmarks for AUV navigation. (a) Cylindrical landmarks for forward sensing. (b) Fiducial markers for downward sensing.

### 3. SLAM Framework

Figure 3 illustrates the overall framework of the proposed visual SLAM system. We have modules for AUV dead-reckoning, image processing, data association, and the SLAM core. In this work, EKF-SLAM [12] is utilized as the SLAM core. Although we have chosen the EKF, other SLAM methods could also be applied without loss of generality, as the same motion and observation models can be utilized. For instance, when implementing graph optimization-based SLAM, information from the dead-reckoning and image processing blocks can create constraints in the front end, while optimization in the back end can update the entire trajectories.

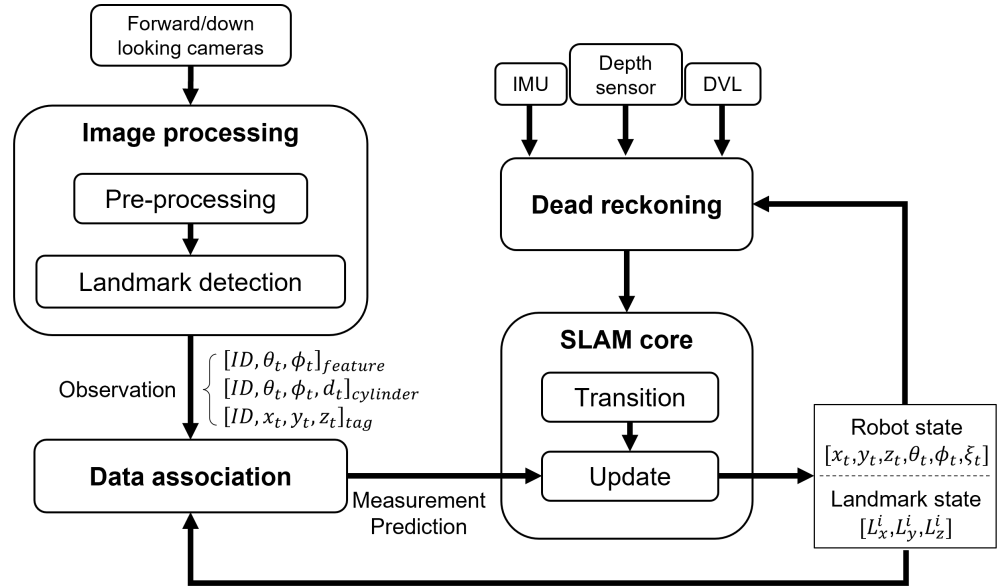


Figure 3. Visual SLAM framework.

We first define the state variables related to the localization and mapping problems. The robot state  $\{\mathbf{x}_t\}$  is defined as  $\text{SE}(3)$  pose as

$$\mathbf{x}_t = \begin{bmatrix} x, y, z, q_x, q_y, q_z, q_w \end{bmatrix}^\top, \quad t = 0, 1, \dots \quad (1)$$

where  $\mathbf{q} = [q_x, q_y, q_z, q_w]^\top$  is a quaternion that represents the vehicle's orientation. The map is then defined as a vector describing the state of each landmarks as follows:

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}^{\text{tag}} \\ \mathbf{m}^{\text{cyl}} \\ \mathbf{m}^{\text{obj}} \end{bmatrix} \quad (2)$$

$$= \begin{bmatrix} \mathbf{m}_1^{\text{tag}\top}, \dots, \mathbf{m}_{N_t}^{\text{tag}\top} \mid \mathbf{m}_1^{\text{cyl}\top}, \dots, \mathbf{m}_{N_c}^{\text{cyl}\top} \mid \mathbf{m}_1^{\text{obj}\top}, \dots, \mathbf{m}_{N_f}^{\text{obj}\top} \end{bmatrix}^\top \quad (3)$$

where  $\mathbf{m}_i^{\text{tag}} \in \text{SE}(3)$ ,  $\mathbf{m}_i^{\text{cyl}} \in \mathbb{R}^3$ , and  $\mathbf{m}_i^{\text{obj}} \in \mathbb{R}^3$  denote the state of the tag, cylindrical, and feature landmarks, respectively, and  $N_t$ ,  $N_c$ , and  $N_f$  denote the total number of the corresponding landmarks registered in a state vector at time step  $t$ . We assume that the landmarks are static during the AUV's operation. The measurement  $\mathbf{z}_t$  then can be defined as follows:



$$\mathbf{z}_t = \begin{bmatrix} \mathbf{z}_t^{\text{tag}} \\ \mathbf{z}_t^{\text{cyl}} \\ \mathbf{z}_t^{\text{obj}} \end{bmatrix} \quad (4)$$

$$\mathbf{z}_t^{\text{tag}} = \begin{bmatrix} ID, x, y, z, q_x, q_y, q_z, q_w \end{bmatrix}^\top \quad (5)$$

$$\mathbf{z}_t^{\text{cyl}} = \begin{bmatrix} ID, d, \theta \end{bmatrix}^\top \quad (6)$$

$$\mathbf{z}_t^{\text{obj}} = \begin{bmatrix} ID, \theta \end{bmatrix}^\top \quad (7)$$

where  $\mathbf{z}_t^{\text{tag}}$ ,  $\mathbf{z}_t^{\text{cyl}}$ , and  $\mathbf{z}_t^{\text{obj}}$  denote the measurement for the tag, cylindrical, and feature landmarks, respectively, and  $\theta$  denotes the bearing of the detected landmark.  $d$  is the distance from the forward camera to the cylindrical landmark, which is calculated as

$$d = f \sqrt{\frac{A_c}{A_p}} \quad (8)$$

where  $f$  denotes the focal length of the forward looking camera,  $A_c$  is the actual cross section area of the observed cylindrical landmark, and  $A_p$  is the pixel area of the landmark in the camera image. Finally, the augmented state variable  $\mathbf{y}_t$  is defined as follows.

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{m}_t \end{bmatrix}. \quad (9)$$

### 3.1. SLAM Formulation

In the EKF SLAM framework, the SLAM procedure consists of two main steps: prediction and measurement update. These steps are executed recursively after the state initialization.

#### 3.1.1. Prediction

In the prediction step, the state of the previous timestep is transformed into the next timestep using the state transition model  $\mathbf{f}(\mathbf{y}_{t-1}, \mathbf{u}_t)$ . We employed the AUV's kinematics model to describe the transition as follows.

$$\begin{aligned} \bar{\mathbf{y}}_t &= \mathbf{f}(\mathbf{y}_{t-1}, \mathbf{u}_t) \\ &= \begin{bmatrix} x_{t-1} + (v_{x,t}^{\text{DVL}} \cos \psi_t - v_{y,t}^{\text{DVL}} \sin \psi_t) \Delta t \\ y_{t-1} + (v_{x,t}^{\text{DVL}} \sin \psi_t + v_{y,t}^{\text{DVL}} \cos \psi_t) \Delta t \\ z_t^{\text{d}} \\ q_{x,t}^{\text{AHRS}} \\ q_{y,t}^{\text{AHRS}} \\ q_{z,t}^{\text{AHRS}} \\ q_{w,t}^{\text{AHRS}} \\ \mathbf{m}_{t-1} \end{bmatrix} \end{aligned} \quad (10)$$

where  $\mathbf{u}_t = [v_{x,t}^{\text{DVL}}, v_{y,t}^{\text{DVL}}, z_t^{\text{d}}, q_{x,t}^{\text{AHRS}}, q_{y,t}^{\text{AHRS}}, q_{z,t}^{\text{AHRS}}, q_{w,t}^{\text{AHRS}}]^\top$  denotes the control input from the DVL, depth sensor, and AHRS with a noise covariance  $\mathbf{Q}_t$ ,  $\Delta t$  denotes the time interval between the current and previous time step, and  $\psi_t$  denotes the heading angle of the AUV. In (10), only the robot state undergoes the state transition and thus is updated. The

map states remain unchanged in the prediction step. The error state covariance  $\mathbf{P}_t$  is also updated in the prediction step as follows.

$$\bar{\mathbf{P}}_t = \mathbf{F}_t \mathbf{P}_{t-1} \mathbf{F}_t^T + \mathbf{Q}_t \quad (11)$$

where  $\mathbf{F}_t$  is the Jacobian of the state transition model.

### 3.1.2. Measurement Update

In the measurement update step, both the robot state and the map states are updated according to the measurement model  $\mathbf{h}(\mathbf{y}_t)$ . As explained in the (5)–(7), the measurement  $\mathbf{z}_t$  is represented in a sensor-fixed frame. The predicted measurement  $\hat{\mathbf{z}}_t$  can be modeled as follows.

$$\hat{\mathbf{z}}_t = \mathbf{h}(\bar{\mathbf{y}}_t) \quad (12)$$

$$= \begin{bmatrix} \hat{\mathbf{z}}_t^{\text{tag}} \\ \hat{\mathbf{z}}_t^{\text{cyl}} \\ \hat{\mathbf{z}}_t^{\text{obj}} \end{bmatrix} \quad (13)$$

$$\hat{\mathbf{z}}_t^{\text{tag}} = \mathbf{h}(\bar{\mathbf{x}}_t, \bar{\mathbf{m}}_t^{\text{tag}}) \quad (14)$$

$$\hat{\mathbf{z}}_t^{\text{cyl}} = \mathbf{h}(\bar{\mathbf{x}}_t, \bar{\mathbf{m}}_t^{\text{cyl}}) \quad (15)$$

$$\hat{\mathbf{z}}_t^{\text{obj}} = \mathbf{h}(\bar{\mathbf{x}}_t, \bar{\mathbf{m}}_t^{\text{obj}}) \quad (16)$$

Update of the vehicle and map states are performed by the Kalman gain calculated from the Jacobian  $\mathbf{H}_m$  of the observation models (14)–(16) as follows.

$$\begin{aligned} \mathbf{H}_m &= \frac{\partial \mathbf{h}(\mathbf{y}_t)}{\partial \mathbf{m}} \\ &= \begin{bmatrix} \mathbf{H}_m^{\text{tag}} & \mathbf{H}_m^{\text{cyl}} & \mathbf{H}_m^{\text{obj}} \end{bmatrix} \end{aligned} \quad (17)$$

where  $\mathbf{H}_m^{\text{tag}}$ ,  $\mathbf{H}_m^{\text{cyl}}$ , and  $\mathbf{H}_m^{\text{obj}}$  denote the Jacobian matrix corresponding to the each landmark state. The Kalman update is performed as follows:

$$\mathbf{S}_t = \mathbf{H}_m \bar{\mathbf{P}}_t \mathbf{H}_m^T + \mathbf{R}_t \quad (18)$$

$$\mathbf{K}_t = \bar{\mathbf{P}}_t \mathbf{H}_m^T \mathbf{S}_t^{-1} \quad (19)$$

$$\mathbf{y}_t = \bar{\mathbf{y}}_t + \mathbf{K}_t (\mathbf{z}_t - \hat{\mathbf{z}}_t) \quad (20)$$

$$\mathbf{P}_t = \bar{\mathbf{P}}_t - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T \quad (21)$$

where  $\mathbf{R}_t$  denotes the sensor noise covariance,  $\mathbf{S}_t$  denotes the innovation covariance, and  $\mathbf{K}_t$  denotes the Kalman gain.

## 4. Experiments

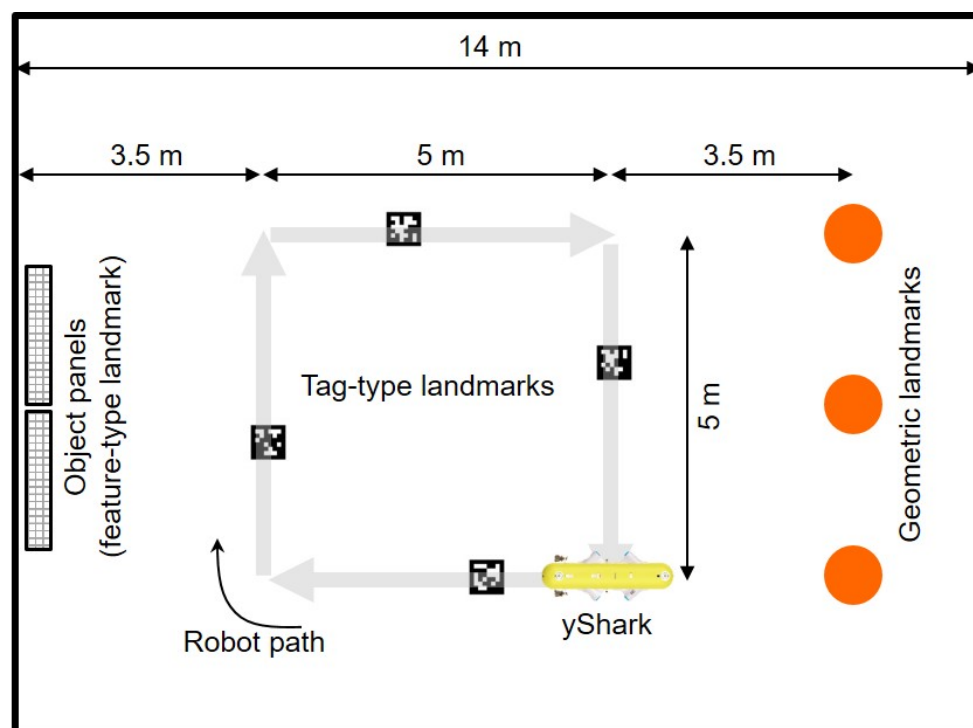
### 4.1. Experimental Setup

Experiments were conducted in a basin environment to validate the proposed SLAM method. Figure 4 illustrates the setup of the landmarks and path plans for the robot during the tests. We installed 4 tag-type landmarks and 3 cylindrical landmarks within the basin. Additionally, we placed two panels with various objects attached to serve as visual features. The yShark II, an AUV developed by the Korea Research Institute of Ships and Ocean Engineering (KRISO), was utilized in this experiment. The robot moved along the square (5 m by 5 m) path 4 times. Tag-type landmarks were observed from a down-looking camera (PointGrey Grasshopper3), and object panels and cylinder-type landmarks were observed

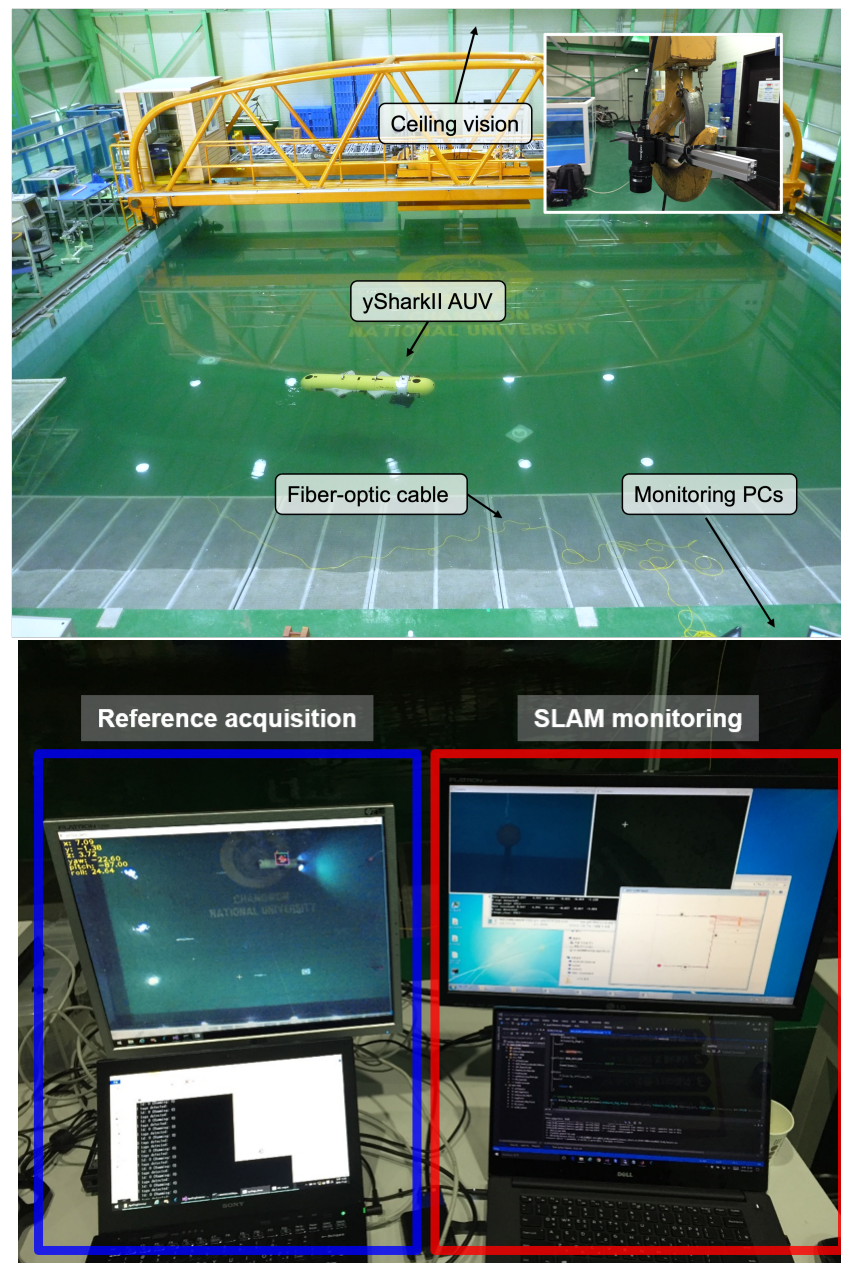
from a forward-looking camera with the same model. Navigation sensors including the AHRS (Microstrain 3DM-GX3-25), DVL (LinkQuest NavQuest600 Micro), and depth sensor (Paroscientific Digiquartz) were used in the calculation of the DR estimation and SLAM prediction. All software for data acquisition, landmark detection, DR, SLAM, and vehicle control was developed in-house in C++.

Figure 5 illustrates our testing setup, located in an engineering basin at Changwon National University.

We implemented a ceiling-mounted vision-based reference acquisition system (VRS) equipped with a wide field of view ( $f = 2.8$  mm) camera and attached one large AprilTag on the AUV's back. In the ground station, remote connections to the AUV's internal PC and the VRS were set through the fiber-optic communication and power-over-ethernet, respectively. The DR and SLAM algorithms were run online and all the sensor data were also saved for the offline evaluations. In the offline evaluation, To generate extreme navigation scenarios, we added synthetic random noises to the DVL and AHRS measurements during the 3rd and 4th loops. Gaussian noise with standard deviation  $\sigma = [0.5 \text{ m}, 10^\circ]$  were periodically (once per 100 visual measurements) added to the translation and rotation states of the vehicle, respectively. Total 5 offline datasets were generated in this way.



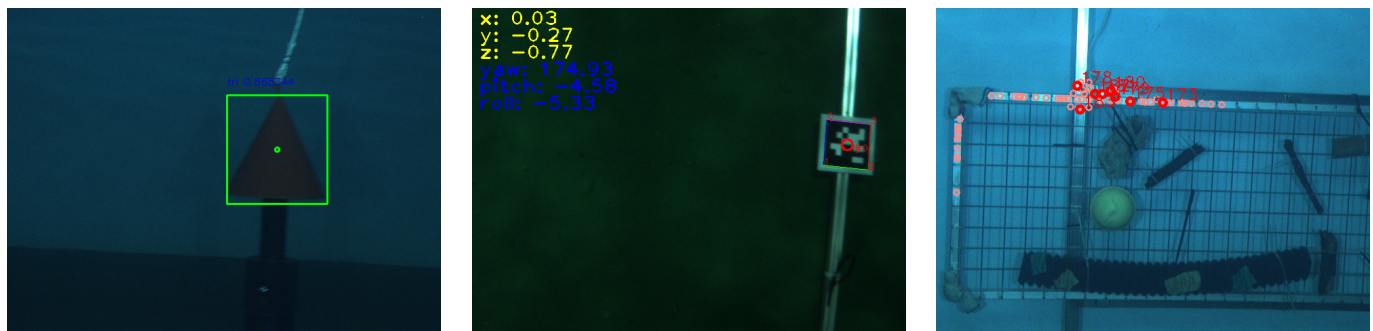
**Figure 4.** Landmark setup and robot path plans.



**Figure 5.** Testing setup. Top: Engineering basin located at Changwon National University with the yShark II AUV. Bottom: ground control station featuring a reference acquisition system and remote access to the AUV's internal computer.

#### 4.2. Results

Examples of the landmark detection results are shown in Figure 6. For cylindrical landmarks, the detected bounding box and its center position are used to construct the measurement vector;  $A_p$  in (8) is calculated as the area of the bounding box; the bearing  $\theta$  in (6) is calculated with the center position. All three types of cylindrical landmarks were successfully detected in our test. Related to the tag landmarks, only 3 to 5 subsequent frames contained the tag images when the AUV moved over the landmark; however, this was sufficient for the update and convergence of the vehicle and map states. Lastly, feature-based landmarks showed that the extracted visual features were dominant around the frame structures, rather than the small objects. This is due to the reflection of the steel frame by the AUV's lighting system.



**Figure 6.** Underwater artificial landmarks and their detection results. From left to right: cylindrical landmarks, tag-type landmarks, and feature-based landmarks.

Localization results from one of the noisy datasets are shown in Figure 7. During the first two loops, the SLAM registered every new landmark detected from the forward and down-looking cameras. Because the error covariance increases during the prediction step (11), the covariance ellipse of the first detected landmark (tag-type) had the smallest value; and this contributed to the significant update whenever a large loop-closure occurred. The covariance for the cylindrical landmarks had an elliptical shape at first (Figure 7b), but converged to the circular shape after a few observations (Figure 7c), as the depth estimation described in (8) are available for this type of landmarks. In contrast, the covariance ellipse for the feature landmarks always had a narrow shape as the bearing information was only available. As the DR also has a good short-term performance, its trajectory is similar to the SLAM results until the second loop (Figure 7b). However, when synthetic noise was added from the 3rd loop, the DR estimate started to diverge with accumulating errors (Figure 7c); while the SLAM trajectory recovered from the erroneous state owing to the information from artificial landmarks.

Results for all 5 datasets are shown in Figure 8. We applied different random noises for each dataset. Estimated trajectories and the absolute trajectory errors in Figure 8f support the superiority of the proposed SLAM method, which maintained the bounded errors throughout the entire trajectories. Table 1 summarizes absolute trajectory errors for DR and SLAM. The RMSE of the SLAM trajectory was 0.67 m for 92 m paths. This improved the positioning error of DR by 41.2%. Similarly, the maximum error was reduced by 43.6% when we used the SLAM.

**Table 1.** Summary of Absolute Trajectory Errors.

Dataset	RMSE (m)		Max. Error (m)	
	DR-Only	SLAM	DR-Only	SLAM
1	1.26	0.68	2.52	1.21
2	1.11	0.66	2.22	1.30
3	1.02	0.67	2.04	1.26
4	1.23	0.64	2.04	1.07
5	1.07	0.69	1.85	1.19
Mean	1.14	0.67	2.13	1.20

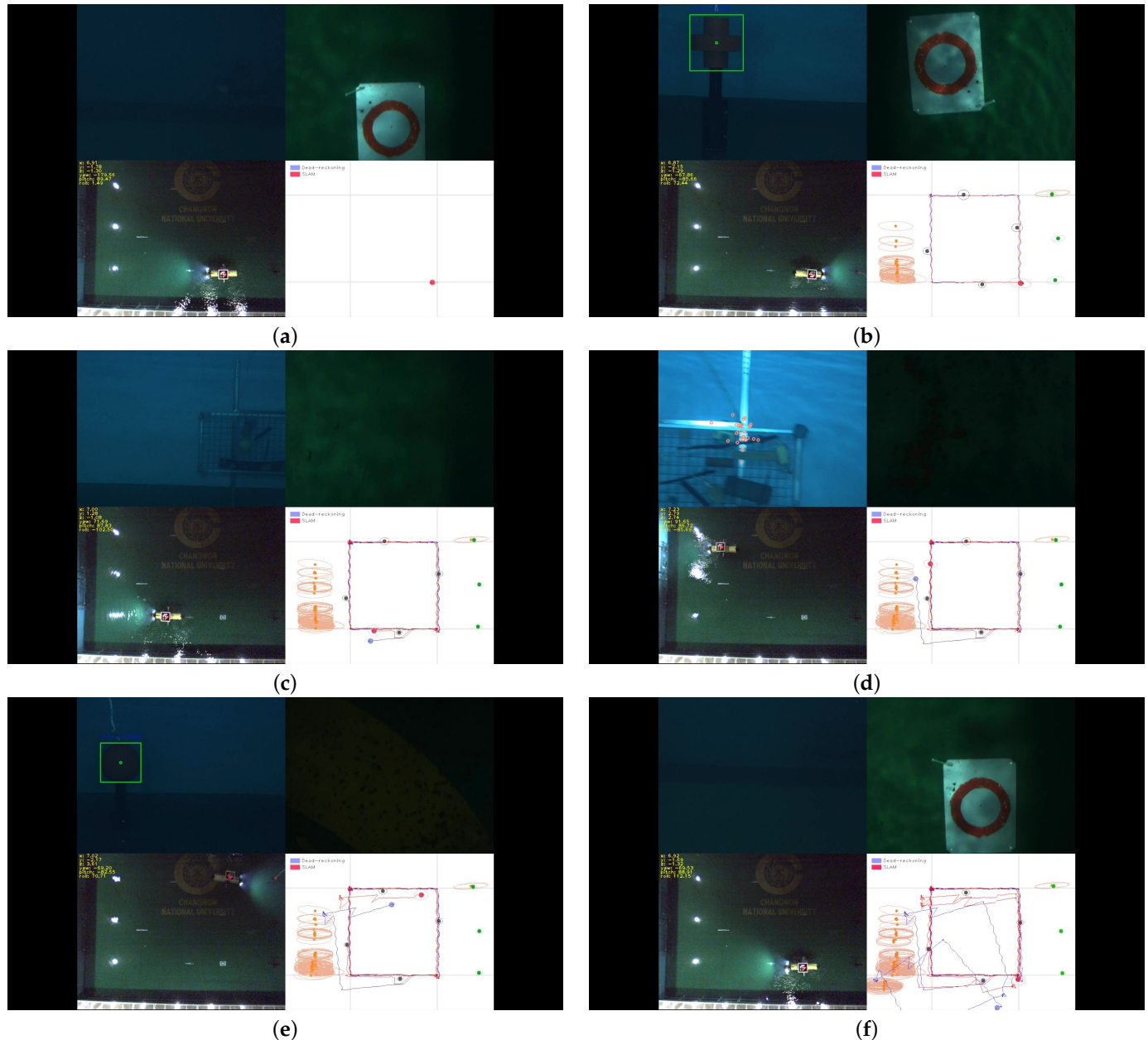
#### 4.3. Discussions

We can verify how each type of landmark contributes to the reduction of position error by examining the ‘jumps’ in the AUV trajectories. This analysis is possible because the layout of the landmarks in the experimental setup typically allows for the detection of one type of landmark at a time. It is clear that tag-type landmarks contribute significantly to pose correction, as illustrated in Figure 7c. The influence of a cylindrical landmark is

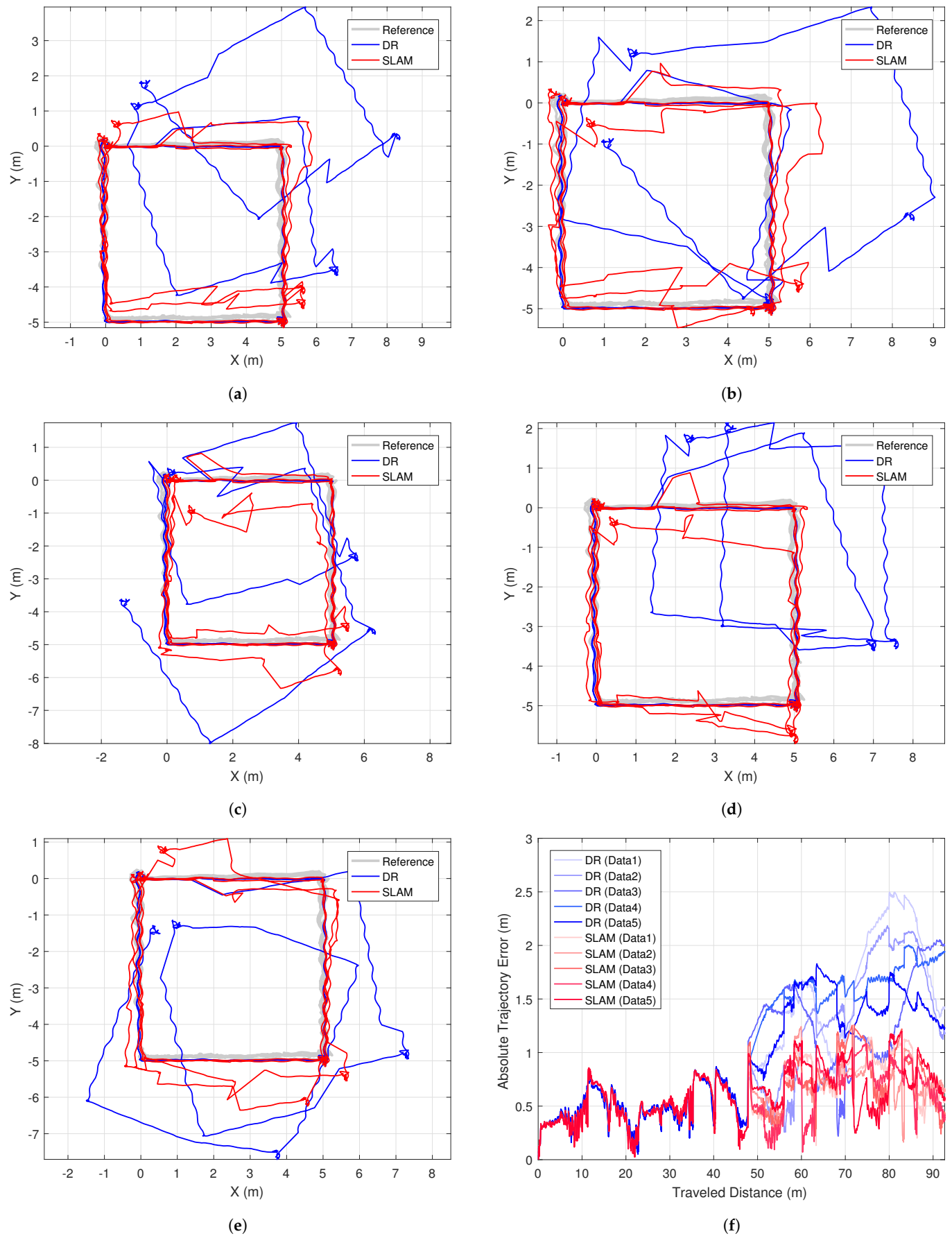


also evident in Figure 7e, where the green circle in the lower right subplot indicates the detected landmark, causing the SLAM pose to jump onto the correct path. The contribution of object-clustered landmarks is harder to explain explicitly; however, we can infer their impact from the fact that the SLAM paths near the object panels exhibit the smallest errors in all cases, as shown in Figure 8.

While the experimental results are promising, there are limitations to consider. The validation was conducted in a controlled basin environment, meaning that the visibility of landmarks may be more challenging in real-world field conditions. To address this issue, deep learning-based image enhancement techniques [30] or multi-modal markers [31] that can be detected by both optical and sonar cameras could be utilized.



**Figure 7.** Estimated robot paths (solid lines in blue: DR, red: SLAM) and mapped landmarks (circles in green: geometric, gray: tag-type, orange: feature-type). Subfigures (a–f) represent sequential snapshots of the results during the AUV's trajectory following.



**Figure 8.** Resulting AUV trajectories (a–e) and summary of estimation errors (f) from five different random noise datasets.

## 5. Conclusions

In this study, visual SLAM for AUVs using artificial landmarks has been investigated. Three types of artificial landmarks were proposed to be used with forward and down-looking cameras available in our AUV. We implemented a conventional EKF-based SLAM framework incorporating a kinematic model of the AUV and landmark observation models. We obtained experimental data from the tests conducted in a basin and validated the performance of the proposed method for the persistent localization scenario. We expect our method to be applied in the area of long-term AUV operations, such as valve-turning in deep-sea platforms or manipulation of nuclear waste in nuclear power plants.

**Author Contributions:** Conceptualization, J.J. and H.-T.C.; methodology, J.J. and Y.L.; software, J.J., Y.L. and H.-T.C.; validation, J.J., Y.L. and H.-T.C.; formal analysis, J.J.; investigation, J.J. and Y.L.; resources, H.-T.C.; data curation, J.J. and Y.L.; writing—original draft preparation, J.J.; writing—review and editing, J.J.; visualization, J.J.; supervision, J.J. and H.-T.C.; project administration, J.J. and H.-T.C.; funding acquisition, J.J. and H.-T.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by research fund of Chungnam National University.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Paull, L.; Saeedi, S.; Seto, M.; Li, H. AUV Navigation and Localization: A Review. *IEEE J. Ocean. Eng.* **2014**, *39*, 131–149. [\[CrossRef\]](#)
2. Zhang, B.; Ji, D.; Liu, S.; Zhu, X.; Xu, W. Autonomous Underwater Vehicle Navigation: A review. *Ocean Eng.* **2023**, *273*, 113861. [\[CrossRef\]](#)
3. Esteba, J.; Cieślak, P.; Palomeras, N.; Ridao, P. Sparus Docking Station: A current aware docking station system for a non-holonomic AUV. *J. Field Robot.* **2024**, *41*, 1765–1779. [\[CrossRef\]](#)
4. González-García, J.; Gómez-Espinosa, A.; Cuan-Urquiza, E.; García-Valdovinos, L.G.; Salgado-Jiménez, T.; Cabello, J.A.E. Autonomous Underwater Vehicles: Localization, Navigation, and Communication for Collaborative Missions. *Appl. Sci.* **2020**, *10*, 1256. [\[CrossRef\]](#)
5. Mandt, M.; Gade, K.; Jalving, B. Integrating DGPS-USBL position measurements with inertial navigation in the HUGIN 3000 AUV. In Proceedings of the 8th Saint Petersburg International Conference on Integrated Navigation Systems, Saint Petersburg, Russia, 28–30 May 2001; pp. 28–30.
6. Xu, J.; Xie, G.; Wang, X.; Ding, Y.; Zhang, S. USV-AUV Collaboration Framework for Underwater Tasks under Extreme Sea Conditions. In Proceedings of the ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 6–11 April 2025; pp. 1–5. [\[CrossRef\]](#)
7. Kinsey, J.C.; Whitcomb, L.L. Preliminary field experience with the DVLNAV integrated navigation system for oceanographic submersibles. *Control. Eng. Pract.* **2004**, *12*, 1541–1549. [\[CrossRef\]](#)
8. Song, J.; Li, W.; Zhu, X.; Dai, Z.; Ran, C. Underwater Adaptive Height-Constraint Algorithm Based on SINS/LBL Tightly Coupled. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–9. [\[CrossRef\]](#)
9. Lee, D.; Kim, G.; Kim, D.; Myung, H.; Choi, H.T. Vision-based object detection and tracking for autonomous navigation of underwater robots. *Ocean Eng.* **2012**, *48*, 59–68. [\[CrossRef\]](#)
10. Kim, D.; Lee, D.; Myung, H.; Choi, H.T. Artificial landmark-based underwater localization for AUVs using weighted template matching. *Intell. Serv. Robot.* **2014**, *7*, 175–184. [\[CrossRef\]](#)
11. Lee, Y.; Choi, J.; Choi, H.T. Experimental results on EKF-based underwater localization algorithm using artificial landmark and imaging sonar. In Proceedings of the 2014 Oceans—St. John's, St. John's, NL, Canada, 14–19 September 2014; pp. 1–3. [\[CrossRef\]](#)
12. Thrun, S.; Burgard, W.; Fox, D. *Probabilistic Robotics*; MIT Press: Cambridge, MA, USA, 2005.
13. Fernández-Madrigal, J.A.; Claraco, J.L.B. *Simultaneous Localization and Mapping for Mobile Robots: Introduction and Methods*; IGI Global: Hershey, PA, USA, 2013.
14. Macario Barros, A.; Michel, M.; Moline, Y.; Corre, G.; Carrel, F. A Comprehensive Survey of Visual SLAM Algorithms. *Robotics* **2022**, *11*, 24. [\[CrossRef\]](#)

15. Williams, S.B.; Newman, P.; Dissanayake, G.; Durrant-Whyte, H. Autonomous underwater simultaneous localisation and map building. In Proceedings of the Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065), San Francisco, CA, USA, 24–28 April 2000; Volume 2, pp. 1793–1798.
16. Roman, C.; Singh, H. A self-consistent bathymetric mapping algorithm. *J. Field Robot.* **2007**, *24*, 23–50. [[CrossRef](#)]
17. Ribas, D.; Ridao, P.; Tardós, J.D.; Neira, J. Underwater SLAM in man-made structured environments. *J. Field Robot.* **2008**, *25*, 898–921. [[CrossRef](#)]
18. Luo, J.; Ko, H.L. UKF-Based Inverted Ultra-Short Baseline SLAM With Current Compensation. *IEEE Access* **2022**, *10*, 67329–67337. [[CrossRef](#)]
19. Cui, J.; Feng, D.; Li, Y.; Tian, Q. Research on simultaneous localization and mapping for AUV by an improved method: Variance reduction FastSLAM with simulated annealing. *Def. Technol.* **2020**, *16*, 651–661. [[CrossRef](#)]
20. Ling, Y.; Li, Y.; Ma, T.; Cong, Z.; Xu, S.; Li, Z. Active Bathymetric SLAM for autonomous underwater exploration. *Appl. Ocean. Res.* **2023**, *130*, 103439. [[CrossRef](#)]
21. Jung, J.; Lee, Y.; Kim, D.; Lee, D.; Myung, H.; Choi, H.T. AUV SLAM using forward/downward looking cameras and artificial landmarks. In Proceedings of the 2017 IEEE Underwater Technology (UT), Busan, Korea, 21–24 February 2017; pp. 1–3. [[CrossRef](#)]
22. Olson, E. AprilTag: A robust and flexible visual fiducial system. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3400–3407. [[CrossRef](#)]
23. Bay, H.; Tuytelaars, T.; Gool, L.V. Surf: Speeded up robust features. In Proceedings of the European conference on Computer Vision, Graz, Austria, 7–13 May 2006; pp. 404–417.
24. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
25. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
26. Leutenegger, S.; Chli, M.; Siegwart, R.Y. BRISK: Binary robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.
27. Sun, J.; Shen, Z.; Wang, Y.; Bao, H.; Zhou, X. LoFTR: Detector-Free Local Feature Matching With Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 8922–8931.
28. Luo, Z.; Zhou, L.; Bai, X.; Chen, H.; Zhang, J.; Yao, Y.; Li, S.; Fang, T.; Quan, L. ASLFeat: Learning Local Features of Accurate Shape and Localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
29. Sarlin, P.E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching With Graph Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
30. Yang, G.; Kang, G.; Lee, J.; Cho, Y. Joint-ID: Transformer-Based Joint Image Enhancement and Depth Estimation for Underwater Environments. *IEEE Sensors J.* **2024**, *24*, 3113–3122. [[CrossRef](#)]
31. Lee, Y.; Choi, J.; Jung, J.; Kim, T.; Choi, H.T. Underwater robot exploration and identification using dual imaging sonar: Basin test. In Proceedings of the 2017 IEEE Underwater Technology (UT), Busan, Korea, 21–24 February 2017; pp. 1–4. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.