IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 16, 2022, accepted May 27, 2022, date of publication May 30, 2022, date of current version June 7, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3179430

Robust Underwater Localization Using Acoustic Image Alignment for Autonomous Intervention Systems

JISUNG PARK AND JINWHAN KIM
Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
Corresponding author: Jinwhan Kim (jinwhan@kaist.ac.kr)

This research was supported by the "Development of autonomous ship technology (20200615)" funded by the Ministry of Oceans and Fisheries (MOF, Korea), and in part supported by the KAIST Institute (N11220120).

ABSTRACT Acoustic imaging sonar can be used as important navigation sensors for underwater intervention systems as they consistently provide spatial information about the surrounding environment, even in limited visibility conditions. However, acoustic imaging sonars are known for high spatial ambiguity and low resolution of their measurements, which makes it challenging to obtain precise navigational information. This paper presents a novel localization method based on sensor fusion using an integrated IMU-DVL system and an acoustic imaging sonar. A sonar simulator is implemented and used to estimate the pose of the robot based on a single acoustic image, and it is formulated as an image alignment problem between a simulated acoustic image and an actual acoustic image. For this, an approximate nearest neighbor search method is employed for initial pose estimation, and a newly developed acoustic image alignment method is applied to obtain more accurate results in a continuous pose space. These methods are then combined to construct an integrated localization system using all the navigation sensors on the robot. The feasibility and utility of the proposed approach is shown through an experimental validation in a test tank.

INDEX TERMS Autonomous intervention system, acoustic imaging sonar, underwater localization.

I. INTRODUCTION

Underwater vehicles, such as remotely operated vehicles (ROVs), are widely used for various intervention tasks, such as subsea construction and inspection (Fig. 1), and these vehicles are referred to as underwater intervention systems. Until recently, the operation of intervention systems was primarily performed by human operators, which has caused various operational problems owing to human errors. Therefore, studies have been conducted on building autonomous intervention systems to increase the efficiency of intervention tasks [1]–[9].

Among the various techniques for autonomous systems, localization is the basic technology for the autonomy of a mobile system, and a reliable autonomous system can be built with the knowledge of the mobile system's characteristics and operating environment. As for intervention systems, an operational environment can be considered a place where man-made objects, such as subsea structures, are intentionally placed by humans. Here, the intervention tasks often require a vehicle to maneuver near the subsea structures and use robotic arms to manipulate them. Accordingly, the relative position between the vehicle and structures is important information and should be accurately given. Considering these operational characteristics, the localization system for

FIGURE 1. Intervention tasks on offshore platforms: The oceaneering NEXXUS ROV performs intervention task on subsea platform.

The associate editor coordinating the review of this manuscript and approving it for publication was Kegen Yu.

the intervention system should have a relative localization capability for the structures, which will be possible through environmental perception using perception sensors.

Underwater vehicles, including intervention systems, are equipped with several perceptions and navigation sensors for their tasks. Among perception sensors, optical cameras and underwater sonars are typically mounted on underwater vehicles, and these sensors provide a significant amount of information to construct a localization system. Optical cameras provide high-resolution color images. However, the use of these sensors is extremely limited owing to light attenuation and turbidity in the underwater environment. Underwater sonar can produce range information without being affected by environmental factors; therefore, it frequently has shown higher usability than optical cameras in several underwater applications.

Underwater sonar is categorized into ranging sonar (e.g., single-beam and multibeam echo sounders) and imaging sonar (e.g., side-scan and multibeam imaging sonar). Multibeam imaging sonar (so-called forward-looking imaging sonar or acoustic camera) produces acoustic images representing the range of environment for a 3-D fan-shaped imaging area. However, the acoustic image does not represent height information above the image plane, and the spatial ambiguity increases rapidly as the detection distance increases. In addition, it makes completely different imaging results depending on the arrangement of the imaging sonar and surrounding objects. Despite these difficulties, compared with other sonars, the acoustic imaging sonar has a high imaging resolution, wide imaging area, high frame rate, and a long imaging distance of approximately 100 m, which is advantageous for detecting near-field objects in the intervention environment. Thus, considering the operational characteristics of the intervention system, the acoustic imaging sonar would be highly useful as a perception sensor for the localization correction of a localization system.

Therefore, this paper proposes a new localization technique using an acoustic imaging sonar for autonomous intervention systems operating near underwater structures, such as underwater wells or manifolds on the seabed. This technique is based on an integrated IMU-DVL system and uses relative pose estimation on acoustic images for localization correction, allowing the vehicle to have mid-range localization capabilities with respect to subsea platforms. The merits of the proposed technique were verified using real data obtained from a test tank. The remainder of this paper is organized as follows. Section 2 describes the related research on underwater localization using a single acoustic imaging sonar. Section 3 addressed a localization strategy for autonomous intervention systems. This section covers pose estimation techniques for a single acoustic image and a way to construct the localization system with an integrated IMU-DVL system. In Section 4, the feasibility of the proposed technique is validated through an experiment using a dataset. At the end of this paper, we summarize and conclude this paper with some future directions.

II. RELATED WORKS

The localization with the imaging sonar has several advantages compared to conventional acoustic-based localization systems (e.g., LBL, SBL, and USBL) in intervention tasks using underwater vehicles. First, the system using the imaging sonar does not require any infrastructure (e.g., transducer beacon), which costs a lot to install, unlike conventional systems. Second, the pose estimation with imaging sonar gets directly the relative pose for the subsea structure of interest, so it is particularly suitable for localization of the intervention system. Third, the imaging sonar is an active sonar, so it does not suffer from the issue of acoustic shadow zone in positioning.

Because acoustic imaging sonars have demonstrated high usability in underwater environments, many researchers have addressed localization problems using acoustic imaging sonars to increase vehicle autonomy. One study proposed a simultaneous localization and mapping (SLAM) implementation using an exact sparse extended information filter (ESEIF) with manual features in acoustic images [10]. The study focused on real-time performance rather than feature detection. Unlike the method with manual features, this method extracts dense features from acoustic images and applies pairwise registration between acoustic images using normal distribution transformation (NDT) [11]. The registrations were combined with onboard velocity, attitude, and acceleration sensors. One approach used the SLAM framework to detect and track features in acoustic images to renavigate a mapped target in a shallow-water ocean environment [12]. Another study introduced the concept of the acoustic structure from motion (ASFM) technique using multiple acoustic images to reconstruct the selected point features [13]. The study demonstrated that this method can be applied in real-time navigation and SLAM for autonomous underwater vehicles in general 3D environments. A study proposed point-based relative pose estimation via bundle adjustment, with the assumption of a fixed sensor elevation [14]. In that study, the accelerated-KAZE (AKAZE) feature, which has been primarily used for optical images, was used to extract features from acoustic images. The study developed a low-cost acoustic-inertial navigation system that efficiently fuses acoustic images and inertial measurements within a tightly coupled extended Kalman filter (EKF) framework [15]. In addition, the study introduced acoustic feature linear triangulation to generate initial estimates for nonlinear least-squares-based feature localization. Some efforts have been conducted to determine loop closures in acoustic images for the SLAM framework. A learning architecture was designed to compare two acoustic images and determine whether they correspond to the same underwater scene [16]. A pose-invariant topological graph was built to represent loop closures between acoustic images in a semi-structured environment [17]. Acoustic landmarks were used for underwater navigation, and they were detected using a technique known as beam slice-based recognition [18]. An acoustic image simulator was used to determine the

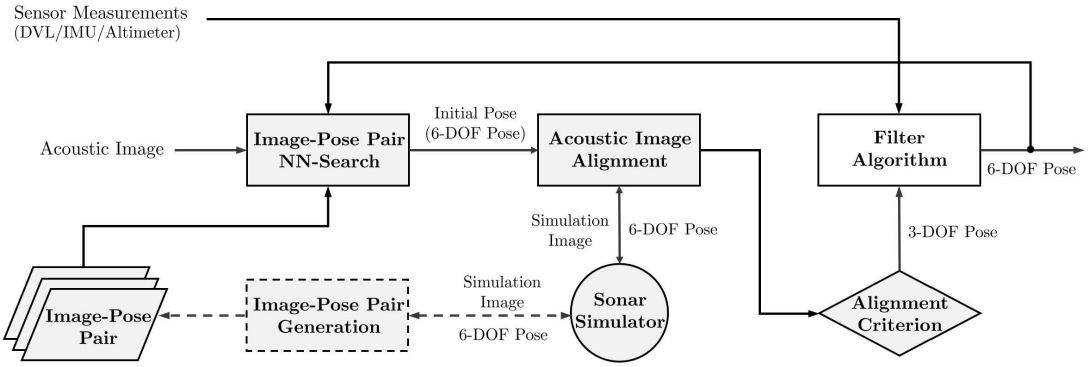


FIGURE 2. Schematic diagram of the proposed localization system: The bold and dotted lines represent the main flow and offline process of the localization system, respectively.

relative 3-D pose of a target whose shape and dimension are known, and it is subsequently employed for the global localization of underwater vehicles [19].

As acoustic imaging sonars provide rich texture information under certain conditions, image registration techniques using acoustic imaging sonars have received considerable research interest. The image registration here is primarily used to create acoustic mosaic images; however, we have confirmed that these techniques can be used to construct an underwater localization system. A Fourier-based technique was used for image registration of acoustic images, avoiding the extraction of features and ensuring a fast implementation [20]. Here, it is formulated using a pose graph, which was incrementally optimized using the g2o framework. The optimized poses were then used to build a mosaic online. Meanwhile, the feature point was influential in the image registration of the acoustic image. This method incorporates the detection of critical features and landmarks. It effectively represents them with a cluster-based Gaussian map, which provides a more efficient representation of the feature blobs in acoustic images [21]. Feature tracking using particle filtering was introduced to register the acoustic image sequences [22]. It begins with the extraction of the intensity, texture, and shape features from an unstructured seabed environment. The Gabor feature and weighted angular projection function (WAPF) were introduced to determine point pairs in acoustic images [23]. Unlike the method that uses feature points, a study attempted to match acoustic images using an image similarity measure. Peripheral mutual information maximization was proposed for acoustic image registration [24]. It was inspired by regional mutual information (RMI), which utilizes the closed-form solution for Shannon entropy. In addition, an optical flow model was introduced to estimate a pixel displacement map between consecutive acoustic images [25]. Using a weighted regularized spline technique, the incremental inter-frame motions were integrated into an attitude trajectory for the acoustic sensor.

Most of the above studies are basically based on image comparison to obtain navigational information from consecutive acoustic images with similar imaging view directions and rich acoustic textures. However, those studies have shown their usefulness in some applications, but they have various technical challenges to be used in localization system for autonomous intervention systems.

III. METHODOLOGY

A. LOCALIZATION STRATEGY

This section addresses the localization strategy enabling autonomous intervention systems to maneuver near subsea structures and perform intervention tasks without the need for external beacons for positioning. The strategy is based on an integrated IMU-DVL system using navigation sensors, a Doppler velocity log (DVL) and an inertial measurement unit (IMU) for the vehicle's underwater maneuvering. It employs a relative pose estimation using an acoustic imaging sonar for localization correction with respect to subsea structures.

An important assumption here is that all sensors are mounted on the intervention system and their positional relationships are known precisely. In addition, subsea structures are designed and deployed by humans; therefore, their shape and layout are entirely known. Consequently, it enables a sonar simulator to simulate the localization space and all possible acoustic images that can be created in the space. Ideally, if the simulation environment sufficiently describes the actual localization space, and if the sonar simulator produces identical simulated images with input acoustic images, we can assume that the pose in which the simulated image is created is the position where the actual acoustic image was obtained.

Based on this assumption, our localization strategy formulates the pose estimation for an input acoustic image as an image alignment problem between the input and simulated acoustic images. Specifically, the pose estimation comprises an image-pose pair group generation, image-pose pair nearest-neighbor (NN) search, and acoustic image alignment. First, the image-pose pair group generation is an offline process. It builds a database that contains pose hypotheses over a bounded localization area and simulated images at each hypothesis using a sonar simulator. Next, the NN search finds an image-pose pair in the database and guesses the pose in which an actual input image was taken. The poses of the

NN search are in discrete pose spaces and may differ from the actual poses. Therefore, the acoustic image alignment tries to find the actual pose by matching the input image and the image generated in the continuous pose space of the simulator, assuming that the actual pose is near the pose from the NN search. Lastly, the pose estimation results are combined with sensor measurements from DVL and IMU in an extended Kalman filter system, and it estimates the six-degree-of-freedom (6-DOF) pose of the intervention system and subsea structures in the global frame.

B. POSE ESTIMATION FOR A SINGLE ACOUSTIC IMAGE

1) IMAGE-POSE PAIR GROUP GENERATION

Image-pose pair group generation samples a certain number of pose hypotheses over a 3-D pose space in a bounded localization area and generates simulated acoustic images for each pose hypothesis using the sonar simulator. As a result, the group contains simulated acoustic images and their 6-DOF pose labels.

The group can not contain all possible hypotheses owing to limited system resources. Therefore, the group generation samples a specific number of poses evenly in each pose dimension. Consequently, it can be a 6^N pose hypothesis if N samplings are performed in each dimension. Underwater vehicles, including intervention systems, typically have small roll and pitch motions, in which case we can only consider 4-DOF for each pose hypothesis. In addition, if we can ignore the vehicle's vertical motion for specific intervention tasks, the pose hypothesis can be a 3-DOF pose. This pair generation requires significant computation due to acoustic image simulation, but it does not affect the real-time capability of the localization system as it is processed offline. After generating all pairs, each pair is indexed in numerical order along with the pose labels. The index is used for the acoustic image NN search for initial pose estimation.

2) IMAGE-POSE PAIR NEAREST NEIGHBOR SEARCH FOR ACOUSTIC IMAGE

Image–pose pair NN search finds a pair with the image most similar to an input acoustic image in the image–pose pair group and then returns a pose label of the pair. Considering a large number of the image–pose pairs in the group, a search without data compression requires significant computation. Hence, the image-pose pair NN search employs an efficient hash-based image search technique using image similarity measure, minhahsing, and locality-sensitive hashing (LSH). Here, minhashing is a dimension reduction method preserving the Jaccard similarity between data. LSH is a clustering technique that uses band partitioning to rapidly allocate similar input data to the same group, called buckets.

Specifically, the NN search comprises image–pose pair clustering and image–pose pair selection. Image–pose pair clustering is an offline process that clusters all pairs in a group based on the image distance using image hashing, minhahsing, and LSH. First, this process creates a hash table composed of values and keys corresponding to the pose labels and image hashes of the pairs, respectively. Each image of m pairs is transformed into n bit vector by applying a specific hashing function and the bit vectors are stacked to form a hash table whose size is an $n \times m$ matrix.

The use of binary hashing loses the information of the original data, but it enables to quickly compare the input data with the hash table. Even after image hashing, n and m are typically very large. The minhashing transforms the input hash table into an $r \times m$ matrix called the signature matrix through r random permutations, where the signature matrix commonly has smaller dimensions than the input hash table while preserving the Jaccard similarity between two hash codes in the table. Subsequently, LSH divides the signature matrix into b bands with h hash codes, compares each band using Jaccard similarity, and assigns pose labels (or keys) with high similarity to the same bucket. Here, LSH treats all pose labels belonging to the same bucket as the same data. Lastly, the pair clustering yields a set of buckets resulting from clustering for the image-pose pair group.

Image–pose pair selection is an online process that finds an image–pose pair with the smallest image distance to the input acoustic image in a pre-generated bucket and then returns the pose label of that pair. Concretely, the pair selection clusters the input image into one of the pre-generated buckets by sequentially applying image hashing, minhahsing, and LSH, similar to the offline process. Because a bucket can have one or more image–pose pairs, we determine the image–pose pair with the smallest image distance to the input image using correlation ratio, a similarity measure suitable for multimodality data comparison. As a result, the pose label searched here is used as an initial guess for image alignment, which is introduced in the next section.

3) ACOUSTIC IMAGE ALIGNMENT

The initial poses of the NN search are in discrete pose spaces and may differ from the actual poses. Therefore, it needs a procedure to find the actual pose in a continuous pose space. Like the NN search, we assume that the pose simulating the image with the smallest image distance from the input image represents the actual position at which the input image was obtained. Therefore, we employ the image alignment technique that tries to find the actual pose by matching the input image and the image generated in the continuous pose space of the simulator, assuming that the actual pose is near the pose from the NN search. The image alignment can be formulated as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}}\, d(S(\mathbf{x}), I^t) \tag{1}$$

The equation determines the acoustic imaging sonar's 6-DOF pose $\hat{\mathbf{x}}$ that minimizes the image distance $d(S(\mathbf{x}), I^t)$ between the acoustic image $S(\mathbf{x}) = [S_r(\mathbf{x}), S_\psi(\mathbf{x})]$ simulated at a pose $\mathbf{x}$ and the input acoustic image $I^t = [I_r^t, I_\psi^t]$ obtained at a time t with an initial pose $\tilde{\mathbf{x}}$. The subscripts r and ψ denote two different imaging coordinates of acoustic

Algorithm 1: Acoustic Image Alignment

Input: $\tilde{\mathbf{x}}$, $I^t = [I^t_r, I^t_\psi]$
Result: $\hat{\mathbf{x}}$
Param: $K, N, \zeta, \mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3]$

$\mathbf{x}_1 \longleftarrow \tilde{\mathbf{x}}$
$F_I \longleftarrow$ FeatureExtraction (I^t)
for $k \leftarrow 1$ **to** K **do**
  for $n \leftarrow 1$ **to** N **do**
    $\lambda_n \longleftarrow$ BidirectionalLineSearch $(\mathbf{x}_k, \mathbf{d}_n, I^t)$
    $\mathbf{x} \longleftarrow \mathbf{x}_k + \lambda_n \mathbf{d}_n$
  end
  $\mathbf{d}_k \longleftarrow$ SetSearchDirection $(\mathbf{x}_k, \mathbf{x})$
  $\lambda_k \longleftarrow$ BidirectionalLineSearch $(\mathbf{x}_k, \mathbf{d}_k, I^t, F_I)$
  $\mathbf{x}_{k+1} \longleftarrow \mathbf{x}_k + \lambda_k \mathbf{d}_k$
  $\hat{\mathbf{x}} \longleftarrow \mathbf{x}_{k+1}$
  $\mathcal{E} \longleftarrow |\mathbf{x}_k - \mathbf{x}_{k+1}|$
  if $\mathcal{E} < \zeta$ **then**
    Return$(\hat{\mathbf{x}})$
  end
end

Algorithm 2: Bidirectional Line Search

Input: $\mathbf{x}_k$, $\mathbf{d}$, I^t, F_I
Result: λ
Param: $M, \gamma, \lambda = [\lambda_1, \lambda_2, \lambda_3]$

$\mathbf{x}_1 \longleftarrow \mathbf{x}_k$
for $m \leftarrow 1$ **to** M **do**
  for $l \leftarrow 1$ **to** L **do**
    $\mathbf{x}_l \longleftarrow \mathbf{x}_m + \lambda_l \mathbf{d}$
    $S \longleftarrow$ AcousticImageSimulation $(\mathbf{x}_l)$
    if InnerLoop **then**
      $\mathcal{D}_l \longleftarrow$ ImageSimilarity (S, I^t)
    else
      $F_S \longleftarrow$ FeatureExtraction (S)
      $\mathcal{D}_l \longleftarrow$ ImageDistance (S, I^t, F_S, F_I)
    end
  end
  $\mathcal{D} \longleftarrow [\mathcal{D}_1, \cdots, \mathcal{D}_L]$, $\Lambda \longleftarrow [\lambda_1, \cdots, \lambda_L]$
  $\mathbf{x}_{m+1}, \Lambda \longleftarrow$ UpdateSamplePoint $(\mathcal{D}, \Lambda)$
  $\lambda \longleftarrow \mathbf{x}_k - \mathbf{x}_{m+1}$
  $\mathcal{E} \longleftarrow |\mathbf{x}_{m+1} - \mathbf{x}_m|$
  if $\mathcal{E} < \zeta$ **then**
    Return(λ)
  end
end

imaging sonar called x-y and r-azimuth images, respectively. The parameters of $\mathbf{x}$ to be estimated can be varied according to the sensor configuration and arrangement, compositions of the surrounding environment, and localization strategy. For example, $\mathbf{x}$ can be a 4-DOF pose by ignoring the roll and pitch motions of the imaging sonar (or vehicle). In addition, if there is no vertical motion or if the altitude is measured directly from the sensor, $\mathbf{x}$ can be a 3-DOF pose by ignoring the vertical motion.

According to literature, an image alignment for acoustic images can be performed using feature-based, intensity-based, and Fourier-based methods. Feature-based methods directly estimate the 2-D geometric transformations of two acoustic images. However, feature-based methods would not work well for two acoustic images that have multimodal relationships, large differences in imaging positions, or sparse textures. Meanwhile, intensity-based methods find transformations for two images with image similarity measure, and Fourier-based methods determine transformations in the frequency domain. These two methods are typical options for texture-rich images and images with a few differences in their imaging position.

Actual acoustic images and simulated images have sparse pixel distributions due to the arrangement of sensors and subsea structures addressed in this paper. Furthermore, they frequently have large different imaging positions, resulting in significant pixel differences between images. Therefore, using just one of the methods above would not be suitable for our image alignment problem because they assume slight pixel differences or dense acoustic texture. In image-alignment problems, the combination of several image distances, each of which accounts for data in a different way, may yield better results. Thus, for our challenging image alignment, we consider the correlation ratio M, which is an intensity-based image distance that measures pixel dependence between two data, and the Chamfer distance C, which indirectly measures geometric differences, even for sparse texture images. Accordingly, the resulting image distance is expressed as the sum of the two distances, as shown in the following equation:

$$d(S_*(\mathbf{x}), I_*) = C(S_*(\mathbf{x}), I_*) + \beta \cdot M(S_*(\mathbf{x}), I_*), \qquad (2)$$

where β is a weighting for correlation ratio. The subscript indicates the imaging type and can be r or ψ. Here, the Chamfer distance measures the spatial similarity between the contours of the two images and takes a distance value greater than zero. This distance requires extracting the contours from both images and then generating a distance function for one of them. As edge pixels with high positive gradients of pixel values on the distance axis in acoustic images are known as representative feature points representing the outline of objects in the image, we use the edge pixels to calculate the Chamfer distance. For this, we employ an edge extraction technique [26] to find edge features and their normal vectors and then cluster the edge features with a range of normal directions and pixel gradients. The correlation ratio measures the functional dependency between two images and takes values between 0 (no functional dependency) and 1 (purely deterministic dependency). In particular, this measure is known to be effective for multimodality image comparison, so we expect it to be effective for our image registration problem.

The r-azimuth image presents pixels with the range and the azimuth axis, and the x-y image uses two distance coordinates. Under an assumption of a static environment, the pixel motion in the x-y image represents the 3-DOF motion of the acoustic imaging sonar. The pixel motion on each axis of the r-azimuth image corresponds to the range and heading motion, respectively. Consequently, we can directly estimate the heading motion of the sonar just by measuring the pixel motion in the azimuth axis, and it will certainly facilitate the acoustic image alignment. Therefore, this paper uses the two image representations to calculate the image distance. The image distance can be expressed as the following equation:

$$d(S(\mathbf{x}), I_t) = \alpha \cdot d(S_r(\mathbf{x}), I_r) + (1-\alpha) \cdot d(S_\psi(\mathbf{x}), I_\psi), \quad (3)$$

where α is a tuning parameter that has different values during the parameter estimation. For example, when estimating the azimuth parameter, β is set to a value less than 0.5 to give more weight to the distance for the r-azimuth image.

Algorithm 1 summarizes the parameter estimation of the image alignment. Image alignment uses the initial pose $\tilde{\mathbf{x}}$ from the NN search and an acoustic image I^t at time t from an actual acoustic imaging sonar. Subsequently, it returns the estimated pose $\hat{\mathbf{x}}$. K and N denote the number of iterations for the parameter estimation and line search, respectively. Here, N corresponds to the degree of freedom of the pose to be estimated and was set to three here. The vector $\mathbf{d}$ is a unit vector representing each degree of freedom of the pose to be estimated and used for line searches in *BidirectionalLineSearch*. This line search is based on the golden-section search or Brent's method and exists in the inner and main loops, respectively. The one for the inner loop determines the acceleration direction of the parameter updates, and the one for the main loop determines the step size of update acceleration λ.

Algorithm 2 is a pseudocode describing how a local line search determines the step size of parameter updates λ, and its inputs are a pose $\mathbf{x}_k$ and vector $\mathbf{d}$ indicating the starting point and direction of the line search, respectively. Meanwhile, this line search employs different cost functions for the inner and main loops of Algorithm 1. Specifically, because the inner loop performs a line search N times that of the main loop, the correlation ratio is used as a cost function to pay more attention to fast computation. Meanwhile, the main loop accelerates parameter estimation, and it should be delicately controlled. Therefore, a combination of the Chamfer distance and the correlation ratio is used as a cost function for more accurate parameter estimation.

4) IMAGE ALIGNMENT CRITERION

A criterion was used to determine whether to use pose estimation for filter updates. Let I^t be an input acoustic image at time t and a simulation image $S(\mathbf{x})$ at pose $\mathbf{x}$ found by image alignment. Then, the criteria can be obtained using the equation below:

$$\eta_r = \frac{C(S_r(\mathbf{x}), I_r^t)}{C(S_r(\mathbf{x}), G(S_r(\mathbf{x})))}, \quad (4)$$

where the denominator is a correlation ratio between a simulation image $S_r(\mathbf{x})$ at pose $\mathbf{x}$ and its image smoothing result $G(S_r(\mathbf{x}))$. The numerator is a correlation ratio between the simulation image $S_r(\mathbf{x})$ and an input image I_r^t. The denominator is the maximum value of correlation ratio we can get when two images are aligned. We empirically found out that the numerators have lower values than the denominators, even though two images were aligned perfectly. As a result, one pose estimation is used for localization correction only when the ratio η_r of the two correlation ratios is greater than a threshold.

C. LOCALIZATION SYSTEM USING EXTENDED KALMAN FILTER

1) SYSTEM DYNAMICS MODEL

The filter system uses the 6-DOF kinematic model of the vehicle (or intervention system) driven by inertial sensor measurements [27], [28]. The state vector of the motion of the vehicle is expressed as

$$\mathbf{x}_v = \begin{bmatrix} x & y & z & \phi & \theta & \psi & u & v & w \end{bmatrix}^T, \quad (5)$$

where the x, y, and z are the positions of the vehicle in the global frame; ϕ, θ, and ψ are the Euler angles in the global frame; and u, v, and w are the linear velocities in the vehicle-fixed local frame. The vehicle pose in the global frame is obtained by integrating the linear accelerations and angular velocities from the IMU. The IMU measurements are represented as

$$\mathbf{z}_{\text{imu}} = \begin{bmatrix} z_{\dot{u}} & z_{\dot{v}} & z_{\dot{w}} & z_p & z_q & z_r \end{bmatrix}^T, \quad (6)$$

where the $z_{\dot{u}}$,$z_{\dot{v}}$, and $z_{\dot{w}}$ are the linear accelerations; and z_p, z_q, and z_r are the angular velocities. The IMU measurements are used to propagate the dynamics of the filter system. To estimate the poses of the vehicle and the subsea structure together, the state vector is composed of the state vector of the structure $\mathbf{x}_m = [x_m\, y_m\, z_m\, \phi_m\, \theta_m\, \psi_m]^T$ and the state vector of the vehicle.

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_v \\ \mathbf{x}_m \end{bmatrix} \quad (7)$$

The system dynamics consist of two motion models: the subsea structure and the intervention system, which are used to estimate the spatial relationship between the intervention system and the subsea structure of interest given the sensor measurements in the state estimator, and that can be represented as

$$\dot{\mathbf{x}} = \begin{bmatrix} \dot{\mathbf{x}}_v \\ \dot{\mathbf{x}}_m \end{bmatrix} = \begin{bmatrix} f(\mathbf{x}_v, \mathbf{z}_{\text{imu}}) \\ 0 \end{bmatrix} + \mathbf{w}, \quad (8)$$

where $f(\mathbf{x}_v, \mathbf{z}_{\text{imu}})$ is the motion model represented in [27]. Here, $\mathbf{w}$ refers to the uncertainty assumed to follow a zero-mean Gaussian distribution. As the subsea structure is

assumed to be stationary, the time derivative $\dot{\mathbf{x}}_m$ of their state vector is set to zero.

2) MEASUREMENT MODEL

The measurement model employs two measurements for the filter update. One is the measurement $\mathbf{z}_v$ composed of the linear velocity (u, v, and w) and altitude d obtained from the DVL and altimeter sensor, respectively. In addition, the measurement model employs the relative pose $\mathbf{z}_m = [\, z_x \;\; z_y \;\; z_\psi \,]^T$ between the subsea structure and the acoustic imaging sonar given by the proposed pose estimation. Therefore, the resulting measurement model can be represented as

$$\begin{aligned}\mathbf{z} &= [\mathbf{z}_v \;\; \mathbf{z}_m]^T + \mathbf{v} \\ &= [u \quad v \quad w \quad d \quad z_x \quad z_y \quad z_\psi]^T + \mathbf{v}, \end{aligned} \tag{9}$$

where $\mathbf{v} = [\mathbf{v}_v \quad \mathbf{v}_m]^T$ is the measurement noise. $\mathbf{v}_v$ is the measurement noise, which is assumed to follow a zero-mean Gaussian distribution with covariance matrix $\mathbf{R}_v$, and $\mathbf{v}_m$ is the measurement noise of the pose estimation with a zero-mean Gaussian distribution with covariance matrix $\mathbf{R}_m$ as follows:

$$\mathbf{R}_m = \begin{bmatrix} \kappa_a^2 \cdot e^{2\cdot a\cdot \eta_r} & 0 & 0 \\ 0 & \kappa_a^2 \cdot e^{2\cdot a\cdot \eta_r} & 0 \\ 0 & 0 & \kappa_b^2 \cdot e^{2\cdot b\cdot \eta_\psi} \end{bmatrix}. \tag{10}$$

Here, κ_a and κ_b represent the sensor distance resolution per pixel of I_r and the azimuth resolution per pixel on the azimuth axis of I_ψ, respectively. η_r and η_ψ are the criterion calculated for I_r and I_ψ as represented in Eq 4. a and b are tuning parameters.

IV. EXPERIMENTAL VALIDATION

This section shows the feasibility and usability of proposed techniques with a dataset obtained in a test tank.

A. EXPERIMENTAL SETUP

The experimental validations were conducted under a scenario in which an intervention system equipped with an acoustic imaging sonar and navigation sensors (DVL and IMU) maneuvers near a cluster of subsea platforms for their intervention tasks.

The validations used a dataset obtained from an acoustic imaging sonar, DVL, and IMU, without using a robotic platform. The dataset was obtained by moving a mounting frame on which the sensors were mounted in the test tank (see Fig. 3). The rotational and vertical motions of the sensors were fixed, and the imaging plane of the acoustic imaging sonar was fixed parallel to the bottom of the tank. According to the validation scenario, two scaled mock-up models of the control panel of the wellhead, an offshore platform for oil or gas production, were built and placed at the bottom of the test tank. As the shape and positional relationship of the mock-up models and the tank were known precisely, 3D computer-aided drawing (CAD) models were easily obtained, and they were used for the sonar simulator to generate simulated acoustic images. We set up a rectangular path

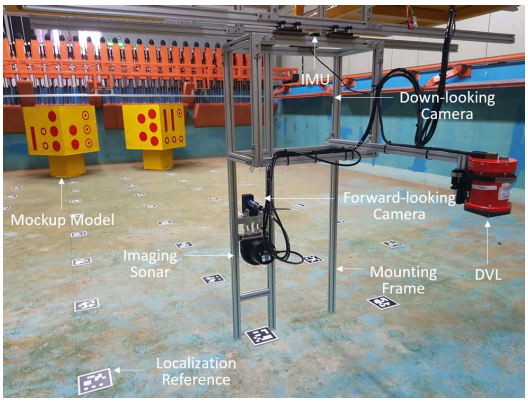


FIGURE 3. Experimental setup for data acquisition: Sensors, mounting frame, two mock-up models and markers for localization reference were deployed in a water tank.

TABLE 1. Specifications for equipment used for data acquisition experiment.

Equipment	Category	Description
Acoustic imaging sonar	Model	Teledyne BlueView P900-45
	Frequency	900 kHz
	Update rate	Up to 15 Hz
	Field-of-view	45°
	Max range	100 m
	Beam width	1 - 20°
	Number of beams	256
	Beam spacing	0.18°
	Range resolution	2.54 cm
Navigation sensors	IMU	Microstrain 3DM-GX-25
	DVL	Linkquest NavQuest 600 Micro
Water basin	Dimensions	12, 10 and 1.5 m of length, width and depth respectively
	Features	Towing platform for longitudinal and transversal motion
Mock-up model	Dimensions	0.9, 0.9 and 1.4 m of length, width and height respectively
	Features	Two cube structures having patterns representing valves, connectors and other features of a control panel for subsea oil wells

for the mounting frame to follow, and this path was positioned near the mock-up models, considering the experimental scenario. The mounting frame had a down-looking camera to obtain relative poses between the acoustic imaging sonar and visual markers placed on the bottom of the tank, and these poses were used to generate a localization ground truth. The specifications for each sensor and the experimental environment are summarized in Table 1. The sensors used here are typically mounted on underwater vehicles for navigation and monitoring purposes.

NN Search has higher performance as the size of the hash table increases, but we cannot consider all possible pairs. Therefore, the performance of techniques using pair groups was observed with a limited number of pairs. We prepared three pair groups with different numbers of pairs for a localization area we have set. Each group was obtained by evenly sampling 5, 10, and 20 pairs per meter for the position dimension and per 5 degrees for the azimuth dimension.

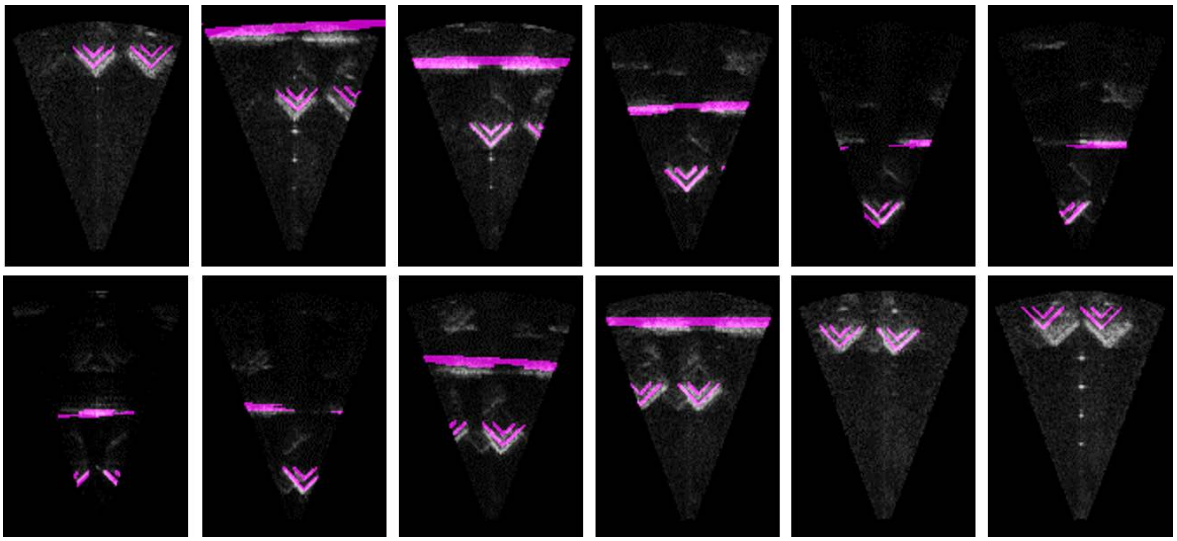


FIGURE 4. Results of image-pose pair NN search: The results were represented with superimposed images between input images (gray) and NN search results (magenta). They are arbitrary selections of input images at which successful pose estimation was performed. Some of the results were exactly the same as the input image, and some estimations were slightly mismatched but did not deviate much from the correct answer.

TABLE 2. Results of NN search: RMSE is represented for the entire localization area and an area within a distance d of 5 meters from the mock-up models. All results were shown according to the different number of pairs. N and N_n are the number of input images and successful NN searches. The subscript 5 refers to attempts occurred within a distance of 5 meters from the mock-up models.

	Entire path					**Mid-range** (d $\leq$ 5 m)				
Sampling size	x (m)	y (m)	yaw (°)	N_n/N	time (sec)	x (m)	y (m)	yaw (°)	$N_{n,5}/N_5$	time (sec)
5	0.338	0.293	4.53	0.21	0.13	0.182	0.208	4.22	0.21	0.12
10	0.234	0.219	4.22	0.25	0.25	0.190	0.172	3.23	0.28	0.26
20	0.221	0.217	3.68	0.25	0.44	0.173	0.161	2.97	0.29	0.41

B. RESULTS AND DISCUSSION

1) RESULTS OF IMAGE NEAREST NEIGHBOR SEARCH

The NN search was evaluated using all acoustic images we obtained in the experiment. Fig. 4 shows overlay images of actual acoustic images and simulated images found by NN search. Here, each image's gray background image is the actual acoustic image, and the magenta foreground image is the simulated image found by NN search. The BlueView p900-45 used here features a relatively long detection range and low imaging resolution. This sensor's acoustic image, In particular, has a low imaging resolution in the far imaging area, as opposed to the oversampled near imaging area. This characteristic is also visible in the acoustic image that we obtained. Consequently, it causes a considerable image difference between the real and simulated images for distant objects and degrades the performance of the NN search, which assumes the same pixel distribution of the two images obtained at the same imaging position. In this regard, the last figure of Fig. 4 depicts an example. Here, the NN search failed to align the simulated image with the actual image, and it can be seen that the shape of the object is different in the two images. On the other hand, NN searches performed well for most image alignments for both images of nearby objects.

Table 2 summarizes the root-mean-squared error (RMSE), success rate, and computation time of NN searches over the entire dataset. Here, The success rate is calculated for NN search results with a particular position and azimuth error that was set considering the resolution of the image-pose pair generation. For instance, if the pose labels of two adjacent pairs have a position difference of 0.5 m, the criterion will be $\sqrt{0.5^2 + 0.5^2}$. As mentioned earlier, NN searches were evaluated with three image-pose pairs that have different resolutions. This is to see the relationship between the number of image-pose pairs and the NN search's accuracy. As expected, the NN search showed a low localization error as the image-pose pair's resolution increased but required more time for the search. Meanwhile, NN search with a small number of pairs occasionally reached the image update rate of actual imaging sonar, though the localization inaccuracy was slightly increased. Overall, the imaging position from the target object had a more significant impact on the NN search's results than the number of pair groups. The NN search showed localization errors that allow the intervention system to maneuver around subsea structures, which is a satisfactory result considering the low resolution of the imaging sonar we used.

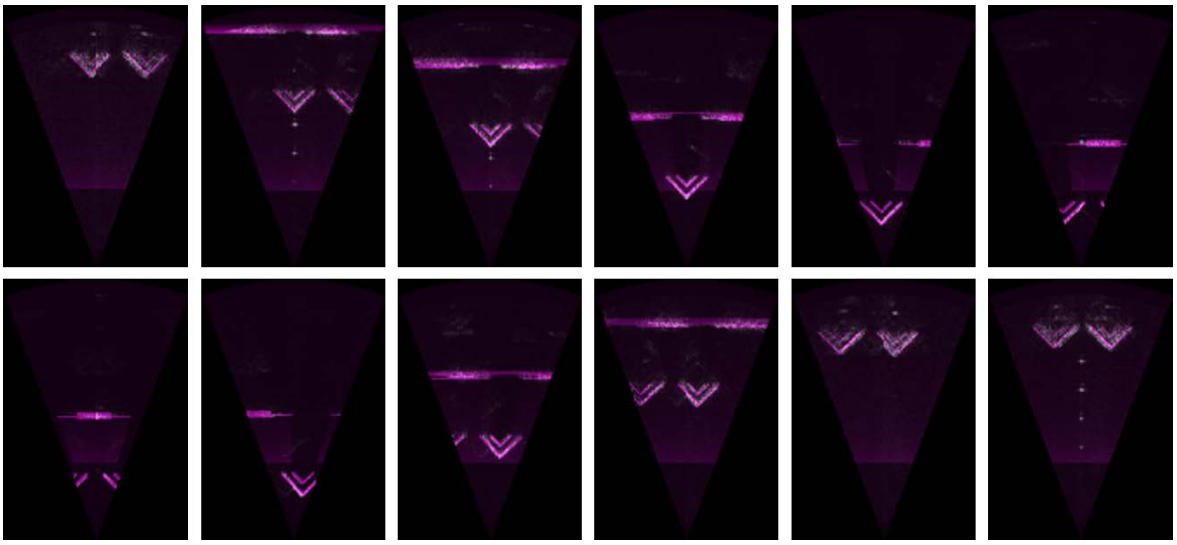


FIGURE 5. Results of acoustic image alignment: The results were represented with superimposed images between input images (gray) and NN search results (magenta). The input images here are the same as that of Fig. 4, and these results were obtained with the initials of the NN search for Fig. 4. All results were well aligned, but with some errors for distant object images.

TABLE 3. Results of acoustic image alignment: The results were calculated in the same way as the NN search results in Table. 2. N_a and N_n denote the number of successful image alignments and NN searches for the entire path. The subscript 5 refers to attempts occurred within a distance of 5 meters from the mock-up models.

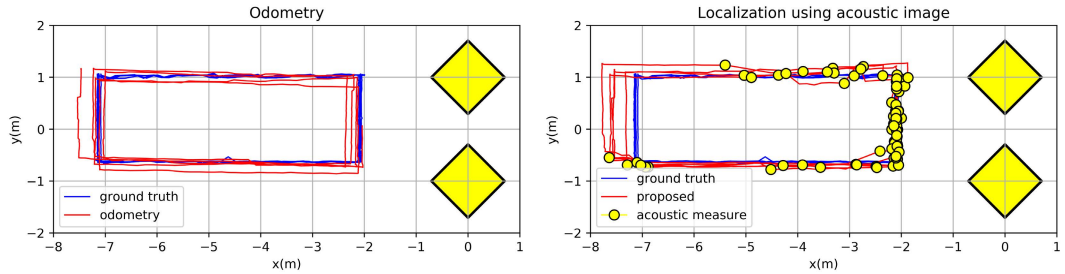
	Entire path					Mid-range (d $\leq$ 5 m)				
Sampling size	x (m)	y (m)	yaw (°)	N_a/N_n	time (sec)	x (m)	y (m)	yaw (°)	$N_{a,5}/N_{n,5}$	time (sec)
5	0.211	0.208	3.50	0.19	5.72	0.142	0.151	2.89	0.23	5.07
10	0.187	0.179	1.63	0.18	4.44	0.120	0.118	1.51	0.30	2.38
20	0.168	0.177	1.21	0.15	4.21	0.106	0.111	1.08	0.31	2.29

2) RESULTS OF ACOUSTIC IMAGE ALIGNMENT

Since image alignment is the post-processing of NN search, it was evaluated with NN search results. Fig. 5 represents the superimposed image of the actual image and the simulated image obtained by image alignment. In each image, the gray background image is the actual acoustic image, and this is equivalent to Fig. 4. The magenta foreground image represents the simulated image returned by image alignment. Fig. 5 shows that the two images are well aligned, which indicates that image alignment can improve the NN search results. Meanwhile, image alignment often failed or had some positional errors (see the first figure in Fig. 5); there are several reasons for this: First, image alignment assumes slight image differences between simulated and actual images at the same imaging position, but our simulator was not able to produce identical images to the actual camera, especially for distant objects. Second, as NN search had relatively high positional errors, image alignment employing the NN search's results in parameter estimation had to begin parameter estimation at incorrect initial values.

Table 3 summarizes the root-mean-squared error (RMSE), success rate, and computation time of image alignment. As with the NN search, this is computed over the entire data we collected. Here, the success rate was calculated for image alignment with a position error of less than 0.07 m, and this was a rather strict criterion given the resolution of the imaging sonar we used. Because the parameter estimation employs the NN search results as initial values, the image alignment's results were represented according to the pair group's resolution. Here, image alignment showed satisfactory error and convergence speed in the high-resolution pair group because NN search at this resolution had high positional accuracy. Image alignment, on the other hand, took longer to converge due to the NN search's poor results in the low-resolution pair group.

Image alignment showed different positional errors depending on some factors we stated, but it was more affected by the imaging position from the object than the pair group's resolution. Overall, image alignment showed a satisfactory ability to make localization corrections for intervention robots. In particular, this is an encouraging result in that global localization is possible with a low-resolution single-shot acoustic image without any image registration process. Meanwhile, image alignment usually works slowly but sometimes approaches the image update rate of a real camera. This could be improved as image alignment allows for a trade-off parameter between speed and accuracy.



(a) Dead-reckoning in rectangular maneuver

(b) Proposed localization in rectangular maneuver

FIGURE 6. Localization results of dead reckoning and proposed localization system: The two rectangles represent the mock-up models; their shape and layout are the same as that of the experiment. In (a), the blue and red lines represent the ground truth and dead reckoning paths, respectively. In (b), the red line indicates the path by the proposed localization. Yellow markers indicate where the relative pose to the mock-up models was measured using pose estimation by image alignment.

TABLE 4. Localization error statistics in Fig. 6: (a) Localization accuracy of the proposed technique and dead reckoning was evaluated for the entire path. (b) Localization error was calculated for paths within 2, 3, 5, and 7 meters distance *d* from the mock-up models.

Path	**Localization**	**Length** (m)	**Duration** (sec)	**Mean**			**SD**		
				x (m)	y (m)	yaw (°)	x (m)	y (m)	yaw (°)
Rectangular	Proposed	51.15	404	0.17	0.14	1.12	0.03	0.02	0.43
	Dead-reckoning	52.28		0.25	0.24	3.78	0.06	0.07	0.51

(a) Localization error statistics

Path	**Localization**	Close-range (d $\simeq$ 2m)		Mid-range (d $\leq$ 3m)		Mid-range (d $\leq$ 5m)		Long-range (d $\leq$ 7m)	
		pos(m)	yaw(°)	pos(m)	yaw(°)	pos(m)	yaw(°)	pos(m)	yaw(°)
Rectangular	Proposed	0.10	1.10	0.12	1.08	0.12	1.11	0.17	1.13
	Dead-reckoning	0.24	3.78	0.25	3.82	0.24	3.81	0.26	3.79

(b) RMSE according to distance from mock-up models

3) RESULTS OF UNDERWATER LOCALIZATION

The proposed localization system is compared to dead reckoning and ground truth system in this section. In this comparison, all localization systems employ the same filter system but use different information for filter update. Dead reckoning, in particular, employs just the DVL measurements to correct the linear velocity in the state vector. In addition to using DVL data, the proposed localization system uses relative poses for the mock-up model, and the ground truth system uses relative poses for the visual marker.

Figs. 6a and 6b show the trajectories the mounting frame traveled, which were estimated by the three localization systems. In both figures, the two yellow squares represent a mock-up model placed in a water tank, and its dimensions and layout are identical to the experimental setup. The blue lines indicate the path estimated by the ground truth system. The mounting frame traveled several times along a rectangular path in the data acquisition experiment. Therefore, the frame's trajectory in each round was slightly different because the operator manually controlled it.

The red line in Fig. 6a represents the trajectory estimated by the dead reckoning. This showed a more significant error than expected, even though the mounting frame moved a short distance. This was because the mounting frame to which all the sensors were attached experienced significant vibration during movement. The red line in Fig. 6b indicates the trajectory estimated by the proposed localization system, and the yellow circle markers indicate the location where localization corrections were done. As can be seen from the previous results, most of the localization correction occurred near the mock-up model, and for this reason, the proposed system showed different localization accuracy according to the distance from the mock-up model.

Table 4 summarizes the localization errors of dead reckoning and the proposed localization system. Overall, our localization system increased localization error as it moved away from the mock-up model, but this is acceptable accuracy, enabling the intervention robot to maneuver near subsea structures and maintain its position on the structures for intervention tasks. In particular, our localization system had a positional error of about 0.1 m near the mock-up model, which is very encouraging considering the resolution of the imaging sonar we used. The proposed system is especially useful for autonomous intervention robots because it can be implemented with only a few onboard sensors without any external infrastructure.

V. CONCLUSION

This study addressed an underwater localization algorithm using an acoustic imaging sonar for an autonomous intervention system, and its feasibility and usability were verified using a dataset obtained in a test tank. The proposed

localization system exhibited localization accuracy for the intervention system to maneuver near offshore platforms and perform autonomous intervention tasks without the requirement for external beacons for positioning. This study is the first attempt to implement global localization using single-shot acoustic images. Although some technical challenges remain to be solved, this study demonstrated the usefulness of localization using an acoustic imaging sonar in implementing autonomous intervention systems, which can significantly improve the overall efficiency and reliability of underwater operations.

REFERENCES

[1] P. Ridao, M. Carreras, D. Ribas, P. J. Sanz, and G. Oliver, "Intervention AUVs: The next challenge," *IFAC Proc. Volumes*, vol. 47, no. 3, pp. 12146–12159, Jan. 2014.

[2] J. Evans, P. Redmond, C. Plakas, K. Hamilton, and D. Lane, "Autonomous docking for intervention-AUVs using sonar and video-based real-time 3D pose estimation," in *Proc. MTS/IEEE OCEANS*, vol. 4, Sep. 2003, pp. 2201–2210.

[3] G. Marani, S. K. Choi, and J. Yuh, "Underwater autonomous manipulation for intervention missions AUVs," *Ocean Eng.*, vol. 36, no. 1, pp. 15–23, Jan. 2009.

[4] P. J. Sanz, P. Ridao, G. Oliver, G. Casalino, Y. Petillot, C. Silvestre, C. Melchiorri, and A. Turetta, "TRIDENT an European project targeted to increase the autonomy levels for underwater intervention missions," in *Proc. MTS/IEEE OCEANS*, Sep. 2013, pp. 1–10.

[5] E. Simetti, G. Casalino, S. Torelli, A. Sperindé, and A. Turetta, "Floating underwater manipulation: Developed control methodology and experimental validation within the TRIDENT project," *J. Field Robot.*, vol. 31, no. 3, pp. 364–385, May 2018.

[6] M. Prats, D. Ribas, N. Palomeras, J. C. García, V. Nannen, S. Wirth, J. J. Fernández, J. P. Beltrán, R. Campos, P. Ridao, P. J. Sanz, G. Oliver, M. Carreras, N. Gracias, R. Marín, and A. Ortiz, "Reconfigurable AUV for intervention missions: A case study on underwater object recovery," *Intell. Service Robot.*, vol. 5, no. 1, pp. 19–31, Jan. 2012.

[7] N. Palomeras, A. Carrera, N. Hurtós, G. C. Karras, C. P. Bechlioulis, M. Cashmore, D. Magazzeni, D. Long, M. Fox, K. J. Kyriakopoulos, P. Kormushev, J. Salvi, and M. Carreras, "Toward persistent autonomous intervention in a subsea panel," *Auto. Robots*, vol. 40, no. 7, pp. 1279–1306, Oct. 2016.

[8] P. Cieslak, P. Ridao, and M. Giergiel, "Autonomous underwater panel operation by GIRONA500 UVMS: A practical approach to autonomous underwater manipulation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 529–536.

[9] N. Palomeras, A. Peñalver, M. Massot-Campos, G. Vallicrosa, P. L. Negre, J. J. Fern ndez, P. Ridao, P. J. Sanz, G. Oliver-Codina, and A. Palomer, "I-AUV docking and intervention in a subsea panel," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 2279–2285.

[10] M. Walter, F. Hover, and J. Leonard, "Slam for ship hull inspection using exactly sparse extended information filters," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2008, pp. 1463–1470.

[11] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 4396–4403.

[12] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, "Relocating underwater features autonomously using sonar-based SLAM," *IEEE J. Ocean. Eng.*, vol. 38, no. 3, pp. 500–513, Jul. 2013.

[13] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 758–765.

[14] Y.-S. Shin, Y. Lee, H.-T. Choi, and A. Kim, "Bundle adjustment from sonar images and SLAM application for seafloor mapping," in *Proc. OCEANS*, Oct. 2015, pp. 1–6.

[15] Y. Yang and G. Huang, "Acoustic-inertial underwater navigation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 4927–4933.

[16] P. O. C. S. Ribeiro, M. M. dos Santos, P. L. J. Drews, and S. S. C. Botelho, "Forward looking sonar scene matching using deep learning," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 574–579.

[17] M. M. Santos, G. B. Zaffari, P. O. C. S. Ribeiro, P. L. J. Drews-Jr, and S. S. C. Botelho, "Underwater place recognition using forward-looking sonar images: A topological approach," *J. Field Robot.*, vol. 36, no. 2, pp. 355–369, Mar. 2019.

[18] J. Pyo, H. Cho, and S.-C. Yu, "Beam slice-based recognition method for acoustic landmark with multi-beam forward looking sonar," *IEEE Sensors J.*, vol. 17, no. 21, pp. 7074–7085, Nov. 2017.

[19] Y. Wang, Y. Ji, D. Liu, H. Tsuchiya, A. Yamashita, and H. Asama, "Simulator-aided edge-based acoustic camera pose estimation," in *Proc. OCEANS*, Feb. 2022, pp. 1–4.

[20] N. Hurtós, S. Nagappa, N. Palomeras, and J. Salvi, "Real-time mosaicing with two-dimensional forward-looking sonar," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 601–606.

[21] M. D. Aykin and S. Negahdaripour, "On feature matching and image registration for two-dimensional forward-scan sonar imaging," *J. Field Robot.*, vol. 30, no. 4, pp. 602–623, 2013.

[22] S. Song, J. M. Herrmann, K. Liu, S. Li, and X. Feng, "Forward-looking sonar image mosaicking by feature tracking," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2016, pp. 1613–1618.

[23] J. Zhang, F. Sohel, H. Bian, M. Bennamoun, and S. An, "Forward-looking sonar image registration using polar transform," in *Proc. OCEANS*, Sep. 2016, pp. 1–6.

[24] S. Song, J. Michael Herrmann, B. Si, K. Liu, and X. Feng, "Two-dimensional forward-looking sonar image registration by maximization of peripheral mutual information," *Int. J. Adv. Robotic Syst.*, vol. 14, no. 6, Dec. 2017, Art. no. 1729881417746270.

[25] B. T. Henson and Y. V. Zakharov, "Attitude-trajectory estimation for forward-looking multibeam sonar based on acoustic image registration," *IEEE J. Ocean. Eng.*, vol. 44, no. 3, pp. 753–766, Jul. 2019.

[26] A. Trujillo-Pino, K. Krissian, M. Alemán-Flores, and D. Santana-Cedrés, "Accurate subpixel edge location based on partial area effect," *Image Vis. Comput.*, vol. 31, no. 1, pp. 72–90, 2013.

[27] J. W. Langelaan, "State estimation for autonomous flight in cluttered environments," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Stanford Univ., Stanford, CA, USA, Jan. 2006.

[28] J. Park, T. Kim, and J. Kim, "Model-referenced pose estimation using monocular vision for autonomous intervention tasks," *Auto. Robots*, vol. 44, no. 2, pp. 205–216, Jan. 2020.

JISUNG PARK received the B.S. degree in mechanical engineering from the Kumoh National Institute of Technology, Gumi, South Korea, and the M.S. degree in ocean systems engineering and the Ph.D. degree in mechanical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. In 2021, he joined the Agency for Defense Development (ADD), Daejeon, where he is currently a Senior Researcher. He has been working in the areas of ocean robotics, vision-based intelligence systems, and underwater mapping and localization using acoustic sensors.

JINWHAN KIM received the B.S. and M.S. degrees in naval architecture and ocean engineering from Seoul National University, Seoul, South Korea, and the M.S. and Ph.D. degrees in aeronautics and astronautics from Stanford University, Stanford, CA, USA. From 1995 to 2000, he was with the Korea Institute of Machinery and Materials and subsequently with the Korea Ocean Research and Development Institute. He was also with Optimal Synthesis Inc., Los Altos, CA, USA, as a Research Scientist. In 2010, he joined as a Faculty Member with the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, where he is currently a Professor with the Department of Mechanical Engineering. He has been working in the areas of mobile robotics and control. He is a member of the American Institute of Aeronautics and Astronautics (AIAA).