2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Geometry-aided Underwater 3D Mapping Using Side-scan Sonar

Yiqiao Yang, Chenglin Pang, Chengdong Wu, Zheng Fang*

***Abstract*— In recent years, the interest in underwater exploration with Autonomous Underwater Vehicles (AUVs) equipped with side-scan sonars (SSS) has grown considerably. However, state-of-the-art SSS Simultaneous Localization and Mapping (SLAM) systems encounter challenges in data association across large viewpoint changes. Additionally, these systems assume that the seabed is a flat surface, leading to significant mapping error in uneven underwater terrains. To address these challenges, we propose a framework that leverages the side-scan sonar geometry to facilitate data association and improve mapping accuracy. The framework begins with a preprocessing module that extracts feature points and provides initial estimates of the elevation angles of the landmarks. Then, a non-consecutive data association module applies epipolar line search to establish correspondences between the current and historical frames. Finally, the mapping module uses side-scan sonar bundle adjustment to recover the positions of the landmarks. The proposed method is evaluated using an underwater terraced fields dataset. Our method achieves over 90% matching rate and reduces the average mapping error from 3.799 to 0.134.**

I. Introduction

In recent years, there has been an increasing interest in underwater exploration [1]. Autonomous Underwater Vehicles (AUVs) have been widely used for seabed mapping [2], underwater monitoring [3], and more. Similar to unmanned systems on the ground, Simultaneous Localization and Mapping (SLAM) technology enables AUVs to autonomously navigate through unknown underwater environments [4].

In underwater environments, electromagnetic-based sensors, such as GPS, cameras, and LiDAR, are limited by the rapid attenuation of electromagnetic waves. In such scenarios, acoustic sonar has become the preferred choice for underwater perception. Low-cost side-scan sonar (SSS) is particularly suitable for large-scale underwater exploration tasks, as it can achieve centimeter-level resolution within a detection range of several hundred meters [5]. Side-scan sonars are installed on both sides of an AUV to scan the seafloor. Typically, the beam width in the direction of the heading is narrow (less than 1°), while it is much wider (greater than 30°) in the direction perpendicular to the heading. SSS records the strength and time delay of the echoes to generate one-dimensional data representations, known as pings [6]. The distance to the target can be calculated using the speed of sound. However, when the AUV travels in a straight line, consecutive pings do not have overlapping regions, making it impossible to establish the data association between consecutive pings. Furthermore, SSS cannot capture the elevation angles of the targets, resulting in the inability to directly obtain the 3D positions of the targets.

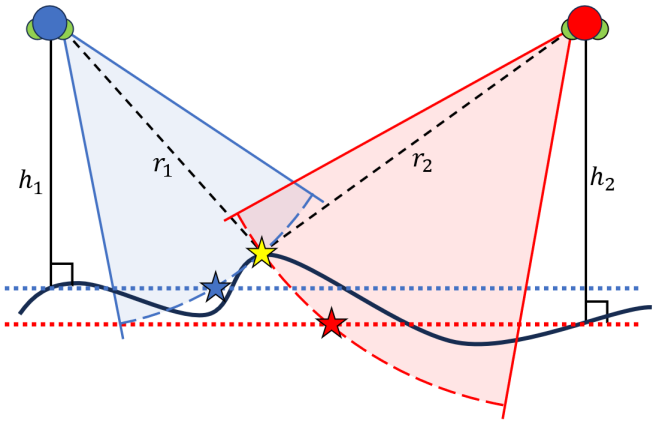


Fig. 1. Mapping error caused by the flat surface assumption: The blue and red circles represent different AUVs. The green circles represent side-scan sonars. Two AUVs observe the same landmark (yellow star) on the seabed (thick black line). When using the flat surface assumption, the blue AUV reconstructs the landmark as the blue star, while the red AUV reconstructs it as the red star.

Advanced approaches for using SSS in SLAM, such as [7], [8], [9], typically transform multiple one-dimensional pings into a two-dimensional image based on the trajectory. The image is called a mosaic image. The features in the image represent the projection of underwater landmarks. Then, these features are associated when loop closures occur. Finally, the associated features are used for pose optimization and mapping. In the mapping process, the elevation of the landmarks is uniformly set to the height of the seabed, which is called the flat surface assumption. However, there are two crucial challenges for SSS SLAM. In scenarios with large viewpoint changes, such as loop closure, the distribution of echo intensities for the same scenario can vary significantly [5]. This results in inconsistent intensity and geometric distortions in images and makes data association difficult. Furthermore, mapping with the flat surface assumption in uneven underwater terrains without optimizing the three-dimensional positions of landmarks can lead to a significant mapping error [10], as shown in Fig. 1.

To address the challenges, we propose a geometry-aided 3D mapping framework specifically designed for side-scan

This work was supported in part by the National Natural Science Foundation of China under Grants 62073066, in part by the Fundamental Research Funds for the Central Universities under Grant N2226001, and in part by 111 Project under Grant B16009. (Corresponding author: Zheng Fang.)
The authors are all with the Faculty of Robot Science and Engineering, Northeastern University, Shenyang 110819, China. Yiqiao Yang and Zheng Fang are also with the National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang 110819, China and also with the Key Laboratory of Data Analytics and Optimization for Smart Industry, Ministry of Education, Northeastern University, Shenyang 110819, China. (e-mail:fangzheng@mail.neu.edu.cn).

sonar, and it can be extended to a full SLAM system. This framework includes preprocessing, non-consecutive data association (NDA), and the side-scan sonar bundle adjustment (SSS BA). The core idea of our method is to establish the SSS projection and back-projection geometric models to assist in epipolar line search for NDA and to construct reprojection error for SSS BA. Our NDA can provide a large number of correct matches in scenarios with significant viewpoint changes, and SSS BA can recover the elevation angle of landmarks to achieve high-accuracy mapping. To validate the performance of our NDA in scenarios with large viewpoint changes, and to assess the mapping accuracy of our framework in uneven underwater terrains, we performed experiments on a 700m×700m underwater terraced fields dataset. Across the entire dataset, our NDA achieves a matching rate of over 90%, and our SSS BA can reduce the average mapping error from 3.799 to 0.134.

In summary, the main contributions of this paper are as follows:

- We establish the projection and back-projection geometric models for side-scan sonar, which provides the interconversion between 2D pixels and 3D landmarks.
- We propose a non-consecutive data association (NDA) method for SSS, which enables strong data association in scenarios with large viewpoint changes.
- We propose a side-scan sonar bundle adjustment (SSS BA) method, which recovers the elevation angles of landmarks. This significantly reduces the mapping error caused by the flat surface assumption.

The rest of this paper is organized as follows. In Sect. II, we discuss related work and how our work is unique compared to existing advanced methods. Sect. III presents our proposed method. Experimental results are shown in Sect. IV. A conclusion is made in Sect. V.

II. Related Work

In typical SLAM systems, robots use sensors to measure the environment and associate the measurement data from different viewpoints. The data association is then used to optimize the robot's pose and create a map of the environment [11]. In this paper, we focus on the data association and mapping tasks using SSS images.

A. Data Association Using Side-scan Sonar Images

In earlier work on matching features across SSS images, Tena Ruiz et al. [12], Fallon et al. [7], and Issartel et al. [13] manually extract features for matching. However, this is not conducive to achieving full autonomy in AUVs. In the context of automatic feature extraction and matching, Ye et al. [14] use nonlinear diffusion filtering to construct a nonlinear scale space, achieving a match rate superior to advanced feature-matching algorithms like KAZE, SIFT, SURF, and ORB. Furthermore, in the field of deep learning, Lindenberger et al. [15] propose LightGlue, which predicts strong matches. However, these optical feature descriptors only maintain intensity consistency and geometric invariance in sonar images when there is a small viewpoint change. This makes them prone to failure in significant viewpoint changes [16], such as loop closure.

In recent years, there has been a growing trend towards incorporating geometric information into image features for data association. For example, MacKenzie et al. [17] use the Lambertian reflectance model to compute the elevation gradient of landmarks, which aids in data association. However, this approach requires significant computational resources when performing a large number of feature associations. Petrich et al. [8] propose a method that utilizes a flat surface assumption and intensity data to construct 4D landmarks. They then employ the Iterative Closest Point (ICP) algorithm for data association. However, using the flat surface assumption in complex underwater terrains can result in inaccurate spatial positions of landmarks, leading to erroneous associations. Zhang et al. [9] integrate the FAST corner detector with SIFT descriptors, utilizing nearest-neighbor search along with sliding compatibility check to enhance the matching process. However, optical descriptors yield a small number of matches.

Our proposed NDA does not require complex computations such as calculating descriptors or using reflectance models. Instead, it leverages the side-scan sonar geometric models to perform epipolar line search and provides a large number of correct matches in the scenario of large viewpoint changes.

B. 3D Mapping Using Side-scan Sonar

Side-scan sonar cannot obtain the elevation angle of landmarks. Therefore, researchers such as Fallon et al. [7] and Bernicola [18] assume a flat seabed surface to estimate the elevation of landmarks. However, Burguera et al. [10] demonstrate that the flat surface assumption can lead to mapping error.

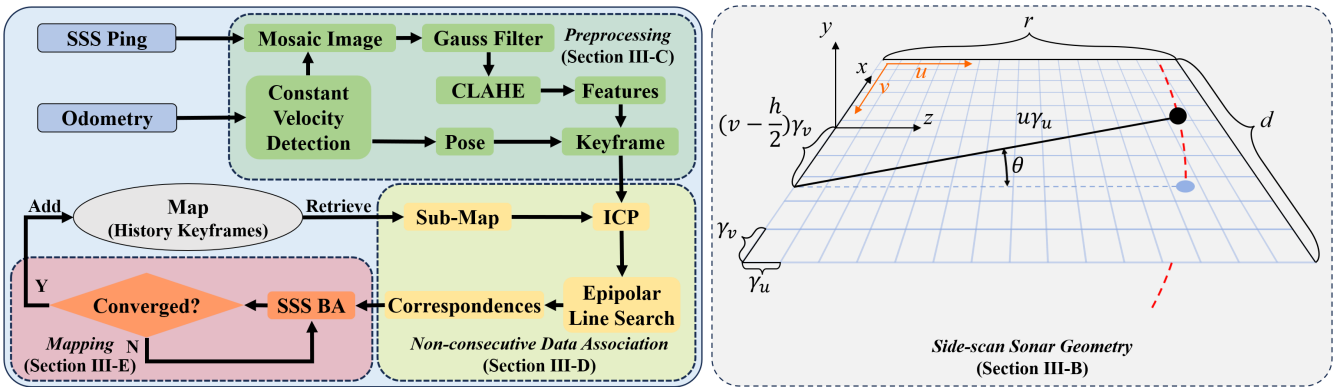
Shang et al. [19] combine depth measurement data with seabed reflectance models to achieve high-resolution and high-accuracy reconstruction of seabed terrain. However, this method requires consideration of the seafloor's reflectivity and the incidence angle of the wavefront, rendering high-accuracy reconstruction in waters with complex underwater topography unreliable. Huang et al. [20] propose Acoustic Structure from Motion (ASFM), which uses multiple forward-looking sonar (FLS) images to recover the 3D position of landmarks. However, the geometry model behind ASFM has not been extended to side-scan sonar.

Inspired by ASFM, we propose a side-scan sonar bundle adjustment (SSS BA) method based on the projection and back-projection geometric models of SSS. Our method allows for the optimization of the elevation angles of landmarks, thereby establishing an accurate 3D map, surpassing the limitations of the flat surface assumption.

III. Methodology

A. Framework Overview

Our workflow is shown in Fig. 2(a). SSS pings and odometry data are input into the preprocessing module, which generates keyframes containing pose and feature points.



(a) Overall pipline

(b) Side-scan sonar geometry

Fig. 2. System overview. (a) The overall pipeline includes three models. The preprocessing module takes SSS pings and odometry data as input, generating processed keyframes. The non-consecutive data association module establishes correspondences between historical keyframes and the current keyframe. The mapping module utilizes SSS BA to optimize the elevation angles of landmarks. Once convergence is achieved, the mapping model updates the keyframes in the map. (b) Side-scan sonar geometry. It is utilized in both the NDA module and the mapping module.

The current keyframe, along with historical keyframes, is then fed into the non-consecutive data association (NDA) module. The NDA module finds loop closure and associates feature points across non-consecutive keyframes. In the mapping module, the associated data is used for SSS BA, and once convergence is achieved, the map is updated with the optimized keyframe. The key algorithms in the NDA and mapping modules utilize our proposed side-scan sonar geometric model, as shown in Fig. 2(b). Firstly, we establish the side-scan sonar geometric model in Sect. III-B. Then, in Sect. III-C, we introduce the preprocessing module. The NDA module is described in Sect. III-D, followed by the mapping module in III-E.

B. Side-scan Sonar Geometry

We assume that the AUV is traveling at a constant velocity over a short period of time. A side-scan sonar image is constructed by assembling $2k+1$ pings ($k = 50$ in our framework). The pose of the keyframe is determined by the sonar's pose at the time the $(k+1)$-th ping is collected. Based on these, we establish the geometric model of the side-scan sonar, as shown in Fig. 2(b) ($k = 5$ in the figure).

The side-scan sonar image can be viewed as the projection of all observed data onto the plane with an elevation angle of 0 in the sonar's coordinate system, and it is scaled along the u and v axes of the image coordinate. We define the scaling factor γ_u for the u axis and γ_v for the v axis of the side-scan sonar image as:

$$\begin{bmatrix}\gamma_u \\ \gamma_v\end{bmatrix} = \begin{bmatrix}\frac{r}{w} \\ \frac{d}{h}\end{bmatrix} \tag{1}$$

where r is the maximum detection range of the side-scan sonar. w is the width of the image, which is the number of sampling points in each ping. d is the distance traveled by the AUV, provided by odometry. h is the height of the image, which is the number of pings used to generate the mosaic image.

The pixel coordinates of a feature point are defined as $p = [u, v]^\top$, and the coordinates of the corresponding 3D landmark in the sonar coordinate system are defined as $P = [X, Y, Z]^\top$. The relationships between them are:

$$P = \begin{bmatrix}X \\ Y \\ Z\end{bmatrix} = \pi^{-1}(p, \theta) = \begin{bmatrix}(v - \frac{h}{2})\gamma_v \\ u\gamma_u \sin\theta \\ u\gamma_u \cos\theta\end{bmatrix} \tag{2}$$

$$p = \begin{bmatrix}u \\ v\end{bmatrix} = \pi(P) = \begin{bmatrix}\sqrt{\frac{Y^2+Z^2}{{\gamma_u}^2}} \\ \frac{X}{\gamma_v} + \frac{h}{2}\end{bmatrix} \tag{3}$$

where $\pi^{-1}(p, \theta)$ is the back-projection function that back-projects pixel coordinates in a 2D image to 3D coordinates in the sonar coordinate system. Conversely, the projection function $\pi(P)$ projects 3D coordinates of a landmark in the sonar coordinate system to 2D coordinates in the image. θ represents the elevation angle of the landmark in the sonar coordinate system. Since θ cannot be directly obtained from the measurement, an initial guess can be obtained based on the flat surface assumption.

We assume that there is a reference frame and a current frame. Given the pixel coordinates p_r as well as the elevation angle θ_r of a point in the reference frame, and the transformation $T_r^c \in SE(3)$ from the reference frame to the current frame, we can combine Eqs. 2 and 3 to derive the reprojection model as:

$$p_c = \pi(T_r^c \pi^{-1}(p_r, \theta_r)) \tag{4}$$

where p_c represents the pixel coordinates of the point being reprojected onto the current frame.

C. Preprocessing

The preprocessing module generates mosaic images from the raw data, extracts feature points, computes initial estimates of landmarks' elevation angles, and sets poses of keyframes.

The preprocessing module receives odometry data to obtain the AUV's motion state. When the AUV travels at a constant velocity, the module mosaics a fixed number of pings together to form the mosaic image. The image is

then smoothed using Gaussian filtering. Contrast Limited Adaptive Histogram Equalization (CLAHE) [21] is applied to enhance the contrast between low-intensity flat seafloor and high-intensity landmarks. Next, the Sobel operator [22] is used to compute the u-axis gradient $d_u(p)$ and the v-axis gradient $d_v(p)$ for all pixels. The image is then thresholded with a threshold value I_{th}. Finally, a score $W(p)$ is computed for all pixels:

$$W(p) = \begin{cases} d_u(p)^2 + d_v(p)^2, & \text{if } I(p) \geq I_{th} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

According to the scores, all the points are sorted and the top 500 points with the highest scores are selected as feature points. The initial estimate of the elevation angle θ_{init} for the corresponding landmark in the sonar coordinate system is computed by the flat surface assumption as:

$$\theta_{init} = \arccos \frac{H \cos\theta_r \cos\theta_p}{R} - \theta_r + \theta_s \tag{6}$$

$$R = u\gamma_u \tag{7}$$

where H is the height of the AUV to the seafloor, which can be measured by a Doppler Velocity Log (DVL) or estimated from the image [23]. θ_r is the roll angle of the AUV, and θ_p represents the pitch angle. Both angles can be obtained from odometry. θ_s is the mounting angle of the side-scan sonar. R is the range measurement of the landmark, which can be computed using Eq. 7. The pose of the keyframe is set according to odometry.

At this point, the preprocessing module outputs a complete keyframe, which includes the pose of AUV, a preprocessed image and 500 feature points with initial elevation angles of corresponding landmarks.

D. Non-consecutive Data Association

The non-consecutive data association (NDA) module performs loop closure detection and associates current keyframes with historical keyframes.

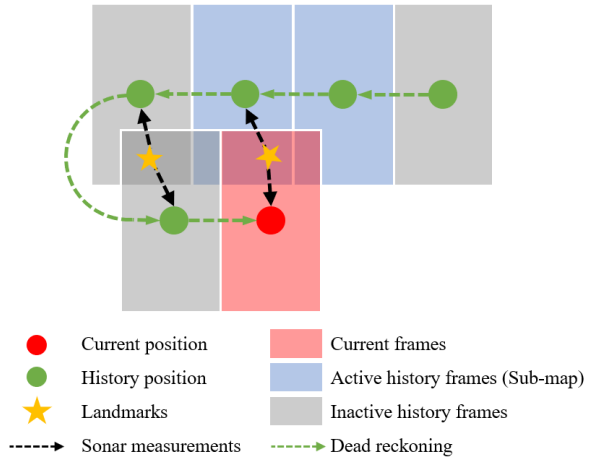


Fig. 3. The strategy of loop closure detection and data association.

The NDA module queries the historical frames' poses that may potentially result in loop closures with the current frame by constructing a KD-tree of the historical frame poses. Historical frames with co-visibility relationships are activated as a sub-map, as shown in Fig. 3. All the feature points of the queried historical frames and the current frame are backprojected to their respective sonar coordinate systems as landmarks using Eq. 2. Based on the relative transformation of two keyframes, the 3D landmarks of the historical frames are transformed into the current frame. Then, the Iterative Closest Point (ICP) registration [24] is performed to eliminate the transformation error approximately, preparing for the calculation of epipolar lines.

The model selects the historical frame from the sub-map based on the criterion of having the largest co-visibility area with the current frame. The uncertainty range of the elevation angle for each feature point in the historical frame is set to $5°$. The maximum and minimum elevation angles are then substituted into Eq. 4 to reproject onto the current frame, resulting in two reprojection points. The line segment formed by these two points is an epipolar line, which effectively approximates the projection of the short uncertain arc. By minimizing the intensity error along the epipolar line, we can find the coordinates of the matching point p_c^* as below:

$$p_c^* = \min_{p_c} \left(I_r - I(p_c)\right)^2 \tag{8}$$

where I_r is the intensity of the feature point in the reference frame. p_c represents the pixel coordinates along the epipolar line to be searched in the current frame. $I(p_c)$ represents the intensity of p_c.

Finally, the NDA module identifies the historical frame that forms a loop closure and establishes the correspondences between the two frames.

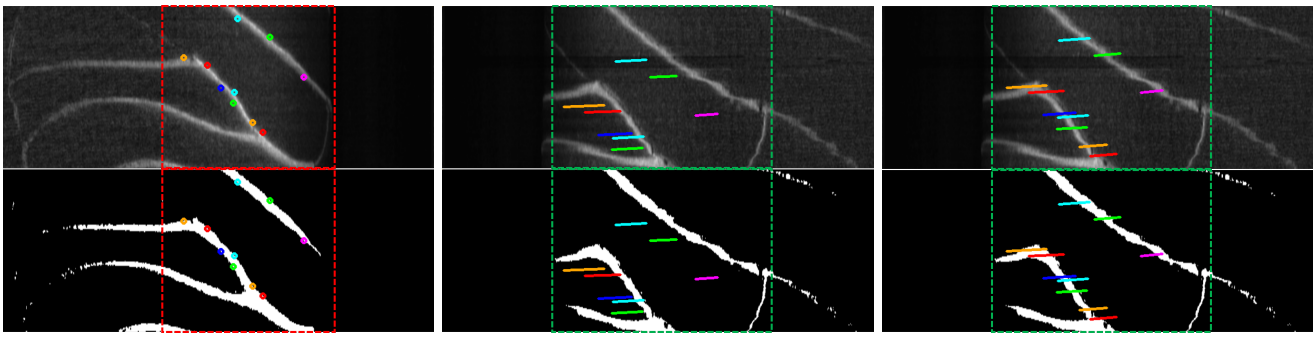
E. Mapping Using Side-scan Sonar Bundle Adjustment

The mapping module receives the correspondences provided by the NDA module. The model constructs the reprojection error according to the correspondences and iteratively optimizes to reduce it. We refer to this process as side-scan sonar bundle adjustment (SSS BA).

As shown in Fig. 3, when loop closure is found, the same landmark cannot appear in more than two keyframes. Therefore, we only perform SSS BA on the two keyframes. The feature points associated with the reference frame and the current frame should correspond to the same landmarks on the seafloor. Therefore, we reproject the feature points from the reference frame into the current frame. The distance between the reprojected points and the associated feature points in the current frame is the reprojection error. The reprojection error is minimized by adjusting the relative transformation and the elevation angles of the feature points. We utilize the reprojection error to construct the least-squares problem as follows:

$$\underset{X}{\operatorname{argmin}} \sum_i \|h_i(X) - z_i\|^2 \tag{9}$$

where the state vector $X = [T_r^c, \theta_0, \theta_1, ...]^\top$ contains all state variables: the relative transformation T_r^c between the two keyframes and the elevation angles θ_k of all landmarks



(a) Feature points in the reference frame (b) Epipolar lines in the current frame (Odom) (c) Epipolar lines in the current frame (Odom+ICP)

Fig. 4. Image preprocessing and epipolar line computation: The first row shows the original images, while the second row shows the images after preprocessing. The wireframe represents the co-visibility between the reference frame and the current frame. (a) Feature points are extracted in the reference frame. (b) Epipolar lines are calculated using the relative transformation provided by odometry (Odom). (c) Epipolar lines are calculated using the relative transformation optimized through ICP registration. The initial guess of the relative transformation is provided by odometry (Odom+ICP).

in the reference frame. $h_i(X) = \pi(T_r^c \pi^{-1}(p_i, \theta_i))$ is the reprojection function in Eq. 4, which reprojects the feature point of the i-th landmark in the reference frame's image into the current frame's image. z_i is the measurment of the i-th landmark in the current frame's image. The optimization is performed using the Ceres optimizer and the Levenberg-Marquardt algorithm [25]. After convergence, the elevation angle θ_r of the landmark in the reference frame is optimized. The elevation angle θ_c of the landmark in the current frame can be calculated as below:

$$P_c = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = T_r^c \pi^{-1}(p_r, \theta_r) \tag{10}$$

$$\theta_c = \arctan \frac{Y_c}{Z_c} \tag{11}$$

where p_r is the measurement of the landmark in the reference frame, and $P_c = [X_c, Y_c, Z_c]^\top$ are the coordinates of the landmark in the current frame.

Thus, the mapping module recovers the elevation angles of the corresponding landmarks in the current frame and the reference frame. Meanwhile, these landmarks contribute to the creation of an accurate sparse 3D map.

IV. Experiment

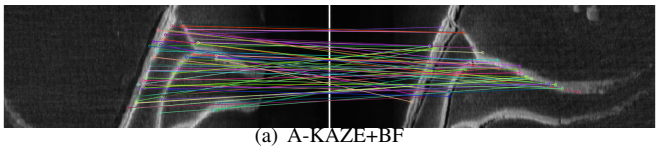
To evaluate the performance of our framework, we collect dataset in a 700m×700m underwater terraced fields. The underwater terraced fields provide rich features, making them suitable for validating the data association and 3D mapping performance in uneven underwater terrains. Our experimental platform uses an AUV with MSS300 multibeam side-scan sonars mounted on both sides. The sonar has a horizontal beam angle of 0.22°, a vertical beam angle of 40°, a maximum detection range of 100 m, and an output frequency of 5 Hz. The AUV is equipped with an integrated inertial navigation system that combines GPS, IMU, depth sensor, and Doppler Velocity Log (DVL) to provide odometry data at a frequency of 1 Hz. SSS pings and odometry data are used as input data for the framework. Since GPS signals cannot be received underwater, the AUV navigates on the water surface. We conduct experiments on the underwater terraced fields dataset. The experimental results demonstrate the effectiveness of each module within our framework and provide a quantitative assessment of the NDA and SSS BA.

A. Preprocessing and Non-consecutive Data Association

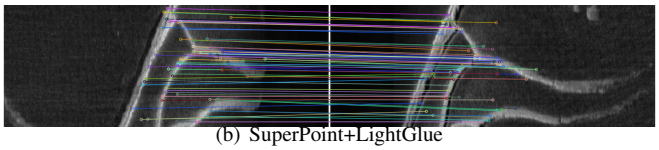
Fig. 4 illustrates the effectiveness of our preprocessing module and how the NDA model works with ICP as well as epipolar line search. The first row of Fig. 4 shows the original images, while the second row shows the images after preprocessing. The reference frame (see Fig. 4(a)) and the current frame (see Fig. 4(b) and Fig. 4(c)) have a co-visibility area.

The original images show that the terrain features under different perspectives exhibit geometric distortions and inconsistent intensity. In contrast, our preprocessing algorithm effectively segments high-intensity underwater features from low-intensity planar areas and performs binarization. Then, it extracts points with significant gradients along the edges of the features in the reference frame, as shown in Fig. 4(a). This process ensures that the features of the same landmark in different perspective images satisfy the intensity consistency, providing favorable conditions for data association.

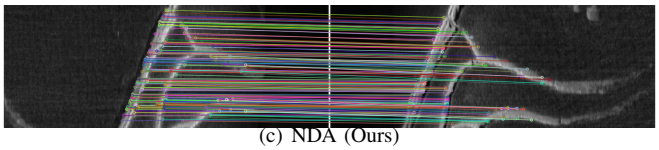
Fig. 4(b) demonstrates the epipolar lines calculated in the current frame using the relative transformation provided by odometry. Since there are transformation error, sensor extrinsic calibration error, and non-strict alignment of timestamps, the calculated epipolar lines do not correctly intersect the white feature area, making it difficult to find correct matches along the epipolar lines. Fig. 4(c) illustrates the epipolar lines calculated using relative transformation corrected by ICP registration. The epipolar lines now accurately intersect the contours of the white feature area. Although the landmarks correlated with feature points may not have accurate spatial positions due to the initialization of the flat surface assumption, the ICP registration still effectively reduces rotational error and translation error along the x-axis in the relative transformation between the two frames. This ensures that



(a) A-KAZE+BF



(b) SuperPoint+LightGlue



(c) NDA (Ours)

Fig. 5. Data association of side-scan sonar images using different methods in a loop closure detection scenario.

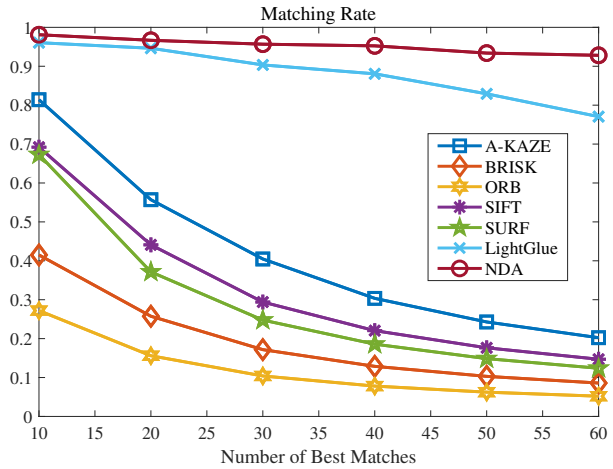
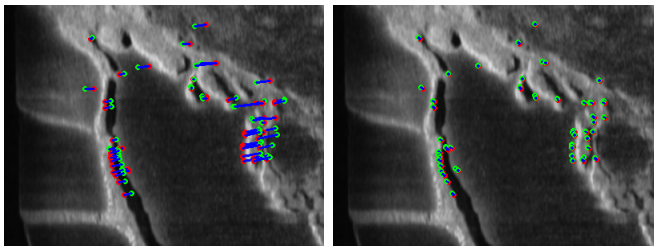


Fig. 6. Matching rate of different methods.

the epipolar line is calculated at the correct position in the current frame.

Fig. 5 illustrates the matching results of A-KAZE, LightGlue, and NDA in a loop closure detection scenario. A-KAZE generates a large number of false matches due to it focuses on local feature regions. This issue becomes significant when the feature distributions differ greatly with large viewpoint changes. LightGlue computes and matches features globally, resulting in a much higher matching rate compared to A-KAZE. However, LightGlue has few matches near the shadow zone and produces unreliable matches in sparsely featured flat underwater areas. This is because LightGlue is trained on camera images and cannot accurately predict the matching relationships of feature points based on sonar geometric changes, particularly in the presence of significant transformation. In contrast, our method utilizes the geometry of the side-scan sonar to calculate epipolar lines, which restricts the search range based on the uncertain range of elevation angles. Moreover, the matching is performed on the preprocessed images, which ensures the



(a) Initial reprojection error (b) Reprojection error after SSS BA

Fig. 7. Reprojection error: The red dots represent the feature points in the current frame, while the green dots represent the reprojection of the feature points of the reference frame. The blue lines depict the correspondences of them. The length of the blue lines is proportional to the reprojection error.

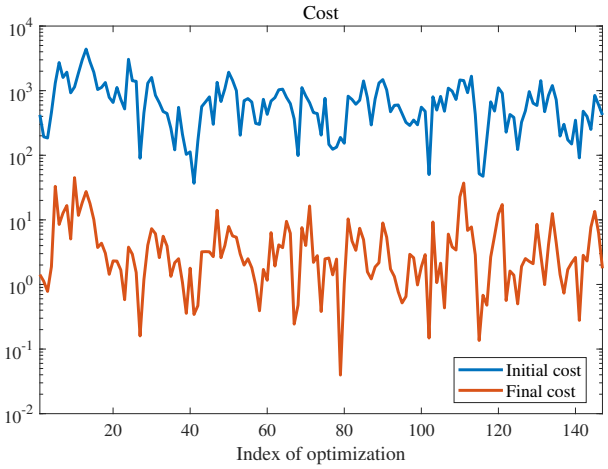


Fig. 8. Initial and final cost of SSS BA.

intensity consistency of corresponding features. Therefore, the NDA can provide a large number of correct matches.

To quantitatively evaluate the matching performance of different methods, we run our framework on the dataset and performed matching using different approaches at each loop closure. For each method, we select the top n pairs of matches with the best matching results and calculate the matching rate. In NDA, we select the top n pairs of matches based on the feature point scores calculated using Eq. 5. In LightGlue, we use SuperPoint to extract feature points and select the top n pairs of matches based on the highest confidence scores. Other algorithms use the brute-force matcher and select the top n pairs of matches with the lowest Hamming distances. Fig. 6 shows the matching rate of different algorithms in the loop closure detection scenario throughout the dataset. Our proposed NDA ensures a high matching rate (over 90%) even with an increasing number of matching pairs, and achieves the highest matching rate among the methods.

B. Mapping

We run our framework on the entire dataset, performing SSS BA with 3 iterations of LM optimization. Fig. 7(a) shows the reprojection error before optimization, where the error is proportional to the length of the blue lines. Fig.

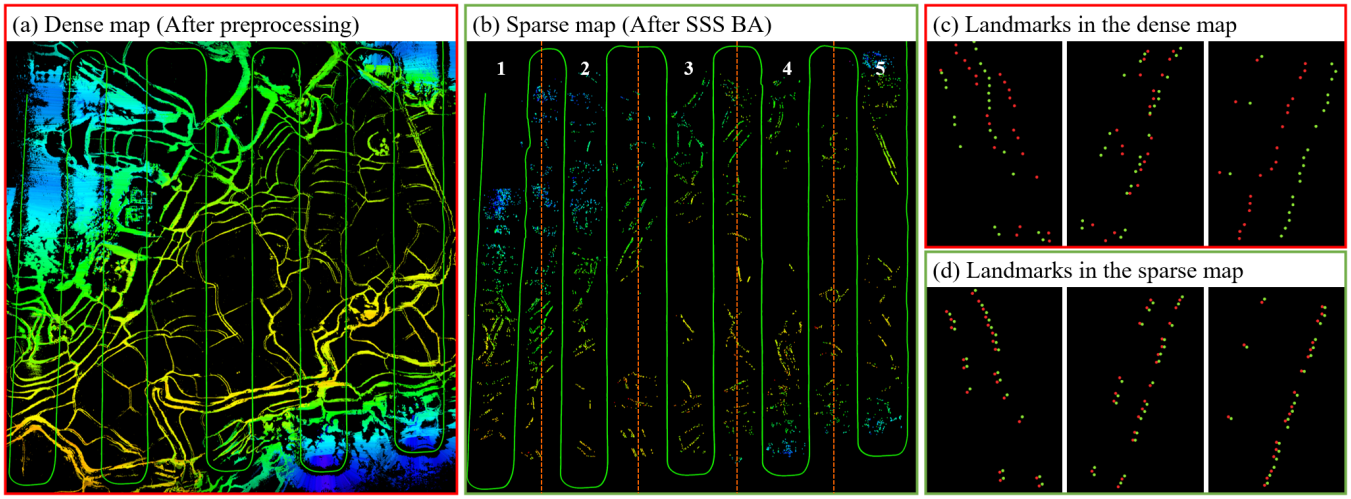


Fig. 9. Mapping Results: (a) The preprocessing module uses the flat surface assumption to produce a dense map. The dense map is inaccurate but affords a clear depiction of the underwater topography. (b) Based on the preprocessing, the framework uses NDA to obtain correspondences and optimizes landmarks through SSS BA. This results in a more accurate sparse map. The sparse map is divided into five sessions. (c) Some landmarks of the dense map. Landmarks of the same color belong to the same keyframe. (d) The landmarks of (c) are optimized by SSS BA in the sparse map.

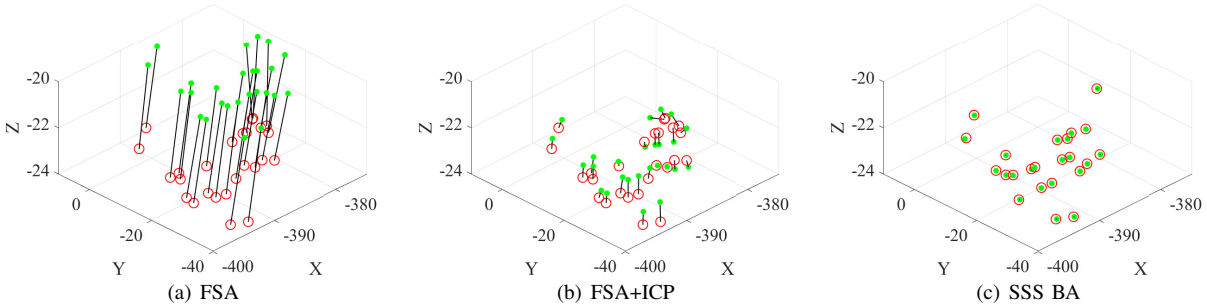


Fig. 10. The 3D mapping error in partial underwater area: The red circular rings represent landmarks of the reference frame, while the green points represent landmarks of the current frame. The solid black lines depict correspondences between the landmarks of the two frames. The length of the lines is proportional to the mapping error.

TABLE I

COMPARISON OF MAPPING ERROR. THE MEANS (μ) AND STANDARD DEVIATIONS (σ) OF MAPPING ERROR ARE GIVEN ($\mu[\sigma]$). METRICS ARE EXPRESSED IN METERS.

Sequence	FSA	FSA+ICP	SSS BA
session1	2.463[1.864]	1.673[1.215]	**0.155[0.178]**
session2	4.577[2.418]	1.513[1.294]	**0.132[0.185]**
session3	3.761[1.547]	1.560[1.500]	**0.124[0.014]**
session4	3.193[1.312]	1.624[1.714]	**0.120[0.018]**
session5	5.000[4.181]	1.620[2.118]	**0.141[0.118]**
average	3.799[2.264]	1.598[1.568]	**0.134[0.129]**

7(b) shows a significant reduction in the reprojection error after optimization. The initial and final cost of SSS BA are shown in Fig. 8. Within the entire dataset, the average initial optimization cost is 764.17, which decreases to an average of 4.91 after 3 iterations of optimization. This demonstrates that our SSS BA can rapidly reduce the reprojection error to optimize.

Next, we evaluate the 3D mapping accuracy. Since ground truth of the seabed topography is not available, we back-project the feature points to the sonar coordinate system as 3D landmarks using Eq. 2, and transform them to the world coordinate system based on the poses of AUV. The corresponding landmarks across different keyframes should exhibit overlap when viewed in the world coordinate system. We run our framework on the entire dataset and the mapping results are shown in Fig. 9. Firstly, we disable the NDA model and SSS BA. The keyframes from the preprocessing model are directly incorporated into the map. This results in a dense map (see Fig. 9(a)), where the elevation angles of landmarks are calculated solely based on the flat surface assumption (FSA). The dense map is not accurate but provides a clear representation of underwater topography. Then, we enable all models of our framework. SSS BA utilizes the correspondences provided by the NDA model to optimize the elevation angles of landmarks, ultimately generating a sparse map (see Fig. 9(b)). The map generated solely using the flat surface assumption does not adjust the 3D positions of landmarks. This leads to poor alignment of point clouds, resulting in a lack of overlap between these point clouds, as shown in Fig. 9(c). In contrast, with the application of

SSS BA, optimizing landmark positions allows a clearer alignment of point clouds, as shown in Fig. 9(d).

We quantitatively evaluate the mapping error using the Euclidean distance between the corresponding landmarks, where a smaller distance indicates higher accuracy. Fig. 10(a) shows that the landmarks generated by the flat surface assumption (FSA) exhibit a significant mapping error. After applying ICP to register the point clouds of two frames, the mapping error decreases, as shown in Fig. 10(b). However, since ICP only adjusts the poses of keyframes and does not optimize the spatial positions of the landmark points, the mapping error is still noticeable. In contrast, SSS BA recovers the unknown elevation angles of the landmarks. This results in corresponding landmarks nearly overlapping, as shown in Fig. 10(c), indicating higher mapping accuracy. The mapping error for the three methods in the five sessions shown in Fig. 9(b) are presented in Table I. Compared to the flat surface assumption, our SSS BA significantly reduces the average mapping error from 3.799 to 0.134.

V. Conclusion

In this paper, we propose a framework that takes advantage of the geometric models of side-scan sonar for underwater 3D mapping. It establishes an interconversion relationship between two-dimensional pixels and three-dimensional landmarks, utilizing epipolar line search for non-consecutive data association and minimizing reprojection error to recover the elevation angles of landmarks.

Our method can provide strong data associations in scenarios with large viewpoint changes and construct an accurate sparse map within uneven underwater terrains. Future work could extend our framework to a full SLAM system with factor graph optimization and facilitate the creation of denser, high-accuracy underwater maps.

References

[1] A. Bodenmann, J. Cappelletto, M. Massot-Campos, D. Newborough, E. Chaney, R. Marlow, R. Templeton, A. B. Phillips, B. J. Bett,

[2] R. B. Wynn, V. A. Huvenne, T. P. Le Bas, B. J. Murton, D. P. Connelly, B. J. Bett, H. A. Ruhl, K. J. Morris, J. Peakall, D. R. Parsons, E. J. Sumner, S. E. Darby, R. M. Dorrell, and J. E. Hunt, "Autonomous Underwater Vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience," *Marine Geology*, vol. 352, pp. 451–468, Jun. 2014.

[3] D. Roper, C. A. Harris, G. Salavasidis, M. Pebody, R. Templeton, T. Prampart, M. Kingsland, R. Morrison, M. Furlong, A. B. Phillips, and S. McPhail, "Autosub Long Range 6000: A Multiple-Month Endurance AUV for Deep-Ocean Monitoring and Survey," *IEEE Journal of Oceanic Engineering*, vol. 46, no. 4, pp. 1179–1191, Oct. 2021.

[4] S. Soylu, P. Hampton, T. Crees, A. Woodroffe, and E. Jackson, "Sonar-Based SLAM Navigation in Flooded Confined Spaces with the IMOTUS-1 Hovering AUV," in *2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV)*. Porto, Portugal: IEEE, Nov. 2018, pp. 1–6.

[5] S. Reed, I. T. Ruiz, C. Capus, and Y. Petillot, "The fusion of large scale classified side-scan sonar image mosaics," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 15, no. 7, 2006.

[6] P. Blondel, *The handbook of sidescan sonar*. Springer Science & Business Media, 2010.

C. Wardell, and B. Thornton, "High-resolution visual seafloor mapping and classification using long range capable AUV for ship-free benthic surveys," in *2023 IEEE Underwater Technology (UT)*. Tokyo, Japan: IEEE, Mar. 2023, pp. 1–6.

[7] M. F. Fallon, M. Kaess, H. Johannsson, and J. J. Leonard, "Efficient AUV navigation fusing acoustic ranging and side-scan sonar," in *2011 IEEE International Conference on Robotics and Automation*. Shanghai, China: IEEE, May 2011, pp. 2398–2405.

[8] J. Petrich, M. F. Brown, J. L. Pentzer, and J. P. Sustersic, "Side scan sonar based self-localization for small Autonomous Underwater Vehicles," *Ocean Engineering*, vol. 161, pp. 221–226, Aug. 2018.

[9] J. Zhang, Y. Xie, L. Ling, and J. Folkesson, "A fully-automatic side-scan sonar simultaneous localization and mapping framework," *IET Radar, Sonar & Navigation*, vol. n/a, no. n/a, 2023.

[10] A. Burguera and G. Oliver, "High-Resolution Underwater Mapping Using Side-Scan Sonar," *PLOS ONE*, vol. 11, no. 1, p. e0146396, Jan. 2016.

[11] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.

[12] I. Tena Ruiz, S. De Raucourt, Y. Petillot, and D. Lane, "Concurrent mapping and localization using sidescan sonar," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 2, pp. 442–456, Apr. 2004.

[13] M. M. Issartel and D. D. GuÃ, "Robust SLAM for side scan sonar image mosaicking," in *OCEANS 2017*, 2017.

[14] X.-F. Ye, P. Li, J.-G. Zhang, J. Shi, and S.-X. Guo, "A feature-matching method for side-scan sonar images based on nonlinear scale space," *Journal of Marine Science and Technology*, vol. 21, pp. 38–47, 2016.

[15] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "LightGlue: Local Feature Matching at Light Speed," in *ICCV*, 2023.

[16] S. Gode, A. Hinduja, and M. Kaess, "Sonic: Sonar image correspondence using pose supervised learning for imaging sonars," *arXiv preprint arXiv:2310.15023*, 2023.

[17] C. M. MacKenzie, M. L. Seto, and Y. Pan, "Extracting seafloor elevations from side-scan sonar imagery for SLAM data association," in *2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2015, pp. 332–336.

[18] L. Bernicola, D. Guériot, and J.-M. Le Caillec, "A hybrid registration approach combining SLAM and elastic matching for automatic side-scan sonar mosaic," in *2014 Oceans - St. John's*, Sep. 2014, pp. 1–5.

[19] X. Shang, J. Zhao, H. Zhang, A. Wang, and X. Wang, "Integration of SSS-Based Reconstructed Results and Bathymetric Data to Obtain High-Resolution and High-Accuracy Underwater Topography," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[20] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Hamburg, Germany: IEEE, Sep. 2015, pp. 758–765.

[21] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, pp. 35–44, 2004.

[22] N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the sobel operator," *IEEE Journal of solid-state circuits*, vol. 23, no. 2, pp. 358–367, 1988.

[23] A. Burguera and G. Oliver, "Intensity correction of side-scan sonar images," in *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*. IEEE, 2014, pp. 1–4.

[24] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[25] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis: proceedings of the biennial Conference held at Dundee, June 28–July 1, 1977*. Springer, 2006, pp. 105–116.