remote sensing

MDPI

Review

An Overview of Key SLAM Technologies for Underwater Scenes

Xiaotian Wang [1,*], Xinnan Fan [2], Pengfei Shi [2], Jianjun Ni [2] and Zhongkai Zhou [2]

[1] School of Computer and Information, Hohai University, Nanjing 210000, China
[2] School of Information Science and Engineering, Hohai University, Changzhou 213002, China; fanxn@hhuc.edu.cn (X.F.)
[*] Correspondence: wxthhu@hhu.edu.cn

Abstract: Autonomous localization and navigation, as an essential research area in robotics, has a broad scope of applications in various scenarios. To widen the utilization environment and augment domain expertise, simultaneous localization and mapping (SLAM) in underwater environments has recently become a popular topic for researchers. This paper examines the key SLAM technologies for underwater vehicles and provides an in-depth discussion on the research background, existing methods, challenges, application domains, and future trends of underwater SLAM. It is not only a comprehensive literature review on underwater SLAM, but also a systematic introduction to the theoretical framework of underwater SLAM. The aim of this paper is to assist researchers in gaining a better understanding of the system structure and development status of underwater SLAM, and to provide a feasible approach to tackle the underwater SLAM problem.

Keywords: underwater vehicles; SLAM; vision sensors; acoustic sensors; deep learning

check for updates

Citation: Wang, X.; Fan, X.; Shi, P.; Ni, J.; Zhou, Z. An Overview of Key SLAM Technologies for Underwater Scenes. *Remote Sens.* **2023**, *15*, 2496. https://doi.org/10.3390/rs15102496

Academic Editors: Yan-Tsung Peng, Wenqi Ren, Jingchun Zhou and Qiuping Jiang

Received: 28 March 2023
Revised: 2 May 2023
Accepted: 4 May 2023
Published: 9 May 2023

1. Introduction

Since the beginning of the 21st century, the importance of applied research on autonomous positioning technology for mobile robots has grown. A global navigation satellite system (GNSS) is an effective and accurate solution for the robots' own positioning and movement trajectory [1]. However, in some environments where GPS is not available or when a priori information is insufficient, other solutions must be found. Simultaneous localization and mapping was introduced to the field of robotics, which entails the robot obtaining environmental information through sensors carried by itself in an unknown environment. During the process of navigation and localization, a structural consistency map of the surrounding environment is built. Compared to traditional localization methods, SLAM is a small-sized and cost-efficient instrumentation method [2,3]. Through the efforts of researchers, SLAM has been implemented in drones, sweeping robots, unmanned vehicles, smart wearable devices, and other devices [4]. Figure 1 below shows the application of SLAM in some scenarios.

Figure 1. The application of SLAM technology in different scenarios.

Unmanned underwater vehicles (UUVs) have become a popular option for underwater exploration due to their safety, portability, and cost-effectiveness. UUVs can be divided into two categories: remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) [5–7]. UUVs are mainly used for marine resource investigation, undersea biology research, underwater structure detection, and marine data collection. The precise positioning and navigation tasks of underwater vehicles are difficult because the underwater environment blocks radio signals such as GPS, and the inertial navigation approach is prone to accumulate errors underwater [8]. Conventional underwater positioning methods such as short baseline (SBL) and ultrashort baseline (USBL) require the installation of a base array with a receiver device in the target area or require a periodic position correction by the AUV [9]. Although these methods are useful, they are costly, and the range of exploration is limited by the beacons. In addition, the vehicles need to surface frequently and often face more exertion. To address these challenges, many researchers have started to investigate how SLAM techniques can be applied to the underwater domain [10,11], thus bringing new possibilities for autonomous positioning and navigation of underwater vehicles.

In 1986, SLAM was first introduced at the IEEE Robotics and Automation Conference to address the spatial uncertainty description and transformation representation. Cheeseman [12] pioneered the use of probabilistic estimation methods for robot localization and mapping, and subsequently, this field of research has been pursued by numerous researchers. SLAM technology has likely gone through three stages of development. From 1986 to 2004, the classical era of SLAM development was mainly focused on the proposal of probabilistic estimation methods, such as the extended Kalman filter (EKF) [13], the particle filter (PF) [14], and the maximum likelihood estimation (MLE) [15]. From 2004 to 2015, SLAM entered the era of algorithm analysis, which included the study of algorithm observability, convergence, sparsity, and consistency. Since 2016, the field has shifted to address the robustness of SLAM [16]. This research includes algorithm robustness, scalability, efficient algorithms under resource constraints, high-level perception, and algorithm adaptiveness. More recently, with the emergence of computer vision and deep learning methods, there has been a shift towards SLAM methods based on deep learning [17–19].

In recent years, the widespread deployment of vision sensors and the rapid advancement of computer vision have led to the emergence of visual SLAM as an alternative to traditional SLAM approaches [20]. Compared to sonar, laser, and infrared range sensors, visual sensors have several advantages in terms of cost, use, and information acquisition capability, and have demonstrated a high accuracy in scenarios with adequate lighting and texture. Depending on the camera type, visual SLAM can be classified into monocular SLAM [21,22], stereo SLAM [23,24], and RGB-D SLAM [25,26]. Additionally, depending on the implementation scheme, visual-based SLAM methods can be divided into feature-based methods (e.g., ORB-SLAM [27–29]), direct methods (e.g., large-scale direct monocular SLAM (LSD-SLAM) [30]), and semidirect methods (e.g., semidirect visual ranging (SVO)) [31,32]. Of the existing systems, ORB-SLAM is a popular visual SLAM solution which utilizes Oriented FAST and rotation BRIEF [33] key-point detectors to match features across successive images. This key-frame-based approach was derived from prior SLAM variants (PTAM [34], DT-SLAM [35]). ORB-SLAM has been tested with mobile robots in various scenarios and has achieved promising results.

In 2000, Williams et al. [36] presented the results of applying a simultaneous localization and map building (SLAM) algorithm to estimate the motion of a submersible vehicle. Scans obtained from an on-board sonar were processed to extract stable point features in the environment, thus constructing a map of the environment and estimating the vehicle's location. This work was the first instance of a deployable underwater implementation of the SLAM algorithm. However, underwater scenes are typically unstructured and difficult to navigate due to the presence of illumination, texture, turbidity, and hydrodynamics. This makes it difficult for camera sensors to accurately extract features, and many land-based algorithms cannot be applied to underwater scenes. To obtain reliable estimates, SLAM for underwater scenes typically requires the use of specialized sensors such as inertial measurement units (IMU), Doppler velocity log (DVL), and depth sensors [37]. Acoustic and laser sensors are also widely used for sensing the underwater environment [38,39]. The fusion of different sensors has become an important approach in solving underwater SLAM. Despite its relative maturity in traditional scenarios, its application in underwater robot localization and navigation is still in its development stage and is a complex research area. Figure 2 illustrates some SLAM algorithms that can be used in underwater environments.

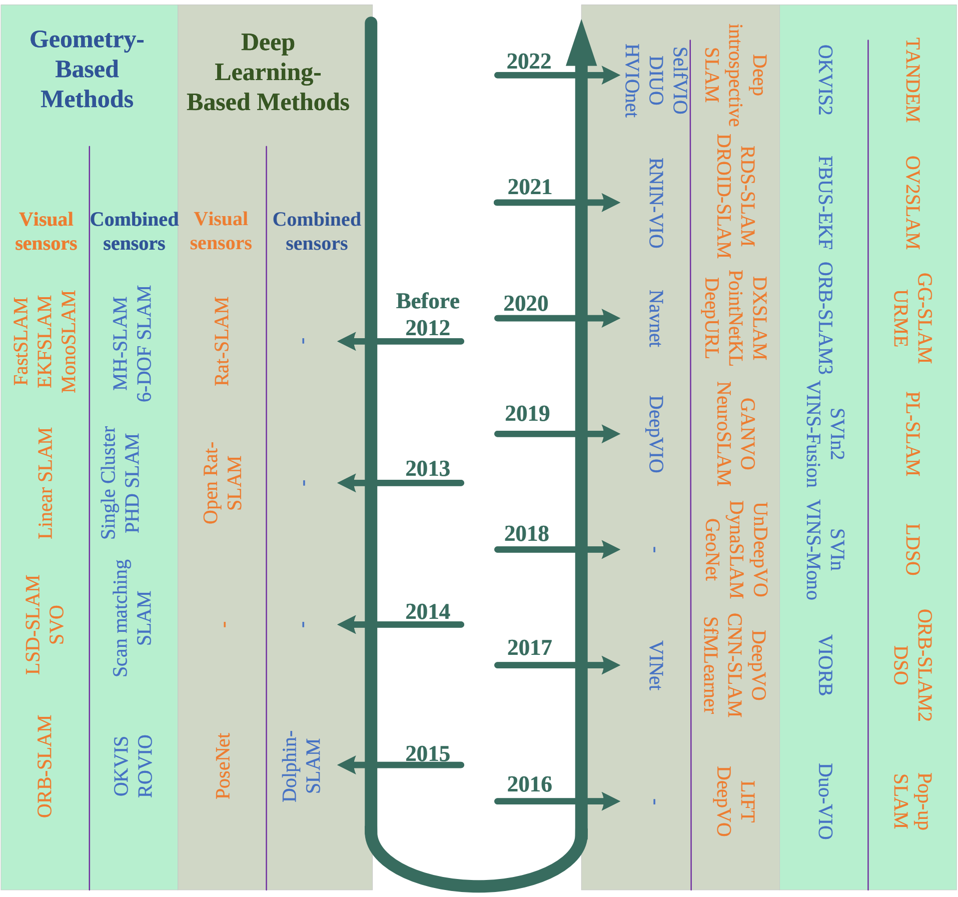

Figure 2. Partial list of SLAM algorithms that have been used in underwater environments in the last decade.

In recent years, underwater SLAM technology has developed rapidly, playing an important role in underwater positioning and surveying. However, it still lacks a systematic evaluation. Therefore, after carefully reviewing the relevant literature of underwater SLAM, this paper summarizes the development of this field. We chose recent related research papers in the underwater environment according to the basic framework, research focus, and development direction of underwater SLAM. The main contributions of this paper are as follows: 1. the development of underwater SLAM is introduced systematically; 2. from the framework of SLAM, the following parts of underwater visual SLAM are mainly introduced: sensors, front-end visual odometry, back-end state optimization, loop closure detection and mapping; 3. we collect the key points and difficult points that need to be solved in the development of underwater SLAM; 4. from the perspective of the application environment and technology development of underwater SLAM, the future development direction of this field is studied. This paper provides a comprehensive review of the key SLAM technologies for underwater robots. It is organized as shown in Figure 3. In Section 1, the research background about underwater robots, SLAM, and underwater SLAM is introduced. Section 2 compares recent underwater SLAM methods based on the theoretical framework of SLAM systems. Section 3 analyzes the main problems of current underwater SLAM. Section 4 makes an estimation of the future development trend and applications of underwater SLAM, in addition to which relevant datasets are presented and relevant experiments are conducted. Finally, the conclusion is summarized in Section 5. This paper provides a comprehensive review of the key technologies of underwater SLAM in terms of research background, field applications, theoretical framework, existing methods, problems, and future trends. It is hoped that this work can provide guidance and help to researchers in related fields.

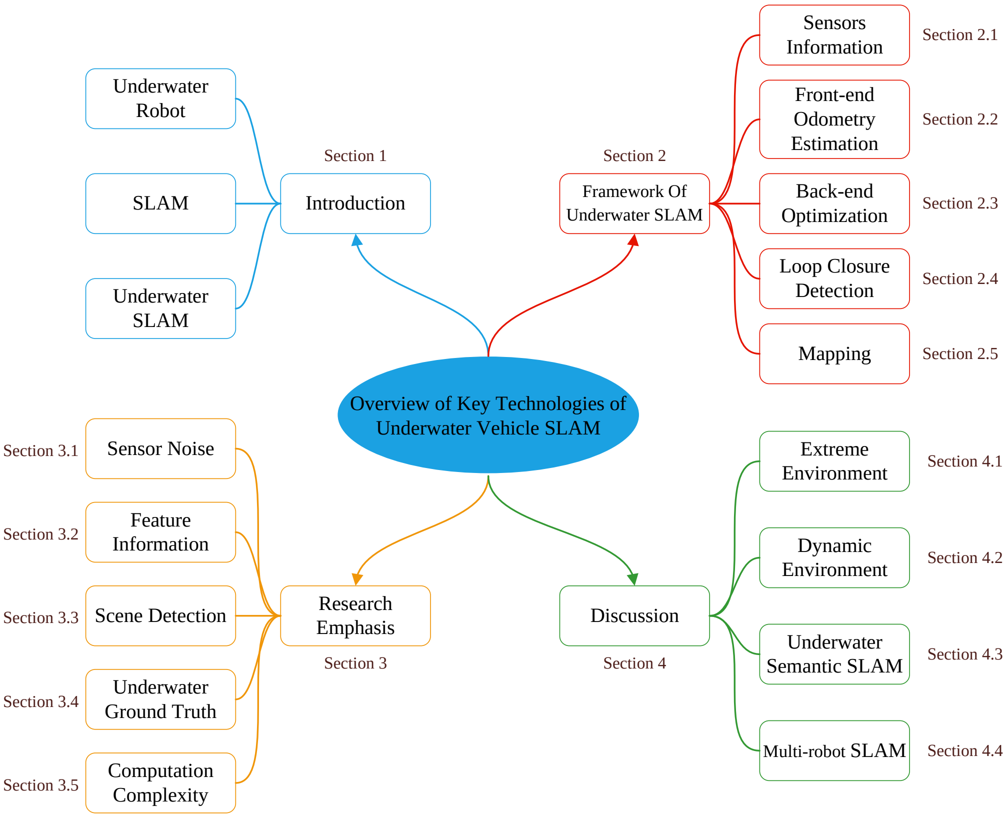

Figure 3. Structure diagram for underwater simultaneous localization and mapping of this paper.

2. Framework of Underwater SLAM

A simultaneous localization and mapping system consists of three processes: perception, localization, and mapping. Proprioception sensors and exteroception sensors are used to obtain information about the environment, thus allowing the robot to estimate and locate its own position and pose. Following this, a map of the environment is generated. The SLAM problem is essentially a state estimation problem, which is mathematically expressed in terms of equations of motion and equations of observation.

$$\begin{cases} x_k = f(x_{k-1}, u_k, w_k) \\ z_{k,j} = h\left(y_j, x_k, v_{k,j}\right) \end{cases} \tag{1}$$

where x_k denotes the robot position at time k, u_k denotes the robot input at time k, w_k denotes the noise at time k, $z_{k,j}$ is the observation data at time k for the jth waypoint, y_j denotes the jth waypoint, and $v_{k,j}$ is the noise at time k for the jth waypoint. Figure 4 is a schematic of the visual slam process.

Visual SLAM is a commonly used sensing method for underwater robots and is one of the main topics of this paper [40]. Figure 5 presents the basic theoretical framework of visual SLAM, which consists of sensor data, front-end, back-end, loop-closure detection, and mapping. Sensor data are collected and sent to the front end, which uses a visual odometer for interframe motion estimation. Loop closure is used to reduce the cumulative error and drift caused by the accumulation of sensor noise and model errors. The back end processes the camera positional and loop-closure detection information measured by the visual odometer at different moments and optimizes it to obtain globally consistent trajectories and maps.

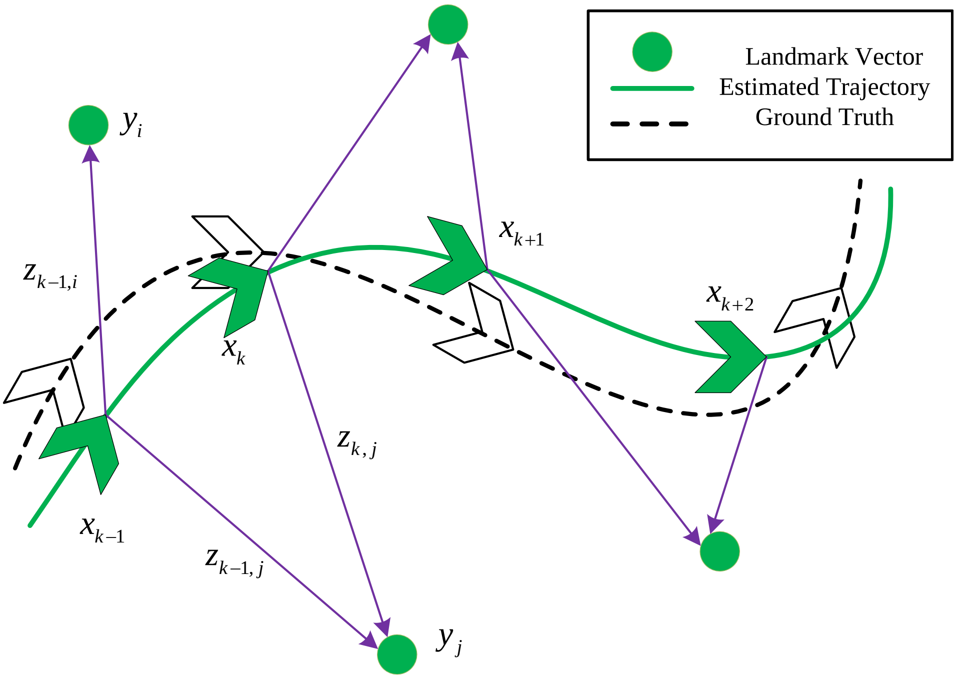

Figure 4. Schematic diagram of the SLAM process. The dashed line is the real trajectory. The solid line is the estimated trajectory. The related description of individual variables are listed under Formula (1).

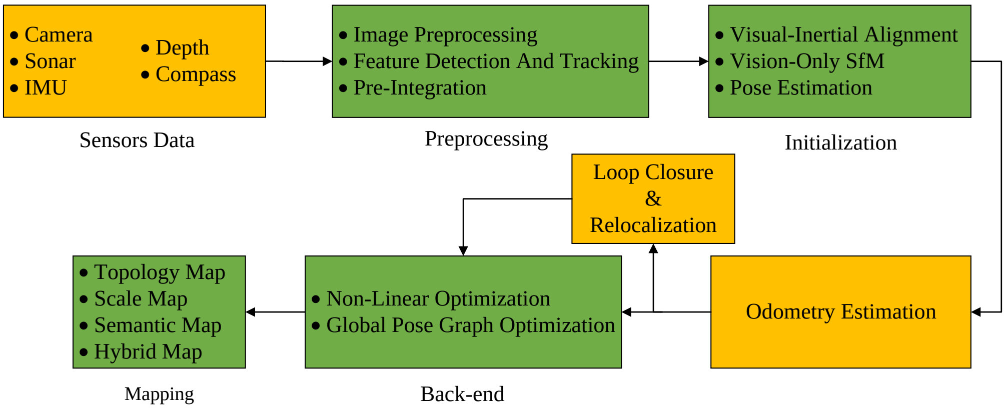

Figure 5. Process of underwater SLAM system.

However, due to the unstructured characteristics and complexity of the underwater environment, conventional SLAM methods can cause various problems when directly used in this domain. To address this problem, this section presents a schematic analysis of the various aspects of underwater SLAM systems based on the framework process of SLAM systems (Figure 5). Moreover, this section summarizes the existing approaches of researchers. Specifically, Section 3.1 describes the improved underwater SLAM from the sensor perspective for the specificity of the underwater environment; Section 3.2 analyzes the front-end odometry estimation with a vision focus; Section 3.3 presents the back-end optimization algorithm for underwater SLAM; Section 3.4 describes the principles and related methods for loop-closure detection in underwater SLAM systems. Finally, Section 3.5 summarizes the mapping approach of underwater SLAM.

2.1. Sensor Information

SLAM systems are capable of deriving the current position, estimating the trajectory, and constructing a map of the environment from the information acquired by their sensors. In comparison to traditional GPS satellite navigation and landmark navigation, a SLAM system relies solely on its sensor information for calculation. The accuracy of mapping is contingent upon the abundance of sensor information regarding the sensed underwater environment. Since GPS and inertial measurement units are not viable options for underwater navigation due to the environment's rejection of them, research has shifted to the use of underwater visual SLAM. Cameras can capture a vast amount of texture information in the underwater environment and are less expensive than the alternatives, making vision-based underwater SLAM a popular research topic [41].

However, underwater scenes are highly unstructured and uncertain. This presents a challenge to underwater robots due to their own structure and the underwater environment, which can lead to blurred imaging, shadowing, and distortion. Consequently, this results in a low signal-to-noise ratio of images, leading to a difficulty in accurately extracting underwater features, errors in matching accuracy, and even a potential tracking failure. Thus, research into underwater SLAM is always challenging, and the selection of sensors and the processing of information are of utmost importance [42].

2.1.1. Proprioceptive Sensors

Proprioceptive sensors are crucial components of underwater robots, enabling them to obtain their own state and position information without external assistance. Commonly used proprioceptive sensors include the depth sensor, Doppler velocity log (DVL), inertial measurement unit (IMU), and compass. The DVL works by transmitting acoustic pulses and receiving Doppler displacement echoes to calculate the vehicle's velocity. The compass provides a directional reference, while the IMU, composed of an accelerometer and a gyroscope, is responsible for measuring acceleration and angular acceleration. The depth sensor, meanwhile, calculates the depth of the vehicle based on the water pressure. With the velocity, orientation, acceleration, and depth information provided by these proprioceptive sensors, underwater robots can estimate their position and pose without external assistance, thereby allowing for the realization of underwater SLAM.

2.1.2. Exteroceptive Sensors

(1) Vision Sensors
SLAM based on vision sensors is an important class of SLAM algorithms, which can be classified into monocular, stereo, and RGB-D SLAM depending on the type of camera used. Additionally, algorithms such as ORB-SLAM3 can also be employed for pinhole and fisheye cameras.

- Monocular Camera
  Monocular cameras use a single lens to generate images, offering advantages such as a low cost, a simple structure, and usability. Mono SLAM was the first implementation of real-time monocular vision SLAM. In underwater scenarios, Hidalgo et al. conducted controlled experiments using ORB-SLAM under different setup conditions, as reported in their paper [43]. The results demonstrated that ORB-SLAM could be used effectively under the conditions of a sufficient illumination, low flicker, and rich scene features. On the other hand, they also indicated that monocular cameras were susceptible to light variations, object motion, and texture blurring when used in underwater scenarios.
  Ferrera et al. [44] presented a new monocular visual odometry method that was robust to turbid and dynamic underwater environments. Results showed that the optical flow method had better tracking performance than the classical descriptor-based methods. The optical flow tracking was further enhanced by adding a retracking mechanism, making it robust to short occlusions caused by environmental dynamics. The algorithm was evaluated on both simulated

and real underwater datasets and could be used in applications such as underwater archaeology. Roznere et al. [45] proposed a real-time depth estimation method for underwater monocular camera images by fusing measurements from a single-beam echosounder. The proposed method matched the echosounder measurements with the detected feature points of the monocular SLAM system and then integrated them into the monocular SLAM system to adjust the visible map points and scale. They implemented the proposed method in ORB-SLAM2 and evaluated its performance in a swimming pool and the ocean to verify the improved effect of image depth estimation, which proved that the method had a certain application value in underwater exploration and mapping.

- Stereo Camera
  In monocular camera SLAM, the scale problem cannot be determined due to the lack of depth information. In contrast, a stereo camera can acquire the distance between the camera and the object using the parallax principle. Mei et al. [46] presented a relative SLAM for the constant-time estimation of structure and motion with a stereo camera system as the only sensor. This approach employed a topological metric representation of relative position sequences based on a heuristic quadtree approach, which allowed for real-time processing while not strictly limiting the size of the maps that could be constructed. Moreover, Pi et al. [47] proposed a visual SLAM method based on a stereo camera as a sensor, leveraging the SURF algorithm for feature detection and matching and the EKF to fuse the feature coordinates and AUV pose to enable motion estimation in real time and feature map construction. Furthermore, Zhang et al. [48] suggested an underwater stereo visual–inertial localization method (FBUS-EKF) based on an open-source benchmark in the EKF framework. This method fused inertial and visual information and eliminated severe noise in order to implement a SLAM system. Experimental results indicated that the typical localization error of the FBUS-EKF method was less than 3%. Thus, stereo cameras hold great promise for the accurate proximity operation and localization of underwater robots.
- RGB-D Camera
  RGB-D cameras can obtain RGB maps and depth maps directly by physical ranging. According to their principles, they can be divided into structured light methods (e.g., Kinect v1) and time-of-flight methods (e.g., Kinect v2). However, existing RGB-D cameras typically use infrared light, which is severely attenuated in underwater environments and has high measurement limitations. As a result, it is difficult to use RGB-D cameras as vision sensors for underwater vision SLAM. Therefore, monocular and stereo cameras remain the most popularly used underwater vision sensors.

(2) Sonar Sensors
The lack of illumination in the underwater environment can significantly impact the quality of the final images. To overcome this issue, sonar can be used to detect and locate objects in the absence of light by exploiting their property of reflecting sound waves. Compared to vision, sound waves demonstrate a smaller attenuation rate and longer propagation distance than light in marine scenes and are not affected by light and geomagnetic interference. Sonar sensors can be categorized into forward-looking sonar (FLS), side-scan sonar (SSS), and acoustic lens sonar (ALS) according to the scanning mode. FLS can be further divided into single-beam sonar and multibeam sonar. The basic principles of sonar SLAM are shown in Figure 6.

- Single-beam Sonar
  Sonar is an essential external detection sensor for simultaneous localization and mapping in underwater vehicles. To this end, a variational Bayesian-based simultaneous localization and mapping method for autonomous underwater vehicle navigation (VB-AUFastSLAM) was proposed based on the Unscented-FastSLAM (UFastSLAM) and the variational Bayesian (VB) approaches [49]. The

proposed algorithm was validated in an open-source simulation environment, and its effectiveness in the marine environment was subsequently verified by constructing an underwater vehicle SLAM system based on an inertial navigation system, a Doppler velocity log (DVL), and a single-beam mechanical scanning imaging sonar (MSIS).

- Multibeam Sonar
  Multibeam sonar (MBS), as sonar for underwater sounding, has become one of the most dominant survey instruments employed in marine activities. The multibeam echosounder (MBES) typically consists of a projector and a hydrophone, which are responsible for transmitting and receiving echo soundings to measure topography. An MBE can have several hundred beams, making it the most suitable sonar sensor for deep water-terrain applications [50]. In [51], a filter-based multibeam forward-looking sonar (MFLS) algorithm for underwater SLAM was presented. Environmental features were extracted using an MFLS and the acquired sonar images were converted to a sparse point-cloud format by threshold segmentation and distance-constrained filtering to avoid a computational explosion problem. Furthermore, the method also fused DVL, IMU, and sonar data of the underwater vehicle to estimate the position of the vehicle and generate an occupancy grid map using a SLAM method based on a Rao–Blackwellized particle filter (RBPF) [52].
- Side-scan Sonar.
  Although MBS has a high resolution, it is a bathymetric tool rather than an imaging system. Side-scan sonar (SSS), with a wider range of applications, is now a commonly used tool for detecting submarine targets such as wrecks, mines, and pipelines. SSS can visually provide acoustic imaging of the seafloor morphology with a relatively high resolution. MBS and SSS have good complementarity in detecting seafloor targets and can improve the accuracy of underwater SLAM. Side et al. [53] described a side-scan sonar SLAM system for online drift compensation for underwater robots. The processing chain consisted of an automatic landmark detector, an automatic data association module, and the SLAM filter. In order to improve the robustness of the whole system while satisfying real-time performance, a batch processing method based on joint compatibility branch and bound (JCBB) was used for data association [54]. The effectiveness of the system was verified in sea trials. Furthermore, there are other sonar systems such as synthetic aperture sonar (SAS) [55] and dual-frequency identification sonar (DIDSON) [56]. A comprehensive survey of sonar SLAM can be found in [57–59].

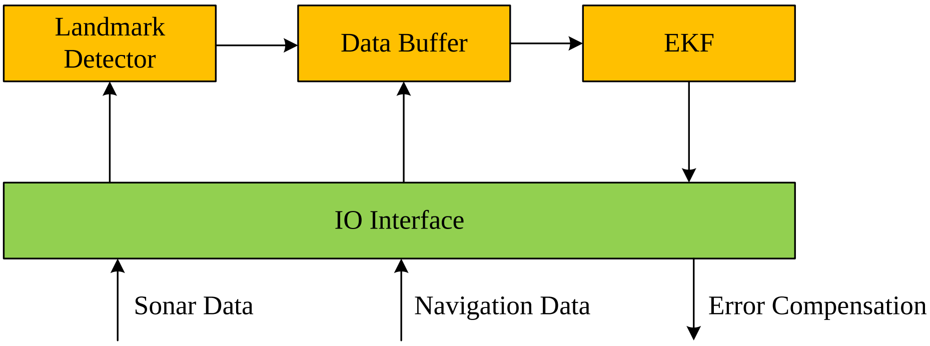

Figure 6. System structure of SLAM based on sonar.

(3) LiDAR Sensors
LiDAR sensors are capable of providing high-frequency range measurements that can operate consistently in complex lighting conditions and optically featureless scenarios [60]. Compared to camera or sonar imaging, laser-scanning imaging can provide higher-resolution 3D measurements of the seafloor in scenes lacking texture underwater. These point cloud data, generated by LiDAR, can provide easy access to

the SLAM system. Moreover, the data generated by LiDAR can be used to accurately map the seafloor and create detailed 3D models. Additionally, LiDAR sensors can be used to detect objects in the environment, allowing for the creation of more accurate navigational maps. Therefore, LiDAR has become a popular choice for seafloor mapping and navigation.

Collings et al. [61] deployed an underwater LiDAR system in parallel with an MBES to survey Kingston Reef of Rottnest Island, Western Australia. In that paper, the relative accuracy and characteristics of underwater LiDAR and multibeam sonar were compared and summarized to map the habitat. Massot et al. [62] proposed a bathymetric SLAM solution for underwater vehicles. The alignment problem of point clouds collected from a single-line-laser structured-light system was solved. In that work, the relative uncertainty in the vehicle localization was reduced by using time-constrained subgraphs. Three translational degrees of freedom and one localization degree of freedom were also used for positional estimation. However, the system could not utilize traditional SLAM image features. Palomer et al. [63] used a 3D underwater laser-scanning system to achieve underwater pipeline structure mapping on a Girona 500 AUV, which can be used for SLAM framework construction.

However, the data quality of LiDAR measurements is susceptible to extreme environments and the point cloud alignment errors caused by the smoothness of the motion. Therefore, the use of single-laser sensors in underwater SLAM environments is more restrictive. Debeunne et al. [64] provides a comprehensive survey on visual–LiDAR SLAM. Solutions using vision, LiDAR, and sensor fusion of both modalities are highlighted.

2.1.3. Multiple Sensors

Single types of sensors often have certain drawbacks when used in underwater SLAM systems. To improve the accuracy and robustness of underwater SLAM, some scholars have integrated multiple sensors into the system. The resulting sensor fusion enables underwater SLAM with a higher accuracy and robustness [39]. Common fusion methods include vision–inertial SLAM (composed of vision and IMU), laser–vision SLAM (composed of laser and vision), and multisensor SLAM (combining sonar, IMU, vision, etc.). Multisensor fusion can be further divided into data layer fusion, feature layer fusion, and decision layer fusion, according to the level of fusion. Coupling complexity can be categorized into loosely coupled, tightly coupled, and ultratightly coupled. Additionally, fusion methods can further be divided into weighted average method, Kalman filter, Bayesian estimation, D-S evidence theory, fuzzy logic, neural network, and so on (as shown in Figure 7).

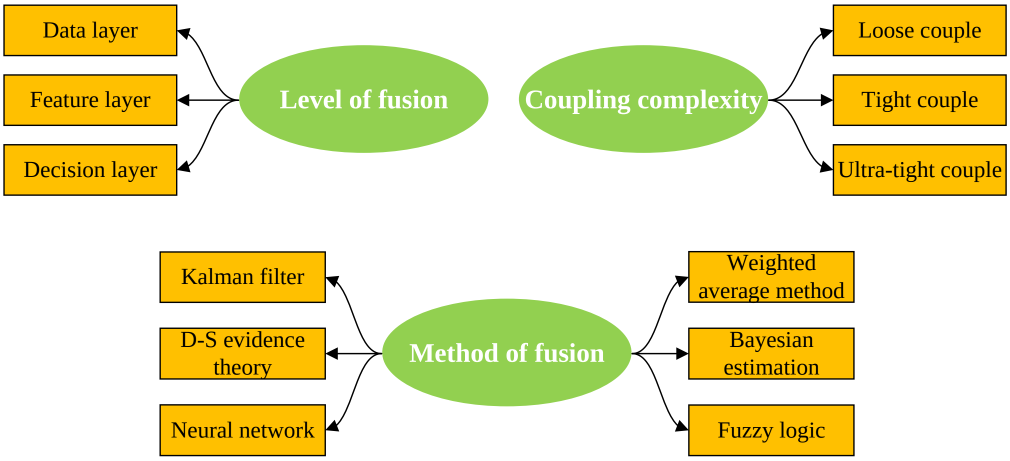

Figure 7. Multiple sensors fusion methods classification.

The two sensing modalities, visual and inertial measurement, provide complementary characteristics that can effectively improve the accuracy of visual–inertial odometry or SLAM. Generally, the main visual–inertial localization methods can be divided into two categories: a class of Kalman filter based methods, such as MSCKF [65] and ROVIO [66]; and a class of graph-based optimization methods, for instance, OKVIS [67] and VINS-Mono [68].

OKVIS was the first relatively complete visual–inertial fusion scheme and has become the basis for many multisensor SLAM algorithms. This method was built with a tightly coupled fusion to best utilize all measurements. Cost functions (combining visual and inertial terms in a fully probabilistic manner) were used for the nonlinear optimization. At the algorithmic level, that paper provided a rigorous probabilistic derivation of the IMU error terms and the respective information matrix, as well as the association of contiguous image frames. At the system level, hardware and algorithms were developed for accurate real-time SLAM. Compared to the filter-based method, a higher accuracy was achieved, although the computational effort was increased. In 2022, Leutenegger proposed OKVIS2 [69], which featured the creation of pose-graph edges through the marginalization of common observations, which could be fluidly turned back into landmarks and observations upon loop closure.

VINS-Mono is a visual–inertial system consisting of a monocular camera and a low-cost inertial measurement unit. In that paper, a robust and versatile monocular visual–inertial state estimator was proposed. A tightly coupled, nonlinear optimization method was used to fuse pre-integrated the IMU measurements and feature observations to obtain a highly accurate visual inertial odometer. The global consistency was enhanced by a four-degree-of-freedom pose map optimization. In addition, the system could save and load maps efficiently for map reuse and map combination.

SVIn2 [70] presented a tightly coupled key-frame-based simultaneous localization and mapping system. The method fused visual sensors, IMU, depth meters, and sonar to address the problem of localization drift and tracking loss, which is particularly relevant for underwater environments. Moreover, a loop-closure detection and relocalization function based on the bag-of-words (BoW) library was implemented to further improve the performance of the system. The structure of SVIn2 is shown in Figure 8.

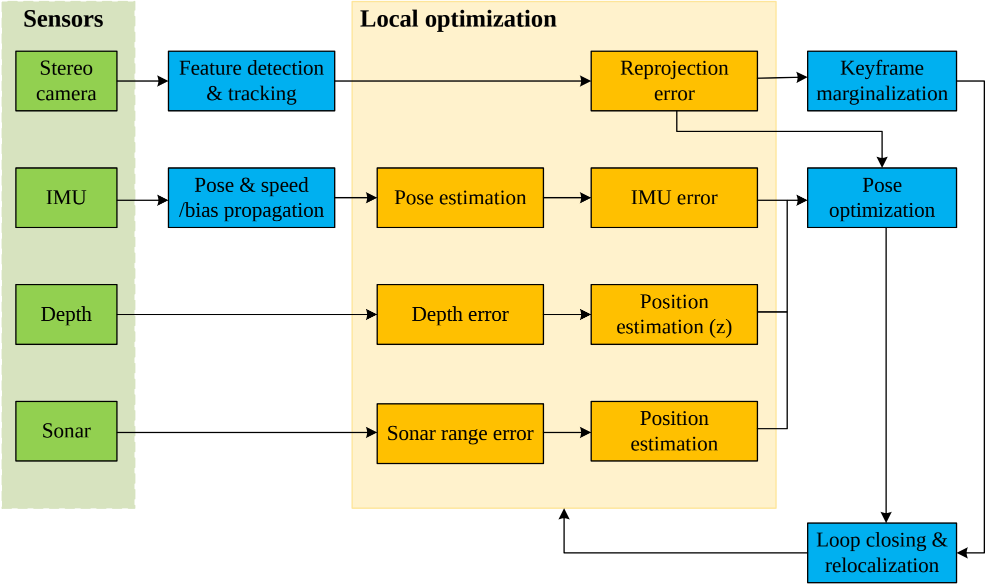

Figure 8. System structure of SVIn2: an underwater SLAM system using sonar, visual, inertial, and depth sensors.

Remark: In addition to improving the accuracy of sensor-acquired information, preprocessing the acquired information (e.g., underwater image enhancement) is another way to improve the effectiveness of SLAM. Commonly used underwater image enhancement algorithms include adaptive histogram equalization (AHE [71]), median filtering (MF [72]), and dark channel prior (DCP [73,74]) algorithms. An extensive review of underwater image enhancement can be found in [75,76].

2.2. Front-End Odometry Estimation

As the front end of the system, the odometer estimate calculates the relative poses by continuously tracking the egomotion of the camera and provides better initial values for the back end. Global trajectories are reconstructed by integrating relative poses for a given initial state. This paper aims to investigate the use of visual data for odometry estimation as they are a common way of perceiving underwater. By exploring this phenomenon, we aim to provide insights into the development of more effective odometry estimation methods for underwater applications.

Visual odometry is a method of camera motion estimation that utilizes the adjacent image information acquired from the camera to determine the relative displacement of static feature points on consecutive frames, thereby calculating relative translation and rotation increments. Furthermore, the final camera motion can be determined by connecting them to a trajectory on a global reference coordinate system. In terms of implementation methods, visual odometry can be divided into two distinct categories: geometry-based methods and deep-learning-based methods.

The geometry-based methods can be further divided into feature-based method and direct method. Among them, the feature method extracts feature points (FAST, SIFT, SURF, ORB etc.) in the image by an algorithm that computes matching feature points in adjacent frames and uses geometric relationships to obtain the rotation matrix and translation matrix of the camera. The goal of the feature point method is to minimize the reprojection error, which is typically achieved by computing the difference between the coordinate values of the pixels. The structure diagram of feature-based visual odometry is shown in Figure 9.

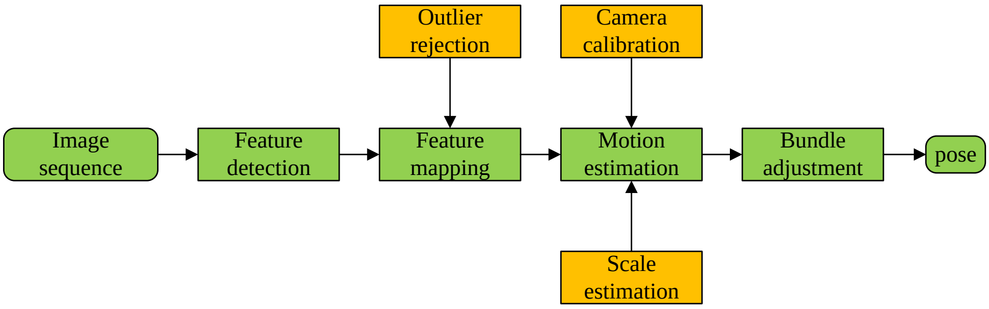

Figure 9. Structure diagram of feature-based visual odometry.

In contrast to the feature-based method, direct methods [77] such as SVO [32] and DSO [78] minimize the photometric error, and the function takes the form of subtracting the grayscale values of the pixels. Direct methods usually match two consecutive images based on the assumption of constant grayscale. The feature point method relies on a more repetitive feature extractor, and correct feature matching. The direct method is more applicable compared to the feature method when there are many repetitive textures in the environment and there is a lack of corner points. However, the direct method is relatively more computationally intensive and usually requires GPU-based computational acceleration. The comparison results between the feature-based method and the direct method are shown in the Table 1.

Unlike geometry-based methods, deep learning methods extract high-level feature representations from images without traditional feature extractors. In underwater scenarios, deep learning methods are often an excellent underwater visual odometry solution.

Early deep learning visual odometry generally consisted in a method of replacing some of the work in the system with deep learning, and such methods were generally referred to as hybrid visual odometry [79]. Subsequently, some end-to-end visual odometry based purely on neural networks were gradually proposed by scholars [80,81]. Based on the availability of ground-truth labels in the training phase, end-to-end visual odometry systems can be further classified into supervised visual odometry and unsupervised visual odometry [82,83]. The visual odometry structure diagram based on deep learning methods is shown in Figure 10.

Table 1. Comparison of feature-based method and direct method of visual odometry.

Classification	Feature-Based Method	Direct Method
Concept	Based on feature point matching; minimize the reprojection error	Based on gray invariant; minimize the luminosity error
Advantage	Strong robustness and high precision	Can be used in scenes with repetitive textures and missing corners
Disadvantage	Both quantity and quality of feature points are required	Sensitive to illumination changes; difficult to realize loop-closure detection and relocation
Characteristic	Data association and pose estimation are decoupled; builds sparse maps; loop-closure detection and relocation are required	Data association and pose estimation are coupled; builds semidense or dense maps; suitable for multisensor fusion

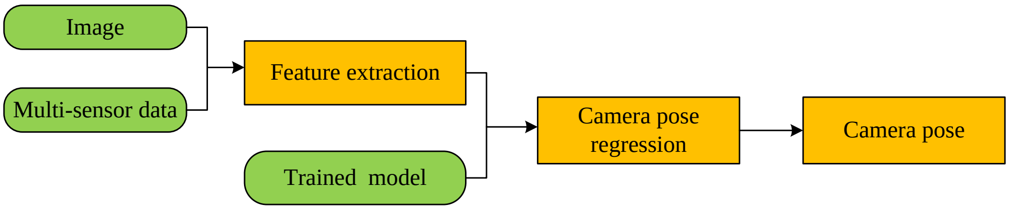

Figure 10. Visual odometry structure diagram based on deep learning methods.

Remark: for a more detailed overview of the VO section, please refer to [84–86].

2.3. Back-End Optimization

SLAM is essentially an estimation of the uncertainty in the robot itself and the surrounding space. As time accumulates, front-end odometry estimates accumulate errors in the estimation of camera poses. Based on the front-end odometry, the back end can perform the state optimization of the entire system at a larger scale and over a longer period of time. For underwater vision SLAM, the mainstream state optimization algorithms include the extended Kalman filter, based on filtering theory, and graph optimization, based on nonlinear optimization theory. To date, the research into the application of these algorithms to the underwater environment has been limited, and further study is needed to determine their effectiveness.

In the SLAM solution process, the linear optimization utilizes a filter model, which is based on Bayesian probability theory. This filter model allows for the conversion of the SLAM problem into the determination of the joint probability of the camera subject pose and the spatial features of the surrounding environment.

The goal of graph-optimization SLAM is to estimate the maximum a posteriori (MAP) of the pose of the mobile robot based on the environmental observations, namely,

$$X^* = argmax(P(X|U)) \tag{2}$$

where X represents the pose of the robot, and U represents the constraint condition. Through the Bayes theorem, we can get:

$$P(X|U) \propto \prod_{i} P(x_{i+1}|x_i, u_i) * \prod_{ij} P(x_j|x_i, u_{ij}) \tag{3}$$

The key to graph-optimization SLAM is to compute the maximum posterior optimization result when the probability distribution is maximized. That is, solving for the maximum posterior probability of the variable to be estimated is transformed into solving for the maximum likelihood estimate.

Loop-closure detection is the commonly used back-end optimization approach (refer to Section 2.4 for details), in addition to BA [87] (minimizing the reprojection error of the map by jointly optimizing the map and the poses) or pose–map optimization (representing the pose–map relationship as a computational map to be optimized) strategies to obtain a more global composition map and pose estimation.

2.4. Loop-Closure Detection

Loop-closure detection is a method for the global optimization of maps, which is used to suppress the error accumulation between the estimated and real values of the camera poses. During the camera movement, the algorithm calculates the similarity between maps to determine whether the camera has reached a visited scene. Upon detection, the information is transmitted to the back end for optimization, ensuring that the estimated trajectory is geometrically consistent and the accumulated error is eliminated. The flow chart of the loop-closure detection is shown in Figure 11.

The underwater scene is highly dynamic and lacks texture and structural features, making feature information from underwater sensors problematic when applied to loop-closure detection compared to structured ground environment information. These factors increase the difficulty of loop-closure detection, likely leading to false or missed detection events. The cumulative error resulting from missed detection events significantly affects the accuracy of positioning and framing. Moreover, the wrong loop-closure results inputted to the back end for optimization can even cause a failure of the entire SLAM system.

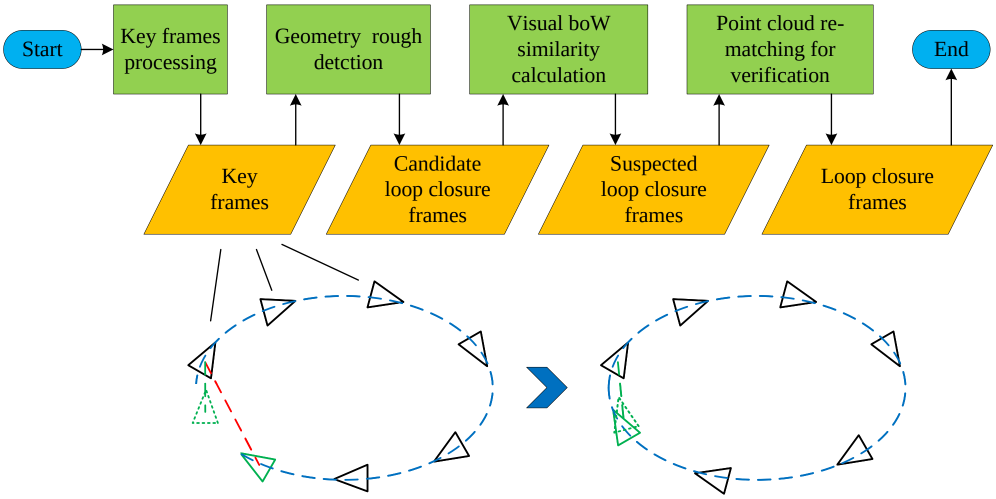
Start
Key frames processing
Geometry rough detction
Visual boW similarity calculation
Point cloud re-matching for verification
End
Key frames
Candidate loop closure frames
Suspected loop closure frames
Loop closure frames

Figure 11. Flow chart of the loop-closure detection.

Traditional loop-closure detection is based on a bag-of-words (BoW) model to store and use the visual features of the detector. Scene recognition is achieved by matching manually designed sparse features or pixel-level dense features [88]. Common methods include DBoW2 [89], FBoW [90], and iBoW-LCD [91].

Bonin et al. [92] presented an experimental evaluation of a hash-based loop-closure detection method for underwater autonomous vehicles utilizing a new global image descriptor called network-hash-based loop closure (NetHALOC). The conversion from images to hashes resulted in significantly fewer data to process, share, compare, and transfer in global navigation, localization, or mapping tasks. To demonstrate the performance of the proposed closed-loop detection method, a comprehensive test was conducted using many real underwater images, which were compared by three different high-quality global image descriptors. The results showed that the proposed method was suitable for underwater visual SLAM applications.

However, the problem was complicated by light, visual range, distortion, and the motion of camera in underwater scenes. Because deep neural networks extract high-level features, these methods are more robust to viewpoint and scene changes. Memon et al. [93] proposed a new hyperdictionary deep learning method. Unlike the traditional BoW dictionary, it made use of more advanced and abstract features. The proposed method did not require the generation of a vocabulary list, which made it memory efficient. Instead, it stored the exact features, which were few and had a small memory usage.

Loop-closure detection based on deep learning has been demonstrated to provide more robust and effective visual features, with better results for position recognition. In the context of underwater scenes, deep learning methods have been shown to exhibit a higher accuracy and stronger robustness than traditional methods. As such, loop-closure detection based on deep learning is an important future research direction for underwater SLAM, offering the potential to achieve an improved accuracy and robustness [94–97].

Remark: Robots are subject to rapid motion or other factors that may lead to image matching and trajectory tracking failure. Consequently, loop-closure detection may not be suitable for such applications [98,99]. As an alternative, relocalization recalculates the camera pose by leveraging the map or repositioning based on map stitching, which provides a more versatile and robust system.

2.5. Mapping

Mapping is an essential component of mobile robotics, as it allows the robot to build a model that accurately reflects the environment. Underwater maps, in particular, have become increasingly important for marine scientific research. These maps typically serve several functions, such as localization, navigation, obstacle avoidance, reconstruction, and interaction. The representation of the final map generated by an underwater SLAM system can vary depending on the method and the implemented functionalities. Generally, these maps can be divided into several categories.

(1) Topology Map

Topological maps possess a high degree of abstraction and are well-suited to environments with large areas and simple structures. This approach represents the environment as a graph in a topological sense, with nodes in the graph corresponding to a feature state or location in the environment. Key frames are utilized as nodes of the map, and common data associations between them are used as edges of the map. By abstracting the map into nodes and edges in line with graph theory, the maps' compatibility with human thinking is improved [100]. Choset et al. proposed a novel approach to simultaneous localization and mapping (SLAM) that utilized the topology of free space to localize the robot on a partially constructed map [101].

Topological maps can be used for path planning, due to their relatively small storage and search space, making them computationally efficient. Furthermore, they enable the utilization of numerous sophisticated and efficient search and inference algorithms [102]. However, topological maps typically lack metric information and are therefore unsuitable for navigation. The use of such maps relies on the identification and matching of topological nodes. If the environment is too similar, topological map methods may have difficulty distinguishing between two points.

(2) Scale Map

- Raster Map
  The raster map divides the 3D environment space into cubes of equal size, each representing an area of 3D space in the real environment. The value of each cube reflects the probability of an obstacle existing in the corresponding 3D space. The raster map preserves as much information as possible about the entire environment, enabling self-positioning, path planning, localization, navigation, and obstacle avoidance. Furthermore, it has great advantages for fusing multisensor information, such as weighted average methods and D-S evidence inference methods. However, as the size of the environment increases, more computation and storage space are required. When the number of rasters increases, for instance in large-scale environments or when the environment is divided in greater detail, the maintenance behavior of the map becomes more difficult. The search space in the localization process is large and, without a better simplification algorithm, real-time performance is poor [103].
- Landmark Map
  Geometric features of the environment are represented using parametric features (e.g., points, lines, and planes). Based on the feature point density, these can be further classified into sparse, semidense, and dense maps. Notably, sparse road maps can only be used for localization, whereas dense maps can be used for navigation and obstacle avoidance functions [104].
- Point Cloud Map
  The environment is described by a large number of three-dimensional spatial points, discretizing all objects in the environment into a dense point cloud [105]. Such point cloud maps are suitable for localization, navigation, obstacle avoidance, and 3D reconstruction. Meanwhile, large-scale environments necessitate a greater amount of computation and more storage space.

(3) Semantic Map
Semantic maps are composed of several distinguishable semantic elements, which can be either scene types or object types. The emphasis is placed on associating semantic concepts with objects in the map, giving them a more abstract meaning. Furthermore, these maps enable mobile robots to act more intelligently and perform more complex interaction tasks [106]. However, the types of objects in different environments often differ. On the one hand, it is not possible to assign semantic concepts to all objects when constructing semantic classes. On the other hand, objects of the same type may differ significantly, while objects of different classes may be more similar, making it difficult to create cognitive maps for complex environments. Furthermore, the complex cognitive map creation algorithm also necessitates a greater computational effort.

(4) Hybrid Map
Currently, no single map representation is capable of adequately meeting all task requirements (localization, navigation, obstacle avoidance, path planning, 3D reconstruction, interaction, etc.) and performance criteria (high accuracy, speed, low computational effort, small storage space, etc.). Consequently, it would be more advantageous to describe the underwater environment using multiple different map representations, thereby harnessing the advantages of each and ultimately achieving different objectives.

3. Research Emphasis and Difficulties

According to the above, we have summarized some selected works on underwater SLAM in Table 2. However, considering the special characteristics of the underwater environment, traditional SLAM algorithms have encountered many problems when expanding to this domain. Thus, there are still many unsolved issues concerning underwater SLAM that require attention.

Table 2. Summary of selected works on underwater SLAM.

	Sensors	Method	Optimization	Loop Closure	Scenario
LSD-SLAM [30]	Mono	Direct	Pose graph	Yes	Large-scale, consistent maps
DSO [78]	Camera	Direct and sparse	Nonlinear joint	No	-
SVO [31]	Mono	Semidirect	Minimize reprojection error	No	-
ORB-SLAM2 [28]	Mono, stereo, RGB-D	Indirect	BA	Yes	Textured environment
ORB-SLAM3 [29]	Mono, stereo, RGB-D, pinhole, fisheye, IMU	Indirect	BA	Yes	Textured environment
ROVIO [66]	IMU, camera	Direct	EKF	No	Employed in UAV
OKVIS [67]	Camera, IMU	Indirect	Marginalization of key frames	No	Hand-held indoor motion, bicycle riding
OKVIS2 [69]	Stereo, IMU	Indirect	Marginalization of common observations	Yes	-
SVIn2 [70]	Stereo, IMU, Depth, Sonar	Indirect	Tightly coupled	Yes	Underwater environments
VINS-Mono [68]	Mono, IMU	Indirect	Tightly coupled and pose graph	Yes	Employed in UAV
MSCKF [65]	Mono, stereo, IMU	Indirect	EKF	No	Real-World environment trajectory
DeepVIO [107]	Stereo, IMU	Self-supervised learning method	-	No	-
SelfVIO [108]	Mono, IMU	Self-supervised learning method	-	No	-
Dolphin SLAM [109]	Sonar, camera, DVL, IMU	Indirect	Bioinspired	Yes	Underwater environments
AEKF-SLAM [110]	Sonar (mainly)	Indirect	AEKF	Yes	Underwater environments
[63]	Laser, AHRS, DVL, pressure sensor	Indirect	EKF	No	Underwater pipe structure
[111]	Mono	Indirect	BA	Yes	Autonomous underwater ship hull inspection

3.1. Sensor Noise

Sensors usually have a limited operating depth, making applications costlier. The working depth and cost of underwater sensors vary depending on the type of sensor and the application scenario [112]. For instance, the operational depth of an underwater IMU is typically limited to a few hundred meters and comes at a low cost. Conversely, sonar systems can operate at depths reaching several thousand meters but are comparatively expensive. Cameras and laser radars generally have shallow operational depths in the tens of meters range, yet their costs remain high. In general, underwater sensors such as visualization sonar (imaging sonar), profiling sonar, DVL, IMU, camera, and depth gauge have distinct characteristics that make it difficult to obtain the same high quality of environmental sensing data as that of land-based LIDAR/cameras. Moreover, the accuracy of sensor data from cameras and other sensors is limited underwater, particularly in environments affected by low light, turbidity, and currents [113,114]. Consequently, the positional estimation and mapping tasks often result in a significant bias due to sensor noise, which varies depending on the situation, especially in large-scale environments where a new calibration is required to achieve better system estimates. Therefore, the modeling of sensor noise is a challenging yet critical task.

3.2. Feature Limitation

Landmark recognition is essential for maintaining the estimated position of a robot, reducing uncertainty in the system, and identifying the position prior to movement. The unstructured nature of underwater environments, however, presents a challenge for feature extraction and matching, as there are few obvious objects or features in most scenes [115]. Additionally, optical sensors, such as cameras, are prone to interference from light and turbidity, while also experiencing decreased imaging results due to the limited field of view in deep-water environments [116,117]. While sonar sensors may provide a larger field of view and improved feature extraction compared to optical sensors, they are still subject to interference from a variety of external factors during actual measurements, including an uneven distribution of the seawater medium, fluctuations generated by the robot's own motion underwater, and noise from marine animals and reverberation interference.

3.3. Scene Detection

Loop-closure detection refers to the ability of a mobile robot to determine when it has reached a location for which a map has been previously constructed, and then to update and correct the originally constructed map, with the main purpose of eliminating the long-term accumulated error of the SLAM system. As previously stated, the underwater environment does not have many structural features, and it undergoes large dynamic changes, making it difficult to perform closed-loop detection, and likely leading to false or missed detection events [118]. Moreover, compared to structured ground environment information, the feature information of underwater vision and acoustics presents additional challenges when used for closed-loop detection [119,120]. These false or missed detection events can lead to cumulative errors, which significantly affect the accuracy of positioning and composition.

3.4. Underwater Ground Truth

Given the GPS-denied nature of underwater scenarios, obtaining the true value of a robot's position and trajectory is a difficult task due to the errors of underwater sensors. Acoustic positioning systems such as LBL [121], SBL, and USBL require extensive base-array installation work and tend to have a greater position uncertainty than the modern resolution of multibeam or interferometric side scan sonar, making them costly. Furthermore, true underwater values are especially difficult to obtain in deep-sea and cave scenarios, resulting in a relatively small number of reliable datasets specific to underwater SLAM. (We have collected some publicly available underwater datasets for use [122–127], as detailed in Appendix A).

3.5. Computational Complexity

The computational complexity of SLAM systems is influenced by the size of the exploration environment and is closely related to the methods used for feature extraction, tracking, data association [128], and filtering. The location landmarks, map elements, and other such features generated by the robot as it moves through the underwater environment are identified and monitored, increasing the uncertainty of the associated computation [129]. Compared to surface environments, underwater environments such as oceans, lakes, and reservoirs have a huge space and complex underwater robot activities, making it difficult to achieve a balance between accuracy and speed for SLAM systems. To date, improving the accuracy in large-scale environments has remained a challenge for underwater SLAM applications.

4. Discussion

In the previous section, we discussed the system framework of underwater SLAM and the current research difficulties. In this section, we study the future development direction of underwater SLAM from the perspective of application environment and technological development, so as to open up new ideas for the research of underwater SLAM.

4.1. Underwater SLAM in Extreme Environments

In complex or confined underwater environments, such as underwater energy storage facilities, docks, flooded tunnels, and sewers, human inspection is often dangerous or impractical, necessitating remote inspection using unmanned underwater vehicles. Implementing such systems, however, becomes more difficult when the size and motion of the robot must be considered [130]. Furthermore, underwater SLAM systems are hindered by the low illumination, turbidity, and lack of features in these environments, severely limiting the capabilities of underwater inspection robots to manually controlled, low-quality visual inspections. This is a key research direction for scholars to explore [131,132].

4.2. Underwater SLAM in Dynamic Environments

Considering the pose calculation principle of SLAM systems, dynamic objects in a SLAM process will seriously affect feature matching and calculation results. On land, SLAM in dynamic environments is a hot research topic [133–135]. However, underwater dynamic phenomena are very common, such as sea creatures, water flow caused by robot motion, bubbles, etc. Maps generated in dynamic environments are more flawed than those generated in static environments. Aiming at underwater SLAM problems in dynamic environments is also more difficult. As an end-to-end feature learning method, deep learning technology provides a new idea for dynamic feature extraction and processing. Unlike direct and feature-based methods that use physical models or geometric theory, deep learning methods provide an alternative to solve problems in a data-driven way. Benefiting from the ever-increasing quantity of data and computational power, these methods can generate accurate and robust systems for tracking motion and estimating the structure of real-world scenes. They are rapidly evolving into a new field of research [136]. However, considering that an underwater dataset is not enough, and the algorithm is facing the problem of scene applicability, the work of underwater dynamic SLAM based on deep learning is slow yet still destined to be an important method to solve dynamic problems in the future (as shown in Figure 12).

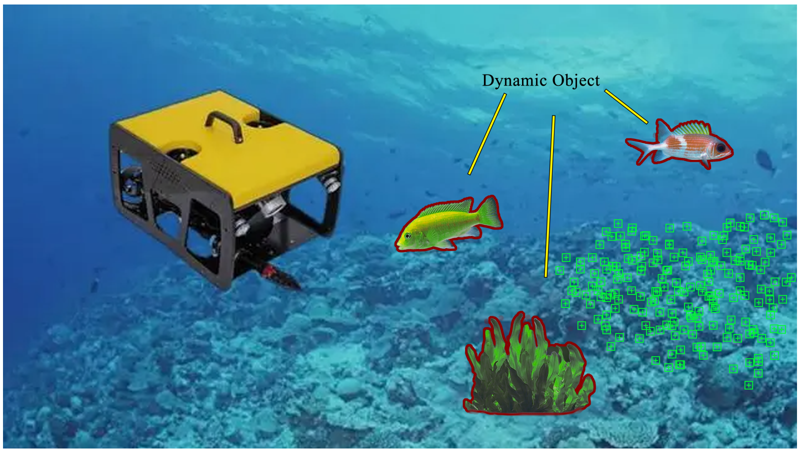

Figure 12. Effect of dynamic, weakly textured objects on underwater SLAM (green squares with crosses inside mean feature points).

4.3. Underwater Semantic SLAM

Most current SLAM systems are based on geometric features and ignore the semantic information of objects in the environment, resulting in a relatively homogeneous system function. With the continuous development of deep learning in the field of computer imaging, researchers have started to incorporate target recognition and semantic segmentation into SLAM to achieve more complex functions. Semantic SLAM refers to a SLAM system that obtains semantic information while acquiring the geometric structure information in the environment during the mapping process, and at the same time recognizing the independent individuals in the environment [137]. This semantic information includes information about the position, pose, and functional attributes of the individuals. With this semantic information, robots can more effectively cope with complex scenarios and refine and improve service tasks. Semantic extraction is currently a deep learning approach that uses deep neural networks to process images, such as segmentation and recognition, and add labels, and the training results determine the final effect of semantic SLAM. In specific underwater environment tasks, such as underwater salvage, underwater defect detection, marine biological analysis, and other underwater tasks, the semantic recognition of captured objects during SLAM is required [138]. However, the accurate generation of image-based obstacle maps in cluttered underwater environments is extremely demanding for the robustness of underwater robotic SLAM systems. Moreover, such recognition can be affected by lighting conditions and moving objects (e.g., schools of fish), which can lead to misjudgments. The presence of a large number of dynamic objects is also detrimental to the final composition. Therefore, there are still many challenges for the application of semantic SLAM in underwater environments [139]. The schematic diagram of underwater semantic segmentation is shown in Figure 13.

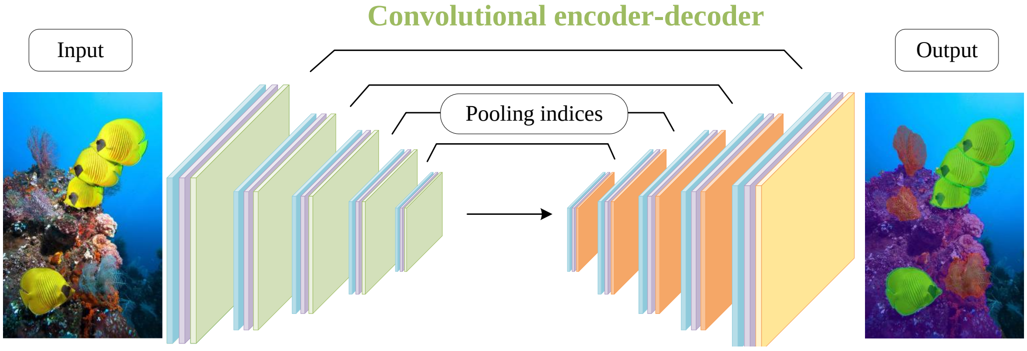

Figure 13. Schematic diagram of underwater semantic segmentation.

4.4. Multirobot Underwater SLAM

Different from SLAM in indoor scenes, underwater scenes have a large area and a single structure, thus relying on a single robot will result in low efficiency [140]. To address this, multirobot SLAM requires technical elements such as a system to build multiple SLAM systems and to collect information from each robot. Moreover, map fusion, an important process in multirobot SLAM, combines multiple local maps estimated by the robot team into a global map [141].

The fusion of multiple local maps is usually done using a loopback detection method that identifies whether multiple robots have visited the same scene. Additionally, a robot federation approach can be employed, allowing another member of a robot team to be observed in the images of that member. However, when multiple robots operate in environments with multiple similar scenes, they are likely to identify different points with similar scenes as the same location, leading to erroneous results from the loopback detection. Furthermore, the viewpoint positions of the cameras mounted on each robot are not guaranteed to be identical, making it difficult to observe the same place. Moreover, the descriptors of each scene differ, thus making it impossible to detect loopback even when the same place is observed. Establishing an effective fusion process of multiple maps is the focus of multirobot underwater SLAM research [142,143].

5. Conclusions

This paper systematically discussed the key technologies of SLAM in underwater scenes, including theoretical background, system framework, existing methods, problems, applications and development trends.

In this paper, the methods of sensor information, front-end odometry estimation, back-end optimization, loop-closure detection, and mapping in the underwater SLAM system framework were summarized. The sensor information part involved proprioceptive sensors, exteroceptive sensors (visual, acoustic, and LiDAR), and multiple sensors; the front-end odometry estimation part involved geometry-based methods (feature methods and direct methods) and deep-learning-based methods. The back-end optimization part involved the extended Kalman Filter based on filtering theory and graph optimization based on nonlinear optimization theory. The loop-closure detection part involved bag-of-words methods and deep-learning-based methods. The mapping part analyzed different types and characteristics of maps. Furthermore, this paper also outlined the specific research difficulties of underwater SLAM, including sensor noise, feature information, scene detection, ground truth, and computational complexity. From the perspectives of extreme environment, dynamic environment, underwater semantic SLAM, and underwater multirobot SLAM, the future research directions and focuses of underwater SLAM were presented.

Our research results link the latest research results in the fields of underwater robotics, computer vision, and machine learning, and provide guidance to future researchers for understanding feasible approaches to apply emerging technologies to solve underwater robot localization and composition problems.

Author Contributions: Conceptualization, X.W. and X.F.; methodology, X.W.; validation, X.F., P.S. and J.N.; formal analysis, X.W. and Z.Z.; resources, X.F. and P.S.; writing—original draft preparation, X.W.; writing—review and editing, X.W., X.F., P.S., J.N. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, Grant/Award Number: 2022YFB4703405; the Fundamental Research Fund for the Central Universities, Grant/Award Number: B220202020; the Fundamental Research Funds for the Central Universities, Grant/Award Number: B220203032.

Data Availability Statement: Datasets relevant to our paper are available online.

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

Abbreviations

The explanations of some abbreviations in the text are shown as follows:

SLAM	Simultaneous localization and mapping	SSS	Side-scan sonar
GNSS	Global navigation satellite system	ALS	Acoustic lens sonar
GPS	Global Positioning System	VB	Variational Bayesian
UUV	Unmanned underwater vehicles	UFast	Unscented-fast
ROV	Remotely operated vehicles	MSIS	Mechanical scanning imaging sonar
AUV	Autonomous underwater vehicles	MBS	Multibeam sonar
SBL	Short baseline	MBES	Multibeam echosounder
USBL	Ultrashort baseline	MFLS	Multibeam forward looking sonar
EKF	Extended Kalman filter	RBPF	Rao–Blackwellized particle filter
PF	Particle filter	JCBB	Joint compatibility branch and bound
MLE	Maximum likelihood estimation	SAS	Synthetic aperture sonar
ORB	Oriented FAST and Rotation BRIEF	DIDSON	Dual-frequency identification sonar
LSD	Large-scale direct	MSCKF	Multistate constraint Kalman filter
SVO	Semidirect visual odometry	FBUS	Fiducial-based, underwater stereo
PTAM	Parallel tracking and mapping	BoW	Bag-of-words
DT	Deferred triangulation	AHE	Adaptive histogram equalization
IMU	Inertial measurement units	MF	Median filtering
DVL	Doppler velocity log	DCP	Dark channel prior
6-DOF	Six-degree-of-freedom	SIFT	Scale-invariant feature transform
LBL	Lone baseline	SURF	Speed up robust feature
ROVIO	Robust visual inertial odometry	SVO	Semidirect visual odometry
VIO	Visual inertial odometry	DSO	Direct sparse odometry
SVIn	Sonar, visual, inertial	MAP	Maximum a posteriori
VINS	Visual–inertial state	BA	Bundle adjustment
CNN	Convolutional neural networks	NetHALOC	Network hash-based loop closure
SfM	Structure from motion	AHRS	Attitude and heading reference system

Appendix A

One of the difficulties of underwater SLAM is the acquisition of underwater truth values, which indirectly leads to a scarcity of datasets dedicated to underwater localization and composition. In this paper, we present several publicly available underwater SLAM datasets that can be used by researchers for experimental comparisons. Refer to Table A1 for details.

The Aqualoc dataset is dedicated to the development of simultaneous positioning and map construction methods for underwater vehicles navigating near the seafloor. Data sequences were recorded in three different environments: a harbor at a few meters depth, and two sites at depths of 270 and 380 m. The data acquisition was performed by ROVs equipped with a monocular monochrome camera, a low-cost inertial measurement unit, a pressure sensor, and a computational unit. The collected data consisted of 17 sequences, provided as ROS packages and used as raw data. The sequence images of the dataset are shown in Figures A1 and A2.

Table A1. Description of underwater SLAM datasets in recent years.

Simulated datasets produced using UWSim (2016) [122]	This paper provides an open collection of seven different simulated datasets produced using an underwater simulation. Those datasets present three trajectories and two simulated seafloor visual data based on real coral reef mosaics.
Underwater caves sonar and vision dataset (2017) [123]	The dataset was collected with an autonomous underwater vehicle test bed in the unstructured environment of an underwater cave. The vehicle was equipped with two mechanically scanned imaging sonar sensors to simultaneously map the cave's horizontal and vertical surfaces, a Doppler velocity log, two inertial measurement units, a depth sensor, and a vertically mounted camera imaging the sea floor for ground-truth validation at specific points.
Datasets collected by an underwater sensor suite (2018) [124]	The proposed sensor suite was used to collect sonar, visual, inertial, and depth data in a variety of environments. More specifically, shipwreck and coral reef data were collected during field trials in Barbados. More data were collected at Fantasy Lake, NC, and at different locales near High Springs, FL.
Aqualoc (2019) [125]	The data sequences composing this dataset were recorded at three different depths: a few meters, 270 m, and 380 m. Seventeen sequences were made available in the form of ROS bags and as raw data. For each sequence, a trajectory was also computed offline using a structure-from-motion library in order to allow the comparison with real-time localization methods.
VAROS synthetic underwater dataset (2021) [126]	Pose sequences were created by first defining waypoints for the simulated underwater vehicle. The scenes were rendered using the ray-tracing method, which generates realistic images by integrating direct light and indirect volumetric scattering. The VAROS dataset version 1 provides images, inertial measurement unit (IMU), and depth gauge data, as well as ground-truth poses, depth images, and surface normal images.
A bathymetric mapping and SLAM dataset with high-precision ground truth for marine robotics (2022) [127]	This paper presents a dataset with four separate surveys designed to test bathymetric SLAM algorithms using two modern sonar sensors, typical underwater vehicle navigation sensors, and a high-precision (2 cm horizontal, 10 cm vertical) real-time kinematic (RTK) GPS ground truth.

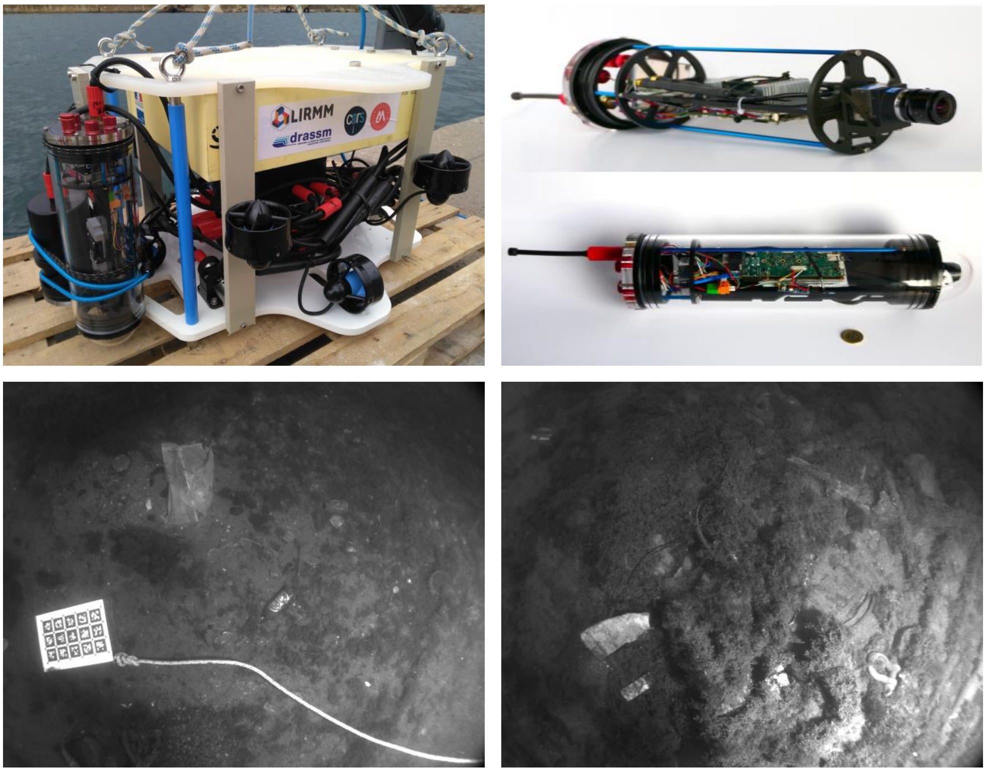

Figure A1. Port sequence of the Aqualoc dataset.

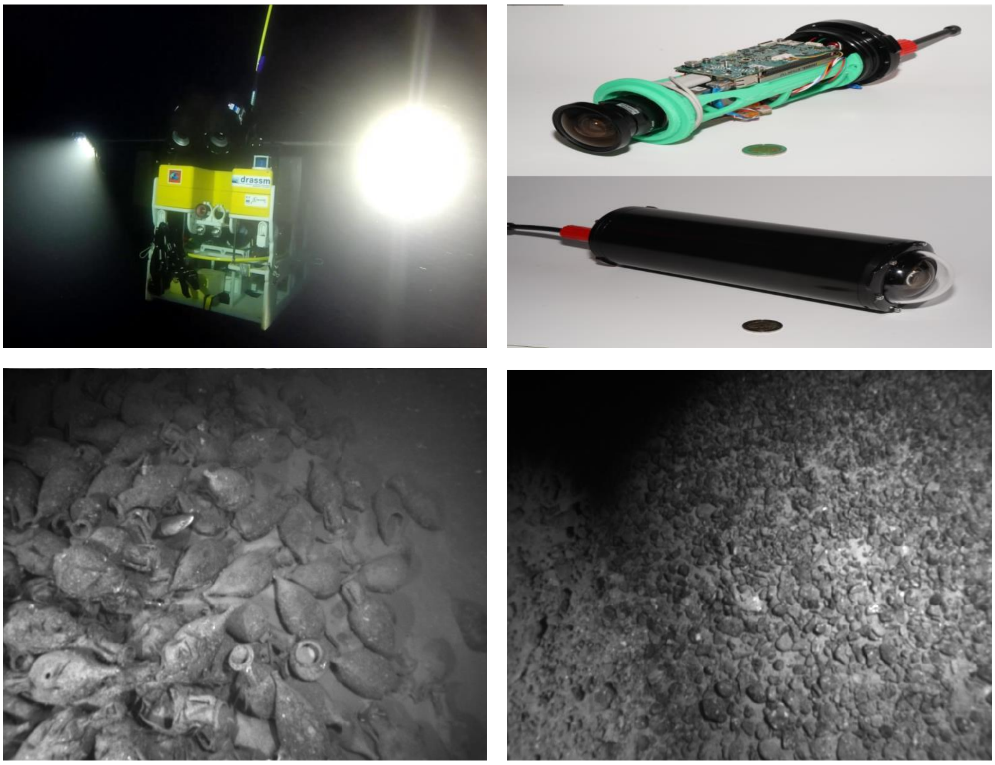

Figure A2. Archaeological sequence of the Aqualoc dataset.

References

1. Zhu, N.; Marais, J.; Bétaille, D.; Berbineau, M. GNSS position integrity in urban environments: A review of literature. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 2762–2778. [CrossRef]
2. Durrant-Whyte, H.; Bailey, T. Simultaneous localization and mapping: Part I. *IEEE Robot. Autom. Mag.* **2006**, *13*, 99–110. [CrossRef]
3. Bailey, T.; Durrant-Whyte, H. Simultaneous localization and mapping (SLAM): Part II. *IEEE Robot. Autom. Mag.* **2006**, *13*, 108–117. [CrossRef]
4. Van Nam, D.; Gon-Woo, K. Solid-state LiDAR based-SLAM: A concise review and application. In Proceedings of the 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Republic of Korea, 17–20 January 2021; pp. 302–305.
5. Cho, H.; Jeong, S.K.; Ji, D.H.; Tran, N.H.; Vu, M.T.; Choi, H.S. Study on control system of integrated unmanned surface vehicle and underwater vehicle. *Sensors* **2020**, *20*, 2633. [CrossRef]
6. Petillot, Y.R.; Antonelli, G.; Casalino, G.; Ferreira, F. Underwater robots: From remotely operated vehicles to intervention-autonomous underwater vehicles. *IEEE Robot. Autom. Mag.* **2019**, *26*, 94–101. [CrossRef]
7. Blidberg, D.R. The development of autonomous underwater vehicles (AUV); a brief summary. *Proc. IEEE Icra* **2001**, *4*, 1–12.
8. Zhao, W.; He, T.; Sani, A.Y.M.; Yao, T. Review of slam techniques for autonomous underwater vehicles. In Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence, New York, NY, USA, 20 September 2019; pp. 384–389.
9. Paull, L.; Saeedi, S.; Seto, M.; Li, H. AUV navigation and localization: A review. *IEEE J. Ocean. Eng.* **2013**, *39*, 131–149. [CrossRef]
10. Burguera, A.; Bonin-Font, F.; Font, E.G.; Torres, A.M. Combining Deep Learning and Robust Estimation for Outlier-Resilient Underwater Visual Graph SLAM. *J. Mar. Sci. Eng.* **2022**, *10*, 511. [CrossRef]
11. Burguera, A.; Bonin-Font, F. *Localization, Mapping and SLAM in Marine and Underwater Environments*; MDPI: Basel, Switzerland, 2022.
12. Smith, R.; Self, M.; Cheeseman, P. Estimating uncertain spatial relationships in robotics. In *Autonomous Robot Vehicles*; Springer: Berlin/Heidelberg, Germany, 1990; pp. 167–193. [CrossRef]
13. Bailey, T.; Nieto, J.; Guivant, J.; Stevens, M.; Nebot, E. Consistency of the EKF-SLAM algorithm. In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006; pp. 3562–3568.
14. Van Der Merwe, R.; Doucet, A.; De Freitas, N.; Wan, E. The unscented particle filter. In Proceedings of the Advances in Neural Information Processing Systems, Cambridge, MA, USA, 1 January 2000; Volume 13.
15. Dutilleul, P. The MLE algorithm for the matrix normal distribution. *J. Stat. Comput. Simul.* **1999**, *64*, 105–123. [CrossRef]
16. Cadena, C.; Carlone, L.; Carrillo, H.; Latif, Y.; Scaramuzza, D.; Neira, J.; Reid, I.; Leonard, J.J. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Trans. Robot.* **2016**, *32*, 1309–1332. [CrossRef]
17. Chen, W.; Shang, G.; Ji, A.; Zhou, C.; Wang, X.; Xu, C.; Li, Z.; Hu, K. An overview on visual slam: From tradition to semantic. *Remote Sens.* **2022**, *14*, 3010. [CrossRef]
18. Lai, T. A Review on Visual-SLAM: Advancements from Geometric Modelling to Learning-Based Semantic Scene Understanding Using Multi-Modal Sensor Fusion. *Sensors* **2022**, *22*, 7265. [CrossRef]
19. Teed, Z.; Deng, J. Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 16558–16569.
20. Macario Barros, A.; Michel, M.; Moline, Y.; Corre, G.; Carrel, F. A comprehensive survey of visual slam algorithms. *Robotics* **2022**, *11*, 24. [CrossRef]
21. Davison, A.J.; Reid, I.D.; Molton, N.D.; Stasse, O. MonoSLAM: Real-time single camera SLAM. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1052–1067. [CrossRef]
22. Tateno, K.; Tombari, F.; Laina, I.; Navab, N. Cnn-slam: Real-time dense monocular slam with learned depth prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6243–6252.
23. Gomez-Ojeda, R.; Moreno, F.A.; Zuniga-Noël, D.; Scaramuzza, D.; Gonzalez-Jimenez, J. PL-SLAM: A stereo SLAM system through the combination of points and line segments. *IEEE Trans. Robot.* **2019**, *35*, 734–746. [CrossRef]
24. Engel, J.; Stückler, J.; Cremers, D. Large-scale direct SLAM with stereo cameras. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 1935–1942.
25. Kerl, C.; Sturm, J.; Cremers, D. Dense visual SLAM for RGB-D cameras. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 2100–2106.
26. Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; Cremers, D. A benchmark for the evaluation of RGB-D SLAM systems. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 573–580.
27. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [CrossRef]
28. Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [CrossRef]
29. Campos, C.; Elvira, R.; Rodríguez, J.J.G.; Montiel, J.M.; Tardós, J.D. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. [CrossRef]

30. Engel, J.; Schöps, T.; Cremers, D. LSD-SLAM: Large-scale direct monocular SLAM. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part II 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 834–849.
31. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 15–22.
32. Forster, C.; Zhang, Z.; Gassner, M.; Werlberger, M.; Scaramuzza, D. SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Trans. Robot.* **2016**, *33*, 249–265. [CrossRef]
33. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
34. Klein, G.; Murray, D. Parallel tracking and mapping for small AR workspaces. In Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, Japan, 13–16 November 2007; pp. 225–234.
35. Herrera, D.C.; Kim, K.; Kannala, J.; Pulli, K.; Heikkilä, J. Dt-slam: Deferred triangulation for robust slam. In Proceedings of the 2014 2nd International Conference on 3D Vision, Tokyo, Japan, 8–11 December 2014; Volume 1, pp. 609–616.
36. Williams, S.B.; Newman, P.; Dissanayake, G.; Durrant-Whyte, H. Autonomous underwater simultaneous localisation and map building. In Proceedings of the Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065), San Francisco, CA, USA, 24–28 April 2000; Volume 2, pp. 1793–1798.
37. Hashemi, M.; Karmozdi, A.; Naderi, A.; Salarieh, H. Development of an integrated navigation algorithm based on IMU, depth, DVL sensors and earth magnetic field map. *Modares Mech. Eng.* **2017**, *16*, 235–243.
38. Menaka, D.; Gauni, S.; Manimegalai, C.T.; Kalimuthu, K. Challenges and vision of wireless optical and acoustic communication in underwater environment. *Int. J. Commun. Syst.* **2022**, *35*, e5227. [CrossRef]
39. Vargas, E.; Scona, R.; Willners, J.S.; Luczynski, T.; Cao, Y.; Wang, S.; Petillot, Y.R. Robust underwater visual SLAM fusing acoustic sensing. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 2140–2146.
40. Fuentes-Pacheco, J.; Ruiz-Ascencio, J.; Rendón-Mancha, J.M. Visual simultaneous localization and mapping: A survey. *Artif. Intell. Rev.* **2015**, *43*, 55–81. [CrossRef]
41. Hodne, L.M.; Leikvoll, E.; Yip, M.; Teigen, A.L.; Stahl, A.; Mester, R. Detecting and suppressing marine snow for underwater visual slam. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 5101–5109.
42. Zhang, S.; Zhao, S.; An, D.; Liu, J.; Wang, H.; Feng, Y.; Li, D.; Zhao, R. Visual SLAM for underwater vehicles: A survey. *Comput. Sci. Rev.* **2022**, *46*, 100510. [CrossRef]
43. Hidalgo, F.; Kahlefendt, C.; Bräunl, T. Monocular ORB-SLAM application in underwater scenarios. In Proceedings of the 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO), Kobe, Japan, 28–31 May 2018; pp. 1–4.
44. Ferrera, M.; Moras, J.; Trouvé-Peloux, P.; Creuze, V. Real-time monocular visual odometry for turbid and dynamic underwater environments. *Sensors* **2019**, *19*, 687. [CrossRef]
45. Roznere, M.; Li, A.Q. Underwater monocular image depth estimation using single-beam echosounder. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 1785–1790.
46. Mei, C.; Sibley, G.; Cummins, M.; Newman, P.; Reid, I. RSLAM: A system for large-scale mapping in constant-time using stereo. *Int. J. Comput. Vis.* **2011**, *94*, 198–214. [CrossRef]
47. Pi, S.; He, B.; Zhang, S.; Nian, R.; Shen, Y.; Yan, T. Stereo visual SLAM system in underwater environment. In Proceedings of the OCEANS 2014-TAIPEI, Taipei, Taiwan, 7–10 April 2014; pp. 1–5.
48. Zhang, P.; Wu, Z.; Wang, J.; Kong, S.; Tan, M.; Yu, J. An Open-Source, Fiducial-Based, Underwater Stereo Visual-Inertial Localization Method with Refraction Correction. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 4331–4336.
49. Mu, P.; Zhang, X.; Qin, P.; He, B. A Variational Bayesian-Based Simultaneous Localization and Mapping Method for Autonomous Underwater Vehicle Navigation. *J. Mar. Sci. Eng.* **2022**, *10*, 1563. [CrossRef]
50. Melo, J.; Matos, A. Survey on advances on terrain based navigation for autonomous underwater vehicles. *Ocean. Eng.* **2017**, *139*, 250–264. [CrossRef]
51. Cheng, C.; Wang, C.; Yang, D.; Liu, W.; Zhang, F. Underwater localization and mapping based on multi-beam forward looking sonar. *Front. Neurorobot.* **2022**, *15*, 189. [CrossRef]
52. Grisetti, G.; Stachniss, C.; Burgard, W. Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE Trans. Robot.* **2007**, *23*, 34–46. [CrossRef]
53. Siantidis, K. Side scan sonar based onboard SLAM system for autonomous underwater vehicles. In Proceedings of the 2016 IEEE/OES Autonomous Underwater Vehicles (AUV), Tokyo, Japan, 6–9 November 2016; pp. 195–200.
54. Neira, J.; Tardós, J.D. Data association in stochastic mapping using the joint compatibility test. *IEEE Trans. Robot. Autom.* **2001**, *17*, 890–897. [CrossRef]
55. Gerg, I.D.; Monga, V. Deep Multi-Look Sequence Processing for Synthetic Aperture Sonar Image Segmentation. *IEEE Trans. Geosci. Remote. Sens.* **2023**, *61*, 4200915. [CrossRef]

56. Belcher, E.; Hanot, W.; Burch, J. Dual-frequency identification sonar (DIDSON). In Proceedings of the 2002 Interntional Symposium on Underwater Technology (Cat. No. 02EX556), Tokyo, Japan, 19 April 2002; pp. 187–192.
57. Jiang, M.; Song, S.; Li, Y.; Jin, W.; Liu, J.; Feng, X. A survey of underwater acoustic SLAM system. In Proceedings of the Intelligent Robotics and Applications: 12th International Conference, ICIRA 2019, Shenyang, China, 8–11 August 2019; Proceedings, Part II 12; Springer: Berlin/Heidelberg, Germany, 2019; pp. 159–170.
58. Fallon, M.F.; Folkesson, J.; McClelland, H.; Leonard, J.J. Relocating underwater features autonomously using sonar-based SLAM. *IEEE J. Ocean. Eng.* **2013**, *38*, 500–513. [CrossRef]
59. Evers, C.; Naylor, P.A. Acoustic slam. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 1484–1498. [CrossRef]
60. Maksymova, I.; Steger, C.; Druml, N. Review of LiDAR sensor data acquisition and compression for automotive applications. *Proceedings* **2018**, *2*, 852.
61. Collings, S.; Martin, T.J.; Hernandez, E.; Edwards, S.; Filisetti, A.; Catt, G.; Marouchos, A.; Boyd, M.; Embry, C. Findings from a combined subsea LiDAR and multibeam survey at Kingston reef, Western Australia. *Remote Sens.* **2020**, *12*, 2443. [CrossRef]
62. Massot-Campos, M.; Oliver, G.; Bodenmann, A.; Thornton, B. Submap bathymetric SLAM using structured light in underwater environments. In Proceedings of the 2016 IEEE/OES Autonomous Underwater Vehicles (AUV), Tokyo, Japan, 6–9 November 2016; pp. 181–188.
63. Palomer, A.; Ridao, P.; Ribas, D. Inspection of an underwater structure using point-cloud SLAM with an AUV and a laser scanner. *J. Field Robot.* **2019**, *36*, 1333–1344. [CrossRef]
64. Debeunne, C.; Vivet, D. A review of visual-LiDAR fusion based simultaneous localization and mapping. *Sensors* **2020**, *20*, 2068. [CrossRef]
65. Mourikis, A.I.; Roumeliotis, S.I. A multi-state constraint Kalman filter for vision-aided inertial navigation. In Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Rome, Italy, 10–14 April 2007; pp. 3565–3572.
66. Bloesch, M.; Omari, S.; Hutter, M.; Siegwart, R. Robust visual inertial odometry using a direct EKF-based approach. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 298–304.
67. Leutenegger, S.; Furgale, P.; Rabaud, V.; Chli, M.; Konolige, K.; Siegwart, R. Keyframe-based visual-inertial slam using nonlinear optimization. In Proceedings of the Robotis Science and Systems (RSS), Berlin, Germany, 24–28 June 2013.
68. Qin, T.; Li, P.; Shen, S. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. [CrossRef]
69. Leutenegger, S. Okvis2: Realtime scalable visual-inertial slam with loop closure. *arXiv* **2022**, arXiv:2202.09199.
70. Rahman, S.; Li, A.Q.; Rekleitis, I. Svin2: An underwater slam system using sonar, visual, inertial, and depth sensor. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 1861–1868.
71. Hitam, M.S.; Awalludin, E.A.; Yussof, W.N.J.H.W.; Bachok, Z. Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In Proceedings of the 2013 International Conference on Computer Applications Technology (ICCAT), Sousse, Tunisia, 20–22 January 2013; pp. 1–5.
72. Lu, H.; Serikawa, S. Underwater scene enhancement using weighted guided median filter. In Proceedings of the 2014 IEEE International Conference on Multimedia and Expo (ICME), Chengdu, China, 14–18 July 2014; pp. 1–6.
73. Yang, H.Y.; Chen, P.Y.; Huang, C.C.; Zhuang, Y.Z.; Shiau, Y.H. Low complexity underwater image enhancement based on dark channel prior. In Proceedings of the 2011 Second International Conference on Innovations in Bio-Inspired Computing and Applications, Shenzhen, China, 16–18 December 2011; pp. 17–20.
74. Hou, G.; Li, J.; Wang, G.; Yang, H.; Huang, B.; Pan, Z. A novel dark channel prior guided variational framework for underwater image restoration. *J. Vis. Commun. Image Represent.* **2020**, *66*, 102732. [CrossRef]
75. Sahu, P.; Gupta, N.; Sharma, N. A survey on underwater image enhancement techniques. *Int. J. Comput. Appl.* **2014**, *87*, 19–23. [CrossRef]
76. Zhou, J.; Yang, T.; Zhang, W. Underwater vision enhancement technologies: A comprehensive review, challenges, and recent trends. *Appl. Intell.* **2023**, *53*, 3594–3621. [CrossRef]
77. Newcombe, R.A.; Lovegrove, S.J.; Davison, A.J. DTAM: Dense tracking and mapping in real-time. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2320–2327.
78. Engel, J.; Koltun, V.; Cremers, D. Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 611–625. [CrossRef]
79. Bruno, H.M.S.; Colombini, E.L. LIFT-SLAM: A deep-learning feature-based monocular visual SLAM method. *Neurocomputing* **2021**, *455*, 97–110. [CrossRef]
80. Zhou, Z.; Fan, X.; Shi, P.; Xin, Y. R-MSFM: Recurrent multi-scale feature modulation for monocular depth estimating. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 12777–12786.
81. Fan, X.; Zhou, Z.; Shi, P.; Xin, Y.; Zhou, X. RAFM: Recurrent atrous feature modulation for accurate monocular depth estimating. *IEEE Signal Process. Lett.* **2022**, *29*, 1609–1613. [CrossRef]
82. Wang, R.; Pizer, S.M.; Frahm, J.M. Recurrent neural network for (un-) supervised learning of monocular video visual odometry and depth. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15 April 2019; pp. 5555–5564.

83. Wagstaff, B.; Peretroukhin, V.; Kelly, J. Self-supervised deep pose corrections for robust visual odometry. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 2331–2337.
84. Teixeira, B.; Silva, H.; Matos, A.; Silva, E. Deep learning for underwater visual odometry estimation. *IEEE Access* **2020**, *8*, 44687–44701. [CrossRef]
85. Wang, K.; Ma, S.; Chen, J.; Ren, F.; Lu, J. Approaches, challenges, and applications for deep visual odometry: Toward complicated and emerging areas. *IEEE Trans. Cogn. Dev. Syst.* **2020**, *14*, 35–49. [CrossRef]
86. Wirth, S.; Carrasco, P.L.N.; Codina, G.O. Visual odometry for autonomous underwater vehicles. In Proceedings of the 2013 MTS/IEEE OCEANS-Bergen, Bergen, Norway, 10–14 June 2013; pp. 1–6.
87. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In Proceedings of the Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms, Corfu, Greece, 21–22 September 1999; Springer: Berlin/Heidelberg, Germany, 2000; pp. 298–372.
88. Qader, W.A.; Ameen, M.M.; Ahmed, B.I. An overview of bag of words; importance, implementation, applications, and challenges. In Proceedings of the 2019 International Engineering Conference (IEC), Erbil, Iraq, 23–25 June 2019; pp. 200–204.
89. Gálvez-López, D.; Tardos, J.D. Bags of binary words for fast place recognition in image sequences. *IEEE Trans. Robot.* **2012**, *28*, 1188–1197. [CrossRef]
90. Uijlings, J.R.; Smeulders, A.W.; Scha, R.J. Real-time bag of words, approximately. In Proceedings of the ACM International Conference on Image and Video Retrieval, New York, NY, USA, 8 July 2009; pp. 1–8.
91. Garcia-Fidalgo, E.; Ortiz, A. ibow-lcd: An appearance-based loop-closure detection approach using incremental bags of binary words. *IEEE Robot. Autom. Lett.* **2018**, *3*, 3051–3057. [CrossRef]
92. Bonin-Font, F.; Burguera Burguera, A. NetHALOC: A learned global image descriptor for loop closing in underwater visual SLAM. *Expert Syst.* **2021**, *38*, e12635. [CrossRef]
93. Memon, A.R.; Wang, H.; Hussain, A. Loop closure detection using supervised and unsupervised deep neural networks for monocular SLAM systems. *Robot. Auton. Syst.* **2020**, *126*, 103470. [CrossRef]
94. Hong, S.; Kim, J.; Pyo, J.; Yu, S.C. A robust loop-closure method for visual SLAM in unstructured seafloor environments. *Auton. Robot.* **2016**, *40*, 1095–1109. [CrossRef]
95. Cattaneo, D.; Vaghi, M.; Valada, A. Lcdnet: Deep loop closure detection and point cloud registration for lidar slam. *IEEE Trans. Robot.* **2022**, *38*, 2074–2093. [CrossRef]
96. Gao, X.; Zhang, T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system. *Auton. Robot.* **2017**, *41*, 1–18. [CrossRef]
97. Merrill, N.; Huang, G. Lightweight unsupervised deep loop closure. *arXiv* **2018**, arXiv:1805.07703.
98. Williams, B.; Klein, G.; Reid, I. Automatic relocalization and loop closing for real-time monocular SLAM. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1699–1712. [CrossRef]
99. Qin, T.; Li, P.; Shen, S. Relocalization, global optimization and map merging for monocular visual-inertial SLAM. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 1197–1204.
100. Blochliger, F.; Fehr, M.; Dymczyk, M.; Schneider, T.; Siegwart, R. Topomap: Topological mapping and navigation based on visual slam maps. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 3818–3825.
101. Choset, H.; Nagatani, K. Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization. *IEEE Trans. Robot. Autom.* **2001**, *17*, 125–137. [CrossRef]
102. Chen, Q.; Lu, Y.; Wang, Y.; Zhu, B. From topological map to local cognitive map: A new opportunity of local path planning. *Intell. Serv. Robot.* **2021**, *14*, 285–301. [CrossRef]
103. Thoma, J.; Paudel, D.P.; Chhatkuli, A.; Probst, T.; Gool, L.V. Mapping, localization and path planning for image-based navigation using visual features and map. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7383–7391.
104. Cheng, W.; Yang, S.; Zhou, M.; Liu, Z.; Chen, Y.; Li, M. Road mapping and localization using sparse semantic visual features. *IEEE Robot. Autom. Lett.* **2021**, *6*, 8118–8125. [CrossRef]
105. Li, J.; Zhan, J.; Zhou, T.; Bento, V.A.; Wang, Q. Point cloud registration and localization based on voxel plane features. *ISPRS J. Photogramm. Remote. Sens.* **2022**, *188*, 363–379. [CrossRef]
106. Nüchter, A.; Hertzberg, J. Towards semantic maps for mobile robots. *Robot. Auton. Syst.* **2008**, *56*, 915–926. [CrossRef]
107. Han, L.; Lin, Y.; Du, G.; Lian, S. Deepvio: Self-supervised deep learning of monocular visual inertial odometry using 3d geometric constraints. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 6906–6913.
108. Almalioglu, Y.; Turan, M.; Saputra, M.R.U.; de Gusmão, P.P.; Markham, A.; Trigoni, N. SelfVIO: Self-supervised deep monocular Visual–Inertial Odometry and depth estimation. *Neural Netw.* **2022**, *150*, 119–136. [CrossRef]
109. Silveira, L.; Guth, F.; Drews, P., Jr.; Ballester, P.; Machado, M.; Codevilla, F.; Duarte-Filho, N.; Botelho, S. An open-source bio-inspired solution to underwater SLAM. *IFAC-PapersOnLine* **2015**, *48*, 212–217. [CrossRef]
110. Yuan, X.; Martínez-Ortega, J.F.; Fernández, J.A.S.; Eckert, M. AEKF-SLAM: A new algorithm for robotic underwater navigation. *Sensors* **2017**, *17*, 1174. [CrossRef]

111. Kim, A.; Eustice, R.M. Real-time visual SLAM for autonomous underwater hull inspection using visual saliency. *IEEE Trans. Robot.* **2013**, *29*, 719–733. [CrossRef]
112. Felemban, E.; Shaikh, F.K.; Qureshi, U.M.; Sheikh, A.A.; Qaisar, S.B. Underwater sensor network applications: A comprehensive survey. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 896832. [CrossRef]
113. Köser, K.; Frese, U. Challenges in underwater visual navigation and SLAM. *AI Technol. Underw. Robot.* **2020**, *96*, 125–135.
114. Amarasinghe, C.; Ratnaweera, A.; Maitripala, S. Monocular visual slam for underwater navigation in turbid and dynamic environments. *Am. J. Mech. Eng.* **2020**, *8*, 76–87. [CrossRef]
115. Guth, F.; Silveira, L.; Botelho, S.; Drews, P.; Ballester, P. Underwater SLAM: Challenges, state of the art, algorithms and a new biologically-inspired approach. In Proceedings of the 5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics, Sao Paulo, Brazil, 12–15 August 2014; pp. 981–986.
116. Muhammad, N.; Strokina, N.; Toming, G.; Tuhtan, J.; Kämäräinen, J.K.; Kruusmaa, M. Flow feature extraction for underwater robot localization: Preliminary results. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 1125–1130.
117. Cong, Y.; Gu, C.; Zhang, T.; Gao, Y. Underwater robot sensing technology: A survey. *Fundam. Res.* **2021**, *1*, 337–345. [CrossRef]
118. Negre, P.L.; Bonin-Font, F.; Oliver, G. Cluster-based loop closing detection for underwater slam in feature-poor regions. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 2589–2595.
119. Hu, M.; Li, S.; Wu, J.; Guo, J.; Li, H.; Kang, X. Loop closure detection for visual SLAM fusing semantic information. In Proceedings of the 2019 Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 4136–4141.
120. Mukherjee, A.; Chakraborty, S.; Saha, S.K. Detection of loop closure in SLAM: A DeconvNet based approach. *Appl. Soft Comput.* **2019**, *80*, 650–656. [CrossRef]
121. Techy, L.; Morganseny, K.A.; Woolseyz, C.A. Long-baseline acoustic localization of the Seaglider underwater glider. In Proceedings of the 2011 American Control Conference, San Francisco, CA, USA, 29 June–1 July 2011; pp. 3990–3995.
122. Duarte, A.C.; Zaffari, G.B.; da Rosa, R.T.S.; Longaray, L.M.; Drews, P.; Botelho, S.S. Towards comparison of underwater SLAM methods: An open dataset collection. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, USA, 19–23 September 2016; pp. 1–5.
123. Mallios, A.; Vidal, E.; Campos, R.; Carreras, M. Underwater caves sonar data set. *Int. J. Robot. Res.* **2017**, *36*, 1247–1251. [CrossRef]
124. Rahman, S.; Karapetyan, N.; Li, A.Q.; Rekleitis, I. A modular sensor suite for underwater reconstruction. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018; pp. 1–6.
125. Ferrera, M.; Creuze, V.; Moras, J.; Trouvé-Peloux, P. AQUALOC: An underwater dataset for visual–inertial–pressure localization. *Int. J. Robot. Res.* **2019**, *38*, 1549–1559. [CrossRef]
126. Zwilgmeyer, P.G.O.; Yip, M.; Teigen, A.L.; Mester, R.; Stahl, A. The varos synthetic underwater data set: Towards realistic multi-sensor underwater data with ground truth. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3722–3730.
127. Krasnosky, K.; Roman, C.; Casagrande, D. A bathymetric mapping and SLAM dataset with high-precision ground truth for marine robotics. *Int. J. Robot. Res.* **2022**, *41*, 12–19. [CrossRef]
128. Song, C.; Zeng, B.; Su, T.; Zhang, K.; Cheng, J. Data association and loop closure in semantic dynamic SLAM using the table retrieval method. *Appl. Intell.* **2022**, *52*, 11472–11488. [CrossRef]
129. Cieslewski, T.; Choudhary, S.; Scaramuzza, D. Data-efficient decentralized visual SLAM. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 2466–2473.
130. Carreno, Y.; Willners, J.S.; Petillot, Y.; Petrick, R.P. Situation-Aware Task Planning for Robust AUV Exploration in Extreme Environments. In Proceedings of the IJCAI Workshop on Robust and Reliable Autonomy in the Wild, Montreal, QC, Canada, 19 August 2021.
131. Watson, S.; Duecker, D.A.; Groves, K. Localisation of unmanned underwater vehicles (UUVs) in complex and confined environments: A review. *Sensors* **2020**, *20*, 6203. [CrossRef]
132. Aitken, J.M.; Evans, M.H.; Worley, R.; Edwards, S.; Zhang, R.; Dodd, T.; Mihaylova, L.; Anderson, S.R. Simultaneous localization and mapping for inspection robots in water and sewer pipe networks: A review. *IEEE Access* **2021**, *9*, 140173–140198. [CrossRef]
133. Xiao, L.; Wang, J.; Qiu, X.; Rong, Z.; Zou, X. Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment. *Robot. Auton. Syst.* **2019**, *117*, 1–16. [CrossRef]
134. Sun, Y.; Liu, M.; Meng, M.Q.H. Improving RGB-D SLAM in dynamic environments: A motion removal approach. *Robot. Auton. Syst.* **2017**, *89*, 110–122. [CrossRef]
135. Ni, J.; Wang, X.; Gong, T.; Xie, Y. An improved adaptive ORB-SLAM method for monocular vision robot under dynamic environments. *Int. J. Mach. Learn. Cybern.* **2022**, *13*, 3821–3836. [CrossRef]
136. Chen, C.; Wang, B.; Lu, C.X.; Trigoni, N.; Markham, A. A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence. *arXiv* **2020**, arXiv:2006.12567.
137. Yu, C.; Liu, Z.; Liu, X.J.; Xie, F.; Yang, Y.; Wei, Q.; Fei, Q. DS-SLAM: A semantic visual SLAM towards dynamic environments. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1168–1174.

138. Himri, K.; Ridao, P.; Gracias, N.; Palomer, A.; Palomeras, N.; Pi, R. Semantic SLAM for an AUV using object recognition from point clouds. *IFAC-PapersOnLine* **2018**, *51*, 360–365. [CrossRef]
139. Wen, S.; Li, P.; Zhao, Y.; Zhang, H.; Sun, F.; Wang, Z. Semantic visual SLAM in dynamic environment. *Auton. Robot.* **2021**, *45*, 493–504. [CrossRef]
140. Chang, Y.; Ebadi, K.; Denniston, C.E.; Ginting, M.F.; Rosinol, A.; Reinke, A.; Palieri, M.; Shi, J.; Chatterjee, A.; Morrell, B.; et al. LAMP 2.0: A robust multi-robot SLAM system for operation in challenging large-scale underground environments. *IEEE Robot. Autom. Lett.* **2022**, *7*, 9175–9182. [CrossRef]
141. Pire, T.; Baravalle, R.; D'alessandro, A.; Civera, J. Real-time dense map fusion for stereo SLAM. *Robotica* **2018**, *36*, 1510–1526. [CrossRef]
142. Paull, L.; Huang, G.; Seto, M.; Leonard, J.J. Communication-constrained multi-AUV cooperative SLAM. In Proceedings of the 2015 IEEE international conference on robotics and automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 509–516.
143. Ji, P.; Li, X.; Gao, W.; Li, M. A Vision Based Multi-robot Cooperative Semantic SLAM Algorithm. In Proceedings of the 2022 34th Chinese Control and Decision Conference (CCDC), Hefei, China, 15–17 August 2022; pp. 5663–5668.

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