REAL-TIME POSITIONING AND TRACKING FOR VISION-BASED UNMANNED UNDERWATER VEHICLES

Jiangying Qin[1,*], Ke Yang[2], Ming Li[1,3,*], Jiageng Zhong[1], Hanqi Zhang[1]

1 State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China - (jy_qin, zhongjiageng, hqzhang)@whu.edu.cn
2 School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan 430079, China - YyangkeK@whu.edu.cn
3 Department of Physics, ETH Zurich, Zurich 8039, Switzerland - mingli39@ethz.ch

Commission IV, WG IV/5

KEY WORDS: Unmanned Underwater Vehicle (UUV), Vision Positioning, Tracking, Deep Learning, Underwater Measurement, SLAM.

ABSTRACT:

Unmanned underwater vehicle (UUV) is a key technology for marine resource exploration and ecological monitoring. How to use vision-based active positioning and three-dimensional perception to realize UUV underwater autonomous navigation and positioning is the basis for UUV's underwater operations. The complexity and unstructured characteristics of seawater bring new challenges to vision-based underwater high-precision positioning. Traditional visual localization algorithms mainly include geometric-based visual localization algorithms (such as ORB-SLAM2) and deep learning-based visual localization algorithms (such as DXSLAM). In this paper, based on the typical marine environment (low brightness, dynamic fish interference, underwater light spot, high turbidity), the experimental analysis and comparison of different visual positioning methods of UUV is carried out, which provides a reference for realizing the real-time localization of UUV, and further provides a better solution for UUV underwater measurement and monitoring operations.

1. INTRODUCTION

In recent years, people have become more and more aware of the importance of the ocean in the fields of resources and ecology, and their interest in ocean exploration has also increased day by day. Unmanned underwater vehicle (UUV) technology is a key technology for marine resource exploration and ecological monitoring. Due to its significant cost advantage, high mobility and ability to complete various complex underwater tasks, it has become the first choice for underwater surveying operations (Chen, 2021; Sahoo et al., 2019). The ability to perceive the three-dimensional environment is the key to UUV underwater operations, which must be solved for UUV's safe navigation and multi-cooperative tasks. High-precision positioning is the basis for realizing environmental perception, and it is also the core issue of ocean exploration and detection.

Conventional UUVs provide noise estimates of motion by being equipped with an Inertial Measurement Unit (IMU) or a Doppler Velocity Log (DVL). Inertial Navigation System (INS) obtains direction information and speed information through inertial sensors to calculate the moving distance of UUV. This method is suitable for long-range tasks and has the advantage of a passive approach, i.e. no signals need to be sent or received from external systems. It does not rely on external references and is widely used in most underwater robot positioning and navigation scenarios. However, the error of inertial sensor will increase over time, eventually leading to significant drift in motion estimates (Mu et al., 2019; Jalal and Nasir, 2021). This type of drift is caused by factors such as ocean currents and the accuracy of the sensor itself, which cannot sense displacement caused by external forces or earth gravity. A possible solution is to use geophysical maps to match sensor measurements, also known as geophysical navigation (GN), which allows longer missions to be completed while maintaining relatively low position errors. However, this method needs to provide a geophysical map, and comparing and matching the map with sensor data will result in high computational cost, which is one of the main reasons for restricting the development of GN (González et al., 2020; Rice et al., 2004). Acoustic Beacon-based System (ABS) has become an effective choice for underwater positioning because of its large sensing range. It obtains the actual position by measuring the flight time of acoustic signals. Acoustic systems are mostly implemented using acoustic repeaters, most of which require complex infrastructure, high deployment costs, expensive sensors, low resolution, and lack of semantic information, making it difficult to meet diverse semantic localization requirements. On the other hand, the velocities of light need to be carefully calibrated before using acoustic localization systems because they are affected by multipath Doppler effects and susceptibility to temperature rise. In addition, for small targets or dynamic targets, acoustic sensors have the problems of difficulty in feature extraction, poor detection accuracy, and high false detection rate, and it is difficult to obtain fine targets, which has also become one of the main reasons that restrict the application of acoustic sensors in the field of marine intelligent detection (Maurelli et al., 2021; Cong et al., 2021). The visual method solves this problem very well. It has rich semantic information, and has clear targets and high resolution. Especially in the coastal zone with good lighting environment, it can not only avoid noise interference caused by the influence of underwater landforms on acoustic sensors, but also collect rich texture and semantic information such as corals and fish when providing environmental perception and positioning information to underwater UUVs. Therefore, in recent years, visual sensors have been increasingly applied in the fields of underwater

mapping and marine life conservation (Hozyn and Zak, 2021; Mohammed et al., 2021).

Due to the unstructured nature of seawater, the influence of complex ocean currents, fluid resistance, and the inability to use GPS signals, traditional terrestrial measurement and remote sensing methods are difficult to directly apply to the underwater environment, which brings new challenges for vision-based underwater high-precision positioning (Xing et al., 2021; Zhu et al., 2020). How to use vision-based real-time perception and active positioning capabilities to realize underwater autonomous positioning and navigation of UUVs is one of the current important tasks.

Based on this, this paper mainly studies the vision-based high-precision positioning of underwater UUV, visual simultaneous localization and mapping (VSLAM), deep learning and other technologies to achieve real-time underwater positioning and 3D perception, and further realize autonomous navigation, positioning and automatic acquisition, and provide a technical basis for multi-task underwater monitoring and surveying.

2. RELATED WORK

Traditional vision-based localization methods are mainly divided into two categories: geometric-based visual localization methods and deep learning-based visual localization methods (Li et al., 2021).

Geometry-based visual localization methods require a pre-built 3D model of the scene. In the positioning process, a 2D-3D matching relationship is established by matching the feature points of the current image frame and the scene model (Lowe, 2004; Bay et al., 2008; Rubleeet al., 2011), RANSAC is used to eliminate the mismatched points (Fischler and Bolles,1981; Chum and Matas, 2005), and finally the PnP algorithm is used to calculate the 6-DOF camera pose (Hesch and Roumeliotis, 2011). There are many classical algorithms in the field of geometry-based visual localization. ORB-SLAM is a monocular SLAM system proposed by Raul et al., in 2015 (Mur-Artal et al., 2017). Based on the PTAM architecture, it adds the functions of map initialization and closed-loop detection, and optimizes the method of key frame selection and map construction. It achieves good results in terms of processing speed, tracking and map accuracy. On the basis of ORB-SLAM, Raul et al. proposed the ORB-SLAM2 algorithm in 2017, which is a complete set of SLAM solutions based on monocular, binocular and RGB-D cameras (Mur-Artal et al., 2017). It can realize the functions of map reuse, loop detection and relocation, and is one of the most excellent geometric-based visual localization algorithms. The ORB-SLAM3(Campos et al., 2021) algorithm is a new SLAM framework proposed by Carlos et al. in 2021. Compared with the monocular version of ORB-SLAM and the stereo version of ORB-SLAM2, ORB-SLAM3 adds an IMU fusion algorithm. It is the first system capable of visual, visual-inertial and multi-map SLAM with monocular, binocular and RGB-D cameras, pinhole and fisheye lens models. Under ideal conditions, the geometry-based visual localization method can accurately estimate the camera pose with high localization accuracy. However, in real-world scenarios, limited by correct and sufficient feature point matching, its localization robustness is poor. Inaccurate camera calibration, inaccurate system modeling, and complex environments (such as dynamic targets, missing textures, and complex lighting) will lead to poor localization accuracy or even impossible localization.

In recent years, deep learning-based visual localization methods have attracted widespread interest. Different from traditional methods that rely on geometric models and mapping relationships to achieve localization tasks, the deep learning-based method proposes a data-driven solution that uses the learning model to construct a mapping function and then regress the camera pose (Chen et al., 2020). Deep learning methods can automatically discover task-relevant features using highly expressive neural networks, which also makes them better suited to various environments that may exist. Deep learning-based visual localization methods can be divided into two categories. The first is to use deep learning to regress the 6DOF camera pose in an end-to-end manner. PoseNet is the first attempt to perform end-to-end global pose regression using convolutional neural networks, it traines neural networks from a single frame of RGB images to regress a 6-DOF camera pose (Kendall et al., 2015). (Walch et al., 2017) proposes a novel CNN+LSTM architecture for camera pose regression for indoor and outdoor scenes. Among them, CNN is used to learn a suitable and robust localization feature representation, and LSTM plays the role of structured dimensionality reduction on the feature vector to improve the localization performance. (Radwan et al., 2018) adopts a multi-task learning approach to exploit the relationship between learned semantics, regression 6-DoF global pose and odometry. (Debaditya et al., 2019) proposes to fine-tune a deep convolutional neural network (DCNN) using synthetic images obtained from a 3D indoor model to regress camera pose. The second category is to use deep learning methods to replace one or more modules in the traditional geometry-based visual localization. On the basis of retaining the traditional geometrical visual localization framework, the introduction of neural networks can further improve its localization performance. (Li et al., 2020) proposes a complete deep learning-based SLAM system that is robust to changes in environment and perspective. It uses HF-Net to extract keypoints, local descriptors and global descriptors of each image, and proposes a global descriptor-based relocalization method. (Tang et al., 2019) proposes to use the GCNv2 deep learning network to generate keypoints and descriptors, using binary descriptor vectors with the same descriptor format as ORB functions, so that it can be used as an alternative in SLAM systems. Visual localization algorithm based on deep learning provides a new possibility for traditional visual localization. Based on its data-driven and high generalization characteristics, it is more robust to complex environments that may exist in real life, and can better adapt to various unstructured environments. However, the accuracy of this type of method is lower than that of the geometry-based visual localization algorithm, which is also one of the important directions for the subsequent improvement.

Based on this, this paper intends to solve the problem that the current UUV still lacks the ability of intelligent visual real-time perception and active positioning by researching key technologies such as UUV's navigation positioning and visual perception. Specifically, this paper compares the classical real-time visual localization algorithms. By simulating various complex visual problems that may exist underwater, experimentally analyze the accuracy, robustness and time performance of different algorithms, so as to provide technical and theoretical reference for the realization of underwater real-time positioning of UUV.

3. METHODOLOGY

At present, most of the geometric-based visual localization algorithms and deep learning-based visual localization algorithms are based on the ground structured environment, and there are few studies on the underwater environment. Based on this, this paper selects the classical geometry-based visual localization algorithm and the deep learning-based visual localization algorithm for research experiments, and compares their localization performance in different underwater environments. Specifically, the geometry-based visual localization algorithm selects the classic ORB-SLAM2 algorithm, and selects DXSLAM to represent the deep learning-based visual localization algorithm. In addition, since this experiment does not set up control points underwater, this paper compares the camera trajectory of the aerial triangulation process as the true value of the trajectory with the visual localization algorithm.

3.1 ORB-SLAM2

ORB-SLAM2 is a complete set of SLAM schemes based on monocular, binocular and RGB-D cameras proposed by Mur-Artal et al. in 2017. It basically continues the algorithm framework of PTAM and can realize the functions of map reuse, loop closure detection and relocation. It has been widely used due to its advantages such as perfection and good generalization.

ORB-SLAM2 innovatively uses three threads to implement SLAM, the tracking thread for real-time tracking of feature points, the local mapping thread for local Bundle Adjustment and the loopback detection thread for global pose graph. Among them, the tracking thread mainly extracts ORB feature points for each image, and compares them with the nearest key frame, calculates the position of the feature points and roughly estimates the camera pose. Or initialize the pose by global relocalization, and then track the reconstructed local map to optimize the pose. The local mapping thread mainly completes the local map construction. This includes inserting keyframes, validating and filtering recently generated map points, and then generating new map points. It uses Bundle Adjustment to solve more accurate camera poses and spatial positions of feature points. The process of visual odometry is completed by the tracking thread and the local mapping thread. The loopback detection thread performs loopback detection on the global map and keyframes to eliminate accumulated errors. Since there are too many map points in the global map, the optimization of this thread does not include map points, but only pose graphs composed of camera poses. The unique three-thread structure of the ORB-SLAM series has achieved good tracking and mapping effects, and can ensure the global consistency of the trajectory and the map, so it has been widely researched and applied.

The main innovations of ORB-SLAM2 are: (1) It is the first open-source SLAM system based on monocular, binocular, and RGB-D cameras, including loop detection, relocation, and map reuse functions; (2) ORB-SLAM2 is based on BA optimization, which achieves higher accuracy than cutting-edge methods based on closest point iteration (ICP), optical and depth error minimization; (3) By using both far and near binocular points and monocular observations, the accuracy is higher than that of directly using the binocular SLAM method; (4) A lightweight relocation mode is proposed, which can achieve effective map reuse in areas that cannot be mapped.

3.2 DXSLAM

DXSLAM is a complete deep learning-based SLAM system proposed by Li et al., which is robust to changes in environment and perspective. It uses image global features for relocalization and uses a new loop closure detection system. Also, DXSLAM is not GPU dependent and can run on CPU.

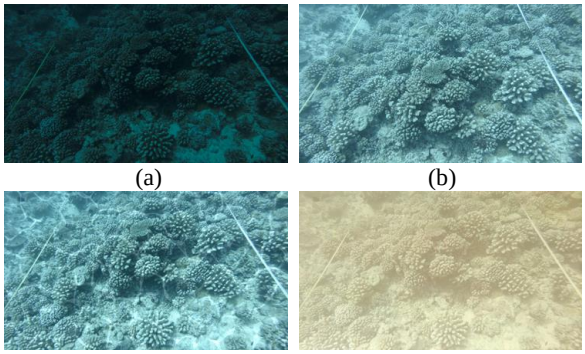
The overall framework of DXSLAM is similar to ORB-SLAM2, and the difference is mainly in the feature point extraction part. It uses HF-Net to extract features from each image frame. The image is first passed through a shared encoder and then through three parallel decoders, which predict keypoint detection scores, dense local descriptors, and global descriptors, respectively. Local features are mainly used for positioning and mapping processes, and global features are mainly used to build an efficient relocation module for fast relocation when system initialization or tracking fails. This method uses a trained bag-of-words model for local feature matching. Further, in order to reduce the system initialization time, FBoW is used to replace the traditional BoW, which greatly improves the system efficiency. Furthermore, since the BoW matching method aggregates local features through the distribution of local features, ignoring their spatial relationships, false matching may occur. Therefore, DXSLAM establishes a highly reliable loop closure detection method based on local features, global features and bag of words.

The main innovations of DXSLAM are: (1) Using HF-Net to extract feature points, the SLAM system has better robustness in the case of changes in the environment and perspective; (2) The introduction of global features makes the system relocation more robust. In addition, FBoW has a higher success rate and a smaller amount of computation than traditional BoW; (3) A loop closure detection method based on global and local features is proposed; (4) It is the first SLAM system based on the deep learning feature point method and can run without GPU.

4. EXPERIMENTS

4.1 Experimental Data and Computing Environment

In order to compare the underwater positioning performance of the two algorithms, this paper conducts experiments on the two algorithms in typical underwater scenes. In the experiment, the processor used is Intel(R) Core (TM) i7-8750H, the memory is 8GB, and the GPU used is GeForce GTX 1060. The dataset includes a variety of typical underwater environmental characteristics, such as low brightness (as shown in figure 1a), dynamic fish interference (as shown in figure 1b), underwater light spots (as shown in figure 1c), and high turbidity (as shown in figure 1d), so as to realize the simulation of real complex underwater environment.



(a) (b) (c) (d)

Figure 1. Schematic of the underwater dataset

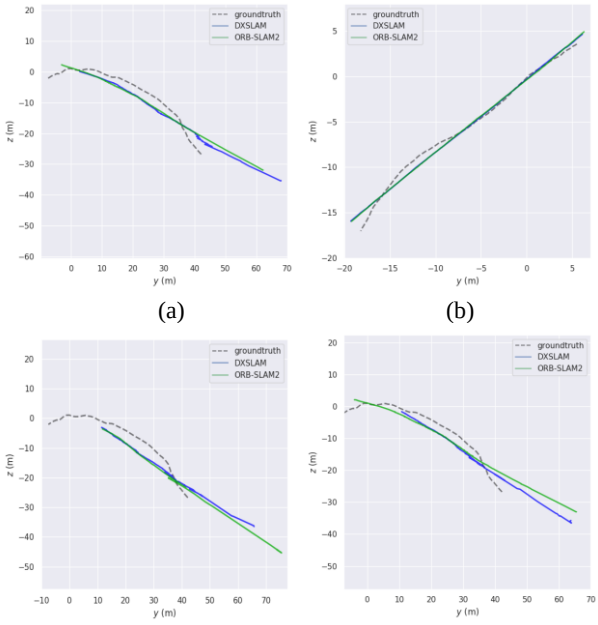
4.2 Experimental Results and Analysis

Table 1 shows the average accuracy results obtained by localizing different underwater datasets with the two methods. For low brightness datasets, the average accuracy of ORB-SLAM2 and DXSLAM are 0.52m and 1.01m, respectively, and the accuracy of ORB-SLAM2 is significantly higher than that of DXSLAM; For the dynamic fish interference dataset, the accuracy of the two is 1.15m and 1.00m, and the DXSLAM positioning accuracy is higher; For the underwater light spots dataset, the accuracy of the two is 0.31m and 0.23m respectively, both of which have obtained higher accuracy but the DXSLAM positioning effect is better; For the high turbidity dataset, the localization accuracy of ORB-SLAM2 is 0.46m, and the localization accuracy of DXSLAM is 0.71m, ORB-SLAM2 achieves better localization results.

	ORB-SLAM2	DXSLAM
low brightness(a)	0.52	1.01
dynamic fish interference(b)	1.15	1.00
underwater light spots(c)	0.31	0.23
high turbidity(d)	0.46	0.71

Table 1. Average precision comparison table (m)

Figure 2 is a comparison diagram of the trajectories calculated by different methods on four sets of data and the real trajectories, in which the black dotted line represents the real trajectory; the blue line represents the trajectory calculated by DXSLAM, and the green line represents the camera trajectory calculated by ORB-SLAM2. On the whole, there are different degrees of trajectory drift in the four sets of data, which may be related to the complexity of the underwater environment. Specifically, for low brightness datasets, the trajectory of ORB-SLAM2 is closer to the groundtruth with less drift at the start and end points. It is worth noting that at the abscissa 40m-45m, the trajectory solved by DXSLAM is jagged, and there is an obvious problem of sudden movement. For the dynamic fish interference dataset, both trajectories have a large drift compared with the real trajectories, especially in the range of -20m to 10m on the abscissa, which may be related to the fact that there are more fish in this part of the video and the movement speed is faster. For the underwater light spot dataset, the error of the two is relatively small, which indicates that the light spot has little influence on the visual positioning, and the DXSLAM trajectory is relatively closer to the real trajectory. Especially at the end point ORB-SLAM2 drifts more and DXSLAM is closer to the true trajectory. For high turbidity datasets, ORB-SLAM2 has better localization results and is closer to the ground truth.

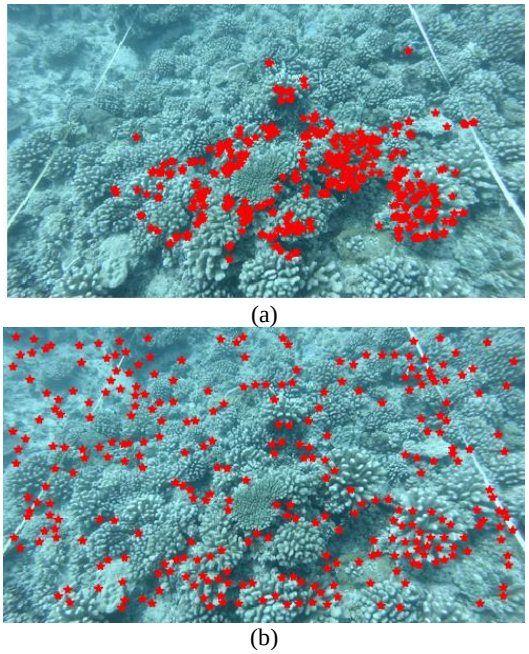


(a) (b) (c) (d)

Figure 2. Comparison of trajectories

Overall, ORB-SLAM2 has higher localization accuracy for low brightness and high turbidity datasets, while DXSLAM localization performs better for dynamic fish interference and underwater light spots datasets. This is because for structured environments such as low brightness and high turbidity, the feature points extracted by ORB-SLAM2 are better, providing better initial conditions for a series of threads including feature matching and pose calculation, and then higher positioning accuracy is achieved. For unstructured environments such as dynamic fish interference and underwater light spots, the randomness and uncertainty of motion and light spots in the environment will lead to large errors in the extraction and matching of feature points. This also shows that the feature extraction method based on deep learning is more robust to complex unstructured environments.

This paper presents the feature extraction results for the dynamic fish interference dataset, as shown in Figure 3. Figure 3a is the feature extraction result of ORB-SLAM2 while figure 3b is the feature point extraction result of DXSLAM, and the red star represents the image feature points extracted by the two algorithms. The interference of dynamic fish brings challenges to traditional geometry-based feature point extraction methods. The specific performance is that in some scenes, the feature points extracted by ORB-SLAM2 are clustered and distributed in the middle of the image, which is also the area where the fish moves, and there are almost no feature points around the image. The feature points extracted by DXSLAM are more evenly distributed, which is more in line with the requirements of high-precision and robust positioning.



(a)

(b)

Figure 3. Comparison of feature extraction

5. CONCLUSIONS

This paper mainly studies the underwater UUV real-time positioning based on vision, and then realizes the underwater autonomous navigation of UUV. Specifically, this paper compares the key performances such as localization accuracy and robustness of the classical geometry-based visual localization method ORB-SLAM2 and the deep learning-based visual localization method DXSLAM in typical underwater environments. The experimental results show that the positioning accuracy of ORB-SLAM2 is higher for the structured environment, but for some unstructured environments such as dynamic fish interference, the robustness of the feature points extracted by ORB-SLAM2 is poor, which affects the positioning accuracy and causes larger errors. For DXSLAM, its localization accuracy is lower than ORB-SLAM2 in structured environments, but it shows stronger robustness in complex scenes. In addition, the experimental results show that the two positioning methods have different degrees of trajectory drift, which also confirms the necessity of focusing on the complexity of underwater environment, such as image preprocessing, so as to realize high-precision real-time UUV positioning and environment perception.

ACKNOWLEDGEMENTS

This research was funded by the National Key R&D Program of China, grant numbers 2018YFB0505400, the National Natural Science Foundation of China (NSFC), grant number 41901407 and the College Students' Innovative Entrepreneurial Training Plan Program, Research on visual navigation, perception and localization algorithm of unmanned underwater vehicle/robot (UUV).

REFERENCES

Bay, H., Ess, A., Tuytelaars, T.,Van, G.L., 2008. Speededup robust features (SURF). Computer vision and image understanding, 110(3), 346–359.

Campos, C., Elvira. R., Rodríguez, J.J.G., Montiel, J.J.M., Tardós, J.D., 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. IEEE Transactions on Robotics, 37(6), 1874-1890.

Chen, C., 2021. Review of Underwater Sensing Technologies and Applications. Sensors, 21(23), 7849-7877.

Sahoo, A., Dwivedy, S.K., Robi, P.S., 2019. Advancements in the field of autonomous underwater vehicle. Ocean engineering, 181(1), 145-160.

Chen, C., Wang, B., Lu, C.X., Trigoni, N., Markham, A., 2020. A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence. arXiv preprint, 2006.12567.

Cong, Y., Gu, C., Zhang, T., Gao, Y., 2021. Underwater robot sensing technology: A survey. Fundamental Research, 1(3), 337-345.

Chum, O., Matas, J., 2005. Matching with PROSAC-progressive sample consensus. Computer vision and pattern recognition (CVPR), IEEE, 1(1), 220-226.

Debaditya, A., Kourosh, K., Stephan, W., 2019. BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. ISPRS Journal of Photogrammetry and Remote Sensing, 150(1), 245-258.

Fischler, M., Bolles, R., 1981. Random sample consensus: A para-digm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381-395.

González-García, J., Gómez-Espinosa, A., Cuan-Urquizo, E., García-Valdovinos, L.G., Salgado-Jiménez, T., Escobedo, J.A.E., 2020. Autonomous underwater vehicles: Localization, navigation, and communication for collaborative missions. Applied sciences, 10(4), 1256-1293.

Hesch, J.A., Roumeliotis, S.I., 2011. A direct least-squares (DLS) method for PnP. International Conference on Computer Vision (ICCV), IEEE, 383-390.

Hożyń, S., Żak, B., 2021. Stereo Vision System for Vision-Based Control of Inspection-Class ROVs. Remote Sensing, 13(24), 5075-5100.

Jalal, F., Nasir, F., 2021. Underwater navigation, localization and path planning for autonomous vehicles: A review. International Bhurban Conference on Applied Sciences and Technologies (IBCAST), IEEE, 2021(1), 817-828.

Kendall, A., Grimes, M., Cipolla, R, 2015. PoseNet: A convolutional network for real-time 6-dof camera relocalization. International Conference on Computer Vision (ICCV), IEEE, 2938-2946.

Li, D., Shi, X., Long, Q., Liu, S., Wang, F., Wei, Q., Qiao, F., 2020. DXSLAM: A robust and efficient visual SLAM system with deep features. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 4958-4965.

Li, M, Qin, J, Li, D, Chen, R., Liao, X., Guo, B., 2021. VNLSTM-PoseNet: a novel deep ConvNet for real-time 6-DOF camera relocalization in urban streets. Geo-spatial Information Science, 24(3), 422-437.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2), 91–110.

Maurelli, F., Krupiński, S., Xiang, X., Petillot, Y., 2021. AUV localisation: a review of passive and active techniques. International Journal of Intelligent Robotics and Applications, 2021(1), 1-24.

Mohammed, A., Kvam, J., Thielemann, J.T., Haugholt, K.H., Risholm, P., 2021. 6D Pose Estimation for Subsea Intervention in Turbid Waters. Electronics, 10(19), 2369-2382.

Mu, X., He, B., Zhang, X., Song, Y., Shen, Y., Feng, C., 2019. End-to-end navigation for autonomous underwater vehicle with hybrid recurrent neural networks. Ocean Engineering, 194(1), 106602-106611.

Mur-Artal, R., Montiel, J.M.M., Tardos, J.D., ORB-SLAM: A Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, 31(5), 1147-1163.

Mur-Artal, R., Tardós, J.D., 2017. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. IEEE Transactions on Robotics, 33(5), 1255-1262.

Radwan, N., Valada, A., Burgard, W., 2018. VLocNet++: Deep Multitask Learning for Semantic Visual Localization and Odometry. IEEE Robotics and Automation Letters, 3(4), 4407-4414.

Rice, H., Kelmenson, S., Mendelsohn, L., 2004. Geophysical navigation technologies and applications. Position Location and Navigation Symposium (PLANS), IEEE, 4(1), 618-624.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. International conference on computer vision, IEEE, 2564-2571.

Tang, J., Ericson, L., Folkesson, J., Jensfelt, P., 2019. GCNv2: Efficient correspondence prediction for real-time SLAM. IEEE Robotics and Automation Letters, 4(4), 3505-3512.

Walch, F., Hazirbas, C., Leal-Taixe, L., Torsten, S., Sebastian, H., Daniel, C., 2017. Image-based localization using LSTMs for structured feature correlation. International Conference on Computer Vision (ICCV), IEEE, 627-637.

Xing, H., Liu, Y., Guo, S., Shi, L., Hou, X., Liu, W., Zhao, Y., 2021. A Multi-Sensor Fusion Self-Localization System of a Miniature Underwater Robot in Structured and GPS-Denied Environments. Sensors, 21(23), 27136-27146.

Zhu, P., Yao, S., Liu, Y., Liu, S., Liang, X., 2020. Autonomous Reinforcement Control of Underwater Vehicles based on Monocular Depth Vision. IFAC-PapersOnLine, 53(2), 9201-9206.