PERINTIS eJournal

PERINTIS eJournal, 2022, Vol. 12 No. 1, pp. 23-34

Visual Simultaneous Localization and Mapping: A review

Article history

Received:
1 May 2022

Received in revised form:
31 May 2022

Accepted:
15 June 2022

Published online:
30 June 2022

*Corresponding author
athif@utm.my

Muhammad Razmi Razali[1], Ahmad Athif Mohd Faudzi[1,2*], Abu Ubaidah Shamsudin[3]

[1]*School of Electrical Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, Johor Bahru, Malaysia*
[2]*Centre for Artificial Intelligence and Robotics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia*
[3]*Fakulti Kejuruteraan Elektrik dan Elektronik, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia*

muhammadrazmi@graduate.utm.my, *athif@utm.my*, *ubaidah@uthm.edu.my*

ABSTRACT

The complexity of Simultaneous Localization and Mapping (SLAM) and drastic technological changes makes development of an autonomous robot system a challenge. As the market demand and awareness of autonomous robot systems are increased, intensive knowledge and information of this technology are essential. The fluidity of information has resulted in a mismatch between SLAM and Visual SLAM where the current approaches are having problems in visual odometry and sensor inputs that describe the common characteristics with different techniques due to the lack of formalism in the knowledge on the localization and mapping. This has caused confusion to the researchers leading to incorrect development of autonomous robot system. Therefore, this review aims to describe and assist researchers to identify the characteristics of SLAM and in the context of the visual odometry characteristics and perception techniques. This review will also discuss the attributes of the visual odometry and the techniques of autonomous robot system from various types of application based on the semantic structure. It is expected that the Visual Simultaneous Localization and Mapping (VSLAM) will give a positive impact to the development of autonomous robot system to meet the market demand.

Keywords: *VSLAM, visual odometry, autonomous robot*

1.0 INTRODUCTION

Simultaneous Localization and Mapping (SLAM) has grown rapidly in the last decade (SLAM). Autonomous mobile robots can use it for a wide range of purposes, including self-exploration in a variety of geographical settings. Resurgent interest in self exploratory autonomous mobile robots has been sparked by recent advances in SLAM. With the SLAM technique, a robot/sensor system uses sensors to gather information about its surroundings and then predicts its position in the environment [1]. To deploy SLAM to be used in the actual world, numerous SLAM algorithms have been developed since its inception. The fundamental problem posed by locating sensors in a global representation has been addressed by a variety of SLAM modalities, including radars, range finders, cameras, and lasers.

As a result, the robot's mission to explore an unfamiliar terrain while avoiding the numerous landmarks and obstacles that it comes across can be quite challenging. Debate rages on concerning the best ways to deal with the high computational cost of combining too many sensors at once from research done by [2]. The goal of this study is to review if limitations are attached no matter which sensor is utilized. According to this research [2], the SLAM algorithms used to address the problem are as problematic as the sensors in SLAM's failure. In light of these advantages, the SLAM technique has seen an uptick in popularity over the past few years, which shows that robot operation can be accomplished without the use of ad hoc localization infrastructure, as an alternative to user built maps [3]. A project by [4] generally focused on the building of maps for inspection and navigation for an autonomous robot. For the method, research done by [4] used LiDAR sensor to provide range between robot and obstacle, then DC motor will drive around the robot and Raspberry-pi will transmit data to PC to perform simultaneous localization and mapping.

All methods should have the ability to allow for loop closures or re-localization, as well as pure localization at the very least. As a result, this technique is limited to doing one specific task for the robot, such as mapping or localization. Mobile robot performance in terms of landmark estimation, robot posterior and location estimates, effective path planning and error reduction are the goals of SLAM approaches. It is essential that self-localization and path planning (SLAM) for unmanned systems be achieved in the field of driverless technology and other mobile robots and drones, as well as in augmented reality and virtual reality [5].

2.0 VSLAM vs SLAM

2.1 VISUAL SLAM (VSLAM)

A significant amount of literature has been published on SLAM. These investigations are aimed towards creating a reasonably continuous map of an area and at determining the location of landmarks and robots on that map. SLAM has also been the focus of empirical investigations that study how the platform of trajectory technique and landmark's locations are integrated and computed online. SLAM studies have been conducted on the topic of localization, and numerous strategies have been presented as solutions [6], [7].

It is becoming increasingly in common in the robotics world to create and explore vision based navigation systems [8]. For example, a real time 3D reconstruction of the environment would allow autonomous cars to function in locations where optical odometry is unavailable (map) [9]. Vision odometry (VO) can calculate a robot's motion (rotation and translation), allowing it to position itself in its surroundings, according to data from a variety of sources. Onboard vision tracks visual landmarks to estimate motion parameters like rotation and translation between two-time instants in the SLAM system's geographic environment. This theory has a major flaw. Visual odometry and low visibility are the principal obstacles in autonomous systems research, limiting sensor input information and restricting the robot's ability to operate.

As an example, the onboard illumination can help in low light settings [10], or LiDAR (light detection and range) and thermal imager are two examples of single to multi sensor modalities [11], according to another significant study. It is nevertheless difficult to navigate in low visibility circumstances, such as smoke or fog. Research done by [12] proposed a method for recognizing and removing fog based on 3D point cloud features and a distance correction method for reducing measurement errors. To reduce misjudgment, laser beam penetration features were added. Support vector machine (SVM) and K-nearest neighbor (KNN) are used to classify point cloud data into ‘fog’ and ‘objects’. In extreme conditions, regular cameras, Radars, or LiDAR are used for VO or SLAM, however they will provide bad conditioned data, making it impossible to forecast a stable robot pose therefore fail to produce the map of the environment.

In literature, many visual SLAM (VSLAM) techniques have been developed and researched. However, far fewer have been proposed that are mature enough to be used on robots in real world settings for the long term [13]. Pure localization, re-localization of a lost track, resource efficiency, loop closure, dependability, and support for a wide range of sensor types are all standard features in 2D SLAM, but not always in VSLAM [14]. For current and future mobile robots, implementations that include these features have a lot of traction. Our research is aimed to speed up the usage of VSLAM in service robots, allowing robotics to be deployed in previously uneconomical or nonplanar applications. "Multirole" VSLAM solutions have never proven reliable and feature complete for general use in the mobile robotics and service robotics fields.

It is the purpose of this research to find reliable VSLAM solutions for service, legged, and mobile robots. In this study, the research on vision-based navigation paradigms, such as visual odometry, will be examined. Our review examines the major design aspects of the primary components of each of the aforementioned paradigms, as well as the advantages and disadvantages of each category, where applicable. The remainder of the paper is laid out as follows. Section 3 begins by laying out the theoretical dimensions of the research and looks at the self-localization scheme in Visual SLAM (VSLAM) environment. The evolution of VSLAM schemes is described in Section 4, which highlights the findings of our assessment and identifies future research considerations in the field. Section 5 will be the conclusion of this review paper.

3.0 ODOMETRY TECHNIQUEODOMETRY TECHNIQUE

3.1 VISUAL ODOMETRY (VO)

For this inquiry, it was thought that understanding the concept of VO would be the best approach to use. VO is an example of a structure from motion source film making technique that is used to recreate a 3D scene from a series of frames. Thus, when the robot returns to a previously seen scene, SLAM techniques can reduce the accumulated pose error by employing a global map of robot postures.

In odometry, sensors are used extensively to gather data, which may be from vision, observation, or inertial sensors. With enough light, a static scene is presented with rich textures that aid in monitoring and extracting motion, and when enough scene is overlapped between consecutive frames, VO is effective. The appearance-based VO estimates the camera position by analyzing the intensity of the picture pixels and minimizing the photometric error. For appearance-based voice over paradigms, the essential process is as depicted in Figure 1.

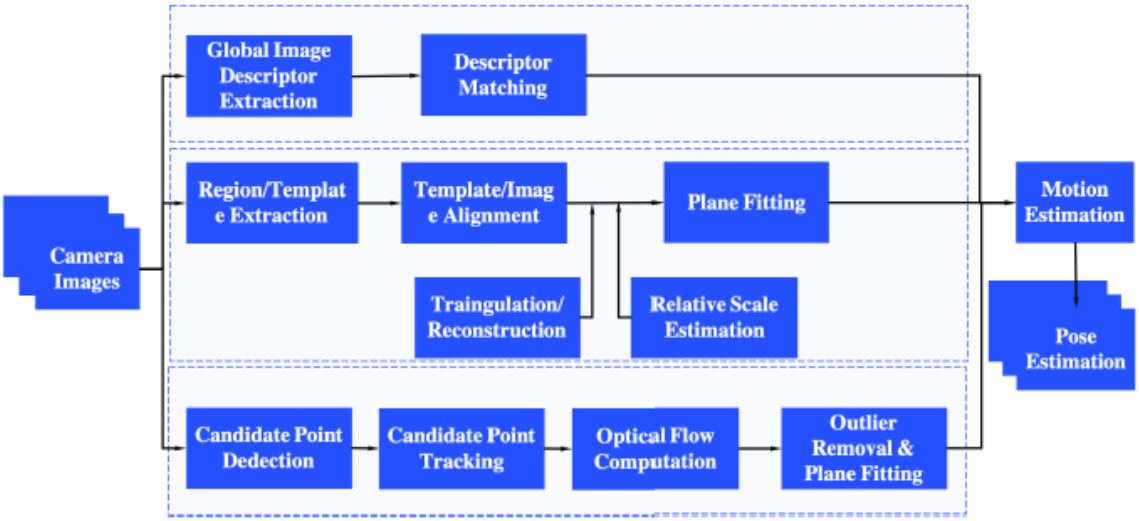

Figure 1 Main pipelines of conventional appearance based [6].

One may separate the VO techniques that use region/template matching and optical flow into two types. Camera poses are concatenated by aligning two consecutive pictures in a regional based motion estimation approach. This technique has been widened in its applicability by analyzing the invariant similarity of tiny areas and using global constraints. As part of the optical flow (OF) based technique, raw visual pixel data is used to assess the pixel intensity change between two consecutive frames from the camera(s) [9]. If a pixel's lighting changes between two frames, the camera's motion would be described by calculating the 2D displacement vector of points projected on the two frames. Some of the limitations of optical based approaches are the strength of the surrounding texture and the computing constraint.

3.2 ARCHITECTURE OF VSLAM

Visual odometry (VO), filtering, graph optimization, loop back detection and mapping are all components of the VSLAM system's architecture, which is comparable to that of the conventional SLAM framework as shown in Figure 2 [15]. There are numerous advantages to the VSLAM technique, as previously stated. In order to support and expand the qualitative analysis, quantitative metrics were examined.

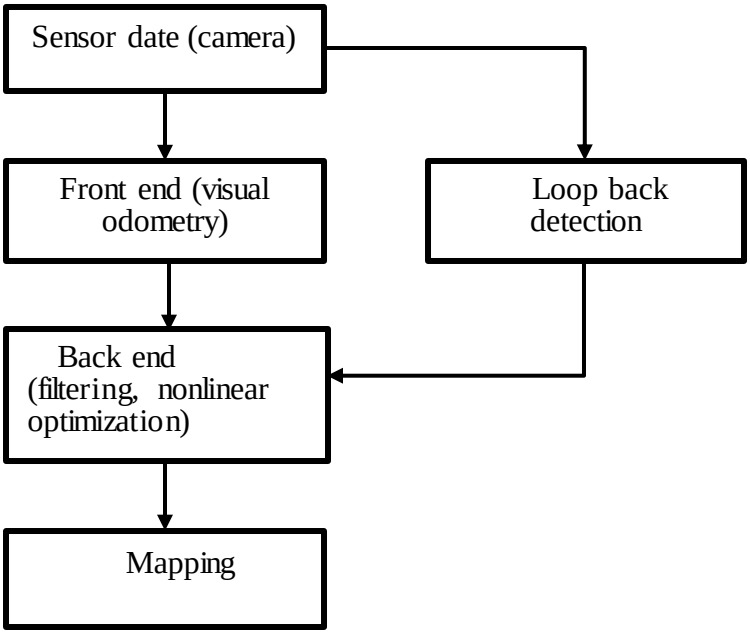

Figure 2 Architecture of Visual SLAM [15]

In the front end, the camera data is abstracted into an easy to interpret model known as VO. As a result, a global trajectory and map are generated by the backend. To tackle the cumulative error problem of VO over time, loop back detection uses the global bag model. Making a map based on a predicted trajectory is what we call mapping. Built maps come in two flavors: metric and topological. The two forms of metric maps are sparse and dense. The most popular types of cameras used in the Visual SLAM system are monocular, binocular (stereo), depth (RGBD), and other panoramic cameras (event camera) as shown in Figure 3.

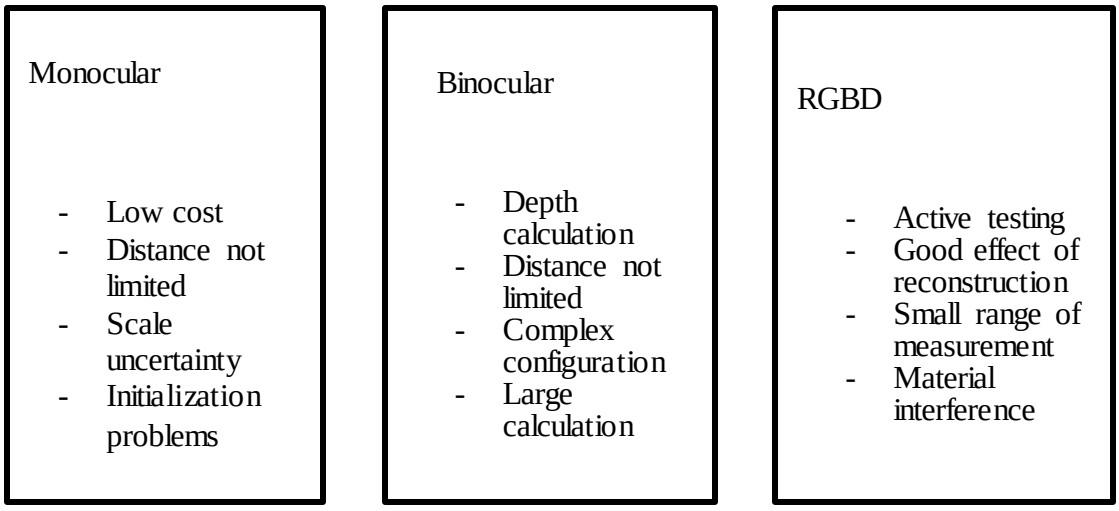

Figure 3 Comparation of three kinds of cameras in Visual SLAM [5]

3.3 FRONT END (VISUAL ODOMETRY)

There are front-end and back-end methods in software architecture [16]. The majority of frontend methods are smoothing, or based on error minimization algorithms [17]. Visual odometry (VO) is another name for a frontend that estimates motion between camera frames and marker locations. According to the observation model, they are designed to process data. Using high precision visual odometry, these methods can be utilized as a precursor to backend procedures or as a standalone procedure. Using these cameras eliminates the need for expensive and complex mathematical equipment without sacrificing quality. It is calculated in VO based on image information acquired from sensors' motion, which is converted into a matrix and solved to determine the motion state of the sensor's associated sensors' orientation and trajectory.

For example, feature points and the photometric method (also known as the direct method) are two of the most commonly used VO solutions. The feature point technique extracts the corner points and surrounding descriptors from each frame image. The invariance of the descriptors around these corner points completes the interframe matching. Finally, the camera positions and epiploic geometry are used to verify the map coordinates. Final adjustments are made to ensure that there is as little reprojection error as possible by fine tuning camera postures and maps. It is possible to obtain the camera position and map directly from the photometric error without having to extract corner points and descriptors. Direct methods cannot describe an image's overall features, as a result. In direct technique closed loop detection, the decrease of cumulative drift has not been satisfactorily solved [5].

3.4 BACK END (FILTERING AND NONLINEAR OPTIMIZATION)

By addressing trajectory and map state estimate constraints from noisy data, the backend considerably improves the collected camera pose and environment data to provide globally consistent motion and environment maps. For backend optimization in the VSLAM system, the extended Kalman filter (EKF) and the graph optimization method are two of the most commonly used techniques [5].

Bayesian estimation is used to determine the present state and confidence of the system based on the previous state and motion input. As a result, the current state of the system is estimated using the observed data and the current state. Because its storage expands exponentially with the square of the state's size, the filter-based optimization method has limited utility in large, unknown situations. Filtering methods include particle filter, extended Kalman filter, and unscented Kalman filter.

The key idea behind nonlinear optimization (graph optimization) is to convert the nonlinear optimization process into a graph, with the vertices representing pose and environment attributes at various times and the edges representing the constraint relationship between vertices [5]. Posture and environment features are solved via a post map optimization strategy that allows the state to be optimized on the vertex to better satisfy the edge's requirements. In order to get an accurate representation of the motion path and surroundings, you must run an optimization algorithm on it.

3.5 BACK LOOP DETECTION

An autonomous system's ability to recognize the current scene and determine whether or not the region has been explored is drastically reduced when it returns to its initial location, thanks to loop back detection. Consistency and reliability should be established between its trajectory and the environmental map by adjusting and/or reducing these disparities when it returns to a certain location. After the optimization technique is completed, the motion trajectory and environment map are the matching maps. It is the most frequent method of loop back detection to use the word bag model, which builds a word table comprising k words by k-mean clustering for the local features received from the image. Images are represented as k-dimensional vectors, which are utilized to determine the scene's uniqueness and identify access points based on the frequency of each word in the word table [5].

3.6 MAPPING

The map necessary to meet the task's requirements is constructed using the camera's track and image. The map can be completed after the location of the road punctuation is known. Using VSLAM, the term "environment map" refers to a collection of all the road punctuation points collected by the autonomous robotics system over the course of a given timeframe. In frontend

detection and back-end optimization, mapping is the process of determining all the possible paths to go there. Positioning services, obstacle avoidance, and environmental reconstruction are the primary roles played by VSLAM mapping. The autonomous robot system should be able to use the current map data and recover from the incorrect (or initial) location when initializing or tracking missed. Metric maps, topological maps, sparse maps, and dense maps are the most common types of maps [5].

4.0 NEW FRONTIERS: SENSORS AND LEARNING

4.1 LiDAR BASED ODOMETRY

LiDAR odometry can be estimated using a variety of techniques. One of the most widely used techniques is Implicit Moving Least Squares SLAM (IMLS-SLAM) [18], which employs a scanning and matching architecture. IMLS is used for surface reconstruction, which is believed to have a better match quality, and an algorithm to sample the 3D images. This work uses only 3D LiDAR sensors for odometry estimation, which is important to mention. LiDAR Monocular Visual Odometry (LIMO), a method proposed by [19] that combines data from both a LiDAR and a monocular camera, is also known as LIMO. Camera features are mapped to LiDAR data, which is used to determine the depth of the scene. Bundle adjustment is used to estimate motion by combining the data.

Another work by [20] provides a revolutionary method called Simultaneous Trajectory Estimation and Mapping (STEAM). Ground truth data is used to train a Gaussian process model. Predicted pose outputs are produced using the estimator and the ground truth, which is a well detected features retrieved from the point clouds. An alternative method for locating an unmanned vehicle in an off road environment, the closet probability and feature grid SLAM (CPFG-SLAM) [21], has been presented for the SLAM challenge. The point cloud features are combined with probability and grid map occupancy probabilities. The optimization function for a match between a point cloud and a grid map is built using expected maximization (EM). Data preprocessing, pose estimation, and updating the feature grid map are all part of this technique. The preprocessing of the point cloud includes filtering and classifying the data. Updating point cloud features includes extracting and updating the grid's probabilities of occurrence of each feature. Levenberg-Marquadt algorithm is then used to run the EM algorithm. It is not only unable to withstand a dynamic environment, but also fails to solve the loop closure problem, despite its great localization accuracy.

4.2 STEREO BASED ODOMETRY

The SOFT-SLAM technique [22], which is based on feature tracking, may be the most advanced stereo visual odometry algorithm currently available. There are two parallel threads used to optimize the feature-based pose graph that is created. Loop closing and global consistency are made possible thanks to the odometry and mapping threads. Using visual odometry instead of bundle adjustment, which is computationally intensive, provides better localization than using bundle adjustment. The ESDSF given by [23] and EKF on Lie Groups

(EKF) method is another stereo approach (LG-EKF) [23]. Based on ESDSF, which is developed from Extended Information Filter, this technique preserves state space geometry by employing sparse information matrix. ESDSF on a Lie Groups is one of the primary features of this technology, which not only has all the advantages of conventional ESDSF but also holds the state space geometry using Lie Groups.

Iterative two stage odometry estimation is presented by [24] in the form of an iterative 2 stage technique. In this algorithm, optical fluxes and reprojection errors induced by the 6 DOF motion are analyzed. They have demonstrated that the optical flow algorithm's reprojection error, which is dependent on the coordinates of the features, is justifiable. A detailed explanation of how the proposed SLAM technique works for mobile robot navigation in outdoor environments is depicted in Figure 4**.**

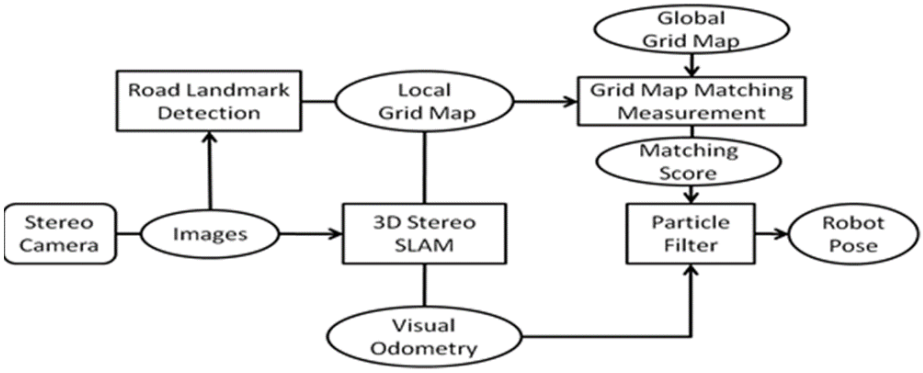

Figure 4 Depict a mobile robot's ability to use stereo SLAM in an outdoor scenario [2]

5.0 CONCLUSION

A survey of several VSLAM techniques was carried out in this work to better understand how researchers are dealing with VSLAM issues and the trends they are seeing. Researchers are interested in VSLAM because it allows for simultaneous mapping and localization. This is a huge step forward in the quest to create mobile robots that can fulfil their goals on their own, without the need for human intervention. In virtue of its inability to challenge the dominance of pose maps in back-end optimization, open-source SLAM systems based on factor map optimization will become more sophisticated in the future as the popularity of related algorithms and open-source code grows. VSLAM's existing difficulties as mentioned above must be addressed to improve this study area. Therefore, a review of recent and foundational VSLAM was undertaken to help us uncover the persistence and recent issues related with existing VSLAM methodologies.

As time has passed, a new breakthrough approach known as VSLAM has been introduced. A key benefit of VSLAM is that it allows mobile robots to carry out mapping and localization tasks at the same time, allowing them to operate more efficiently in dynamic environments. This has spawned new study areas such as events camera, deep learning, and semantic VSLAM

because of the numerous hurdles that remain in dynamic adaptability and generalization ability as well as the improved sensing and multi sensor integration capabilities. There are various faults and concerns that need to be addressed in VSLAM before it can be regarded a comprehensive solution in theory, as previously stated. In summary, VSLAM is a promising solution, but how far can the built VSLAM algorithm effectively meet the main purpose of the VSLAM technique in making the mobile robot really autonomous remains to be seen.

Acknowledgement

The research has been carried out under program Research Excellence Consortium (JPT (BPKI) 1000/016/018/25 (57)), Consortium of Robotics Technology for Search and Rescue Operations (CORTESRO) provided by Ministry of Higher Education Malaysia (MOHE). The authors also acknowledge Universiti Teknologi Malaysia (UTM), vote no (4L930) for providing facilities and support to complete this research.

REFERENCES

[1] A. Singandhupe and H. M. La (2019) A Review of SLAM Techniques and Security in Autonomous Driving. *IEEE International Conference on Robotic Computing (IRC)*, 602-607.

[2] O. Agunbiade and T. Zuva (2018) Simultaneous Localization and Mapping In Application to Autonomous Robot. *International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, 1-5.

[3] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira (2016) Past, Present, and Future of Simultaneous Localization and Mapping: Toward The Robust-Perception Age. *IEEE Transactions on Robotics*, 32, 1309-1332.

[4] A. R. M. Kamil, A. A. Faudzi, and Z. H. Ismail, (2019) Ducts Inspection Mobile Robot Using Simultaneous Localization and Mapping. *International Conference on Universal Wellbeing (ICUW)*, 204.

[5] C. Z. Sun, B. Zhang, J. K. Wang, and C. S. Zhang (2021) A Review of Visual SLAM Based on Unmanned Systems. *International Conference on Artificial Intelligence and Education (ICAIE)*, 226-234.

[6] S. Poddar, R. Kottath, and V. Karar (2019) Motion Estimation Made Easy: Evolution and Trends in Visual Odometry. *Recent Advances In Computer Vision*, Ed: Springer, 305-331.

[7] S. A. Mohamed, M.-H. Haghbayan, T. Westerlund, J. Heikkonen, H. Tenhunen, And J. Plosila (2019) A Survey on Odometry for Autonomous Navigation Systems. *IEEE Access*,7, 97466-97486.

[8] Y. D. Yasuda, L. E. G. Martins, and F. A. Cappabianco (2020) Autonomous Visual Navigation for Mobile Robots: A Systematic Literature Review. *ACM Computing Surveys (CSUR)*, 53, 1-34.

[9] Y. Alkendi, L. Seneviratne, and Y. Zweiri (2021) State of The Art In Vision-Based Localization Techniques for Autonomous Navigation Systems. *IEEE Access*, 9, 76847-76874.

[10] C. Papachristos, S. Khattak, and K. Alexis (2017) Autonomous Exploration of Visually-Degraded Environments Using Aerial Robots. *International Conference on Unmanned Aircraft Systems (ICUAS)*, 775-780.

[11] Y.-S. Shin and A. Kim (2019) Sparse Depth Enhanced Direct Thermal-Infrared SLAM Beyond The Visible Spectrum. *IEEE Robotics and Automation Letters*, 4, 2918-2925.

[12] A. U. Shamsudin, K. Ohno, T. Westfechtel, S. Takahiro, Y. Okada, and S. Tadokoro (2016) Fog Removal Using Laser Beam Penetration, Laser Intensity, and Geometrical Features for 3D Measurements In Fog-Filled Room. *Advanced Robotics*, 30, 729-743.

[13] X. Shi, D. Li, P. Zhao, Q. Tian, Y. Tian, Q. Long (2020) Are We Ready for Service Robots? The Openloris-Scene Datasets for Lifelong SLAM. *IEEE International Conference on Robotics and Automation (ICRA)*, 3139-3145.

[14] A. Merzlyakov And S. Macenski (2021) A Comparison of Modern General-Purpose Visual SLAM Approaches. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 9190-9197.

[15] F. R. Fabresse, F. Caballero, I. Maza, and A. Ollero (2016) Robust Range-Only SLAM for Unmanned Aerial Systems. *Journal of Intelligent & Robotic Systems*, 84, 297-310.

[16] G. Grisetti, R. Kümmerle, C. Stachniss, And W. Burgard (2010) A Tutorial on Graph-Based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2, 31-43.

[17] M. Kuzmin (2018) Review, Classification and Comparison of The Existing SLAM Methods for Groups of Robots," *Conference of Open Innovations Association (FRUCT)*, 115-120.

[18] J.-E. Deschaud (2018) IMLS-SLAM: Scan-To-Model Matching Based on 3D Data, *IEEE International Conference on Robotics and Automation (ICRA)*, 2480-2485.

[19] J. Graeter, A. Wilczynski, And M. Lauer (2018) Limo: Lidar-Monocular Visual Odometry. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7872-7879.

[20] T. Y. Tang, D. J. Yoon, And T. D. Barfoot (2019) A White-Noise-On-Jerk Motion Prior For Continuous-Time Trajectory Estimation on SE (3). *IEEE Robotics and Automation Letters*, 4, 594-601.

[21] K. Ji, H. Chen, H. Di, J. Gong, G. Xiong, J. Qi (2018) CPFG-SLAM: A Robust Simultaneous Localization and Mapping Based on LIDAR In Off-Road Environment. *IEEE Intelligent Vehicles Symposium (IV)*, 650-655.

[22] I. Cvišić, J. Ćesić, I. Marković, And I. Petrović (2018) SOFT - SLAM: Computationally Efficient Stereo Visual Simultaneous Localization and Mapping for Autonomous Unmanned Aerial Vehicles. *Journal of Field Robotics*, 35, 578-595.

[23] K. Lenac, J. Ćesić, I. Marković, and I. Petrović (2018) Exactly Sparse Delayed State Filter on Lie Groups for Long-Term Pose Graph SLAM. *The International Journal of Robotics Research*, 37, 585-610.

[24] M. Buczko, V. Willert, J. Schwehr, And J. Adamy (2018) Self-Validation for Automotive Visual Odometry. *IEEE Intelligent Vehicles Symposium (IV)*, 1-6.