Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

Collaborative Monocular Visual SLAM for Multi-Robot

Khandana Ali Khan[1], Shafaque Saira Malik[2]
[1]Department of Computer Science, University of Balochistan, Quetta, Pakistan. khandanaali11@gmail.com
[2]Department of Computer Science, University of Balochistan, Quetta, Pakistan. shafaque.malik@gmail.com

DOI: 10.5281/zenodo.6844405

ABSTRACT

Collaborative SLAM is an amazing extension of single robot locations where multiple robots with monocular cameras work together in a dynamic environment to build one global map. The global map is later used by the multiple moving robots to localize themselves on the map. The application of collaborative SLAM can be used in various fields that include collaborative military tasks, search and rescue, agricultural planting, multi-robots working together to improve efficiency, and many others.

Generally, every existing collaborative SLAM method uses an offline technique to process the collected data in the indoor environment. The indoor environment has limited space and lacks GPS connectivity. In this paper, we aim to give a step toward the usage of two drones equipped with monocular cameras and a standard laptop as the server for monitoring indoor workplaces. We worked on Simultaneous localization and mapping standard architecture with building the centralized global SLAM by the micro aerial vehicles such as Tello in our case. We investigated the method and localization of the drone on the global map.

Keywords: Micro aerial vehicle (MAV), Collaborative SLAM, ROS, Visual odometry.

Cite as: Khandana Ali Khan & Shafaque Saira Malik. (2022). Collaborative Monocular Visual SLAM for Multi-robot. *LC International Journal of STEM (ISSN: 2708-7123)*, 3(2), 19–30. https://doi.org/10.5281/zenodo.6844405

INTRODUCTION

SLAM has been performing outstandingly in making maps and localizing itself in the real-time environment as well as simulation. These devices can be used in a wide range of fields, including agriculture, search and rescue, environmental monitoring, surveillance, and inspection. Many various SLAM methods utilizing cameras, laser scanners, and other sensor types have been proposed by researchers. SLAM using the camera is referred to as visual SLAM. Visual SLAM became a very practical approach to solving the SLAM problem because it is based on visual information only like PTAM and ORB-SLAM. Visual SLAM methods are most commonly used for single-robot use cases.

Many robot technologies are supposed to operate together in large-scale environments to accomplish challenging tasks that can only be accomplished by a few robots because of the increased need for robotics applications to loop closer and global bundle adjustments are stable if used correctly. Multiple-agent systems can also improve map accuracy by requiring only a small area to be covered by each agent, resulting in faster mapping and a smaller amount of drift. The loop detection method can be used to quickly fix this.

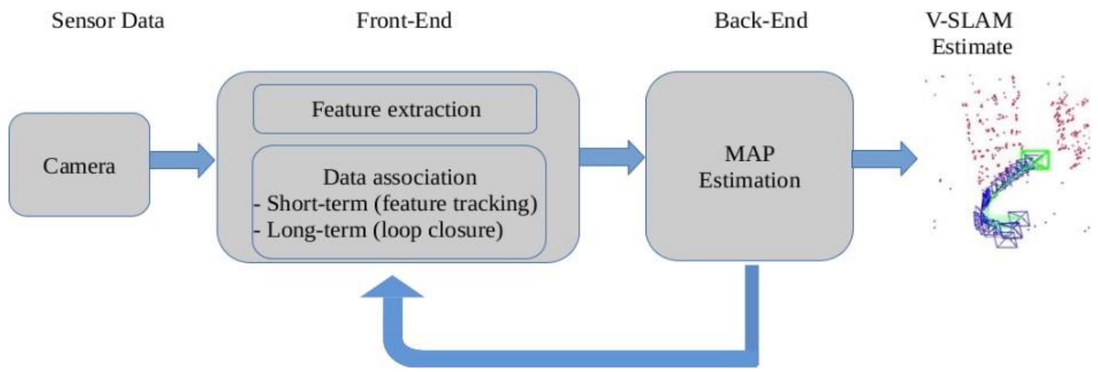


Fig. 1. Simultaneous localization and mapping standard architecture.

In a visual SLAM system, the architecture is made up of two major components: the front end and the back end. Monocular and stereo camera sensors are used to collect data at the front end. Monocular only one camera is used in visual SLAM. Two cameras are used in stereo visual SLAM. In terms of cost, computing complexity, and adaptability, each has its pros and downsides. In the front end, as depicted in Figure.1, the feature extraction phase is performed. It is also necessary to link these elements to 3D locations using map points and landmarks. A video feed must also keep track of map coordinates. By recognizing sites that have been previously encountered, long-term association lowers drift (loop detect). The back end then uses the observation data from the front end to solve optimization problems or state estimation issues, which involve estimating parameters that characterize the location of a landmark in the environment or the location of the robot inside it. Tracked objects, their locations, and relations, and the camera position in the world are estimated using SLAM.

Due to the lack of GPS, these indoor workplaces are physically constrained and there is difficulty with localization. In our system, we move forward with the use of drones for indoor workplaces. Look into the placement of drones in workplaces. Using a small drone equipped with a monocular camera. Office, class, corridor, and hall monitoring. ORB-SLAM2 was discovered to be the optimal method for these workplaces.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

The remainder of the paper is structured as follows. The purpose of the paper is introduced in section I. Literature review in section II. Methodology in section III. The experiments and results are described in section IV. Finally, Conclusions are presented in section V.

LITERATURE REVIEW

Durrant et al. [1] defined SLAM as a procedure by which a drone builds a continuous map of its surrounding and at the same time localize itself on the map. Although the technique can be done in real-time, the outcome is often improved with post-processing.
Davison et al. [2] presented a Bayesian framework that processes image information of single-hand waved camera with real-time localization via mapping a sparse set of natural features. The motion modeling technique is used for mapping distinct features. Which are used to estimate the camera motion.
They additionally [3] presented the first feature-based monocular visual SLAM which is called Monoslam. Real-time monocular SLAM solution. The algorithm creates a sparse map of landmarks within a probabilistic framework. It has been improved to make localization easier.
T. J.chong et al. [4] focus on different sensor information in a variety of environments to solve the SLAM problem. There are many types of sensors such as sonar sensors, laser sensors, visual sensors (monocular and stereo-based vision sensors), and RGB-D sensors.
Mur-Artal et al. [5] In this study author presented ORB-SLAM which is a feature-based SLAM framework by equipping the robot with a camera that works in real-time in small and large scale, indoor and outdoor areas that applies the same functionality for all SLAM operation, include tracking, mapping, re-localization, and loop detection. The map selection of the point and keyframes for the reconstruction leads to high robustness and track-able map that only develops when the scene content changes. The system efficiency on the inside is less than 1cm and in large outside situations, it's a few meters.
Additionally, Mur-Artal et al. [6] ORB-SLAM2 is based on our monocular feature-based ORB-SLAM for stereo and RGB-D cameras. This system works with the same ORB-SLAM features for mapping, tracking, and place recognition tasks. The system is comprised of three main threads running in parallel. A movement BA is used to track the camera and locate it with each frame by seeking feature matches to the local map and minimizing the inaccuracy in the re-projection of the camera location. In addition to performing local BA, the management and optimization of the local map are both accomplished through local mapping.

Collaborative Visual SLAM

Zou et al. [7] demonstrated that a collaborative VSLAM can manage dynamic environments with a multi-camera. These multiple hand-held cameras can function independently on different portions. To produce a global map, all cameras work together.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

A multi-MAV SLAM construct built a motion system source is demonstrated by Forster et al. [8] A keyframe builds a visual odometry system where each onboard agent feeds a fresh keyframe to the main server. If a server interaction between two maps is detected, they are integrated and optimized globally. While this is likely the first solution to deal with multi-MAV configuration, it doesn't give the agent any information, Allowing that to benefit from the optimum map and posture analysis.

Schmuck et al. [9] have shown a centralized framework for collaborative monocular SLAM small multi-UAVs to take up the role of agents. The agents onboard memory is limited, an agent works independently, in parallel. The central server gathers all of the information from agents and merged the maps of all agents. Agents can work on updating information by conveying merged and optimized information back to them, resulting in better and more consistent estimation.

Li et al. [10] The collaborative ORB-SLAM, a team of robots working together to find a new environment, was shown by the author. Robots are allocated to explore different elements of the organization and build local maps in the CORB-SLAM architecture.

Vemet al. [11] for a team of micro autonomous drones with forward-facing monocular-camera, the author showed and accessed a collaborative localization framework for the group using Microsoft air-sim software. In lieu of actual UAVs, the simulation will be used for all testing.

Schmuck et al. [12] present centralized collaborative SLAM architecture for autonomous agents with a monocular camera, communication units, and mini processing board. In this system, each agent running onboard visual odometry CCM-SLAM maintains their independence as individuals, although a central server with significantly better computational power allows their collaboration by gathering all of their experience, merging and optimizing their map, and conveying information back to them. An in-depth examination of benchmarking dataset focuses of the scalably and robustness of CCM-SLAM in case of data loss and transmission delay that is usual in a real mission.

Liu et al. [13] the author presents a collaborative monocular SLAM that can be used on a variety of IOS smartphones. All thanks to a centralized design. The agent is capable of exploring the world on its own, performing visual-inertial-odometry online, and then transferring all measurement data to a central server with additional processing capacity. When necessary, the server communicates with the agents and keeps a record of all of the data. A variety of datasets from Euroc and real-world scenarios. In comparison to VINS- Mono, the proposed system's mapping and integration accuracy are comparable.

Decentralized Visual SLAM

Jimenez et al. [14] The proposed model showed that it could handle decentralized navigation tasks, boosting the system's autonomy. This approach also demonstrated the model's versatility

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

when integrated into a framework, allowing the MARS system to be designed with agents that were not available at system startups, such as wireless agents, monitoring agents, or other robotic agents.

Cieslewski et al. [15] Proposed an efficient framework for decentralized SLAM based on decentralized place recognition and optimization algorithms. It does not rely on communication with a central ground station design to reduce the bandwidth used by each robot in visual SLAM based on real-world data but in simulation.

Yu et al. [16] The ORB-SLAM algorithm was executed on a CPU. DSLAM is used in multi-robot applications where agents can communicate environment information and locations.

Duboiset al. [17] based on rigid, condensed, and pruned visual-inertial packets, authors offer three approaches for sharing visual-inertial information. They also suggest a shared collaborative decentralized SLAM architecture to manage the calculation, exchange, and integration of such packets and testing those methods on the EuRoC dataset as well as our own Air-Museum dataset. The experiment revealed that the suggested method enables agents to create, trade, and integrate consistent visual-inertial packets as well as enhance trajectory prediction accuracy by several centimeters.

Distributed Visual SLAM

Egodagamage et al. [18] In this research, a distributed computing framework is needed because agents are included in the construction of such maps. An individual agent may create a map of its own local area. This may be used to create a map of a bigger region once merged. A system that produces a semi-dense global map of the environment using numerous monocular agents with unknown relative starting positions also uses an appearance-based technique to detect map overlaps.

Chen et al. [19] The author presents a multi-agent distributed monocular SLAM based on the efficient map of a large-scale environment with several robots working together. It is suggested that the monocular multiple-agent SLAM may be achieved without any previous information or big map overlaps by using a relative posture computation and map merging approach.

Zhang et al. [20] In this framework for a team of robots to map the large-scale environment with great efficiency, an efficient distributed SLAM is developed that relies on robust monocular SLAM techniques. Each user share all keyframe and accompanying feature analyses with all other users, and all data from all devices are processed to build a single pose graph showing the complete system on all devices.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

METHODOLOGY

The proposed design system's various functional components shown in Fig. 2 are discussed in detail below.

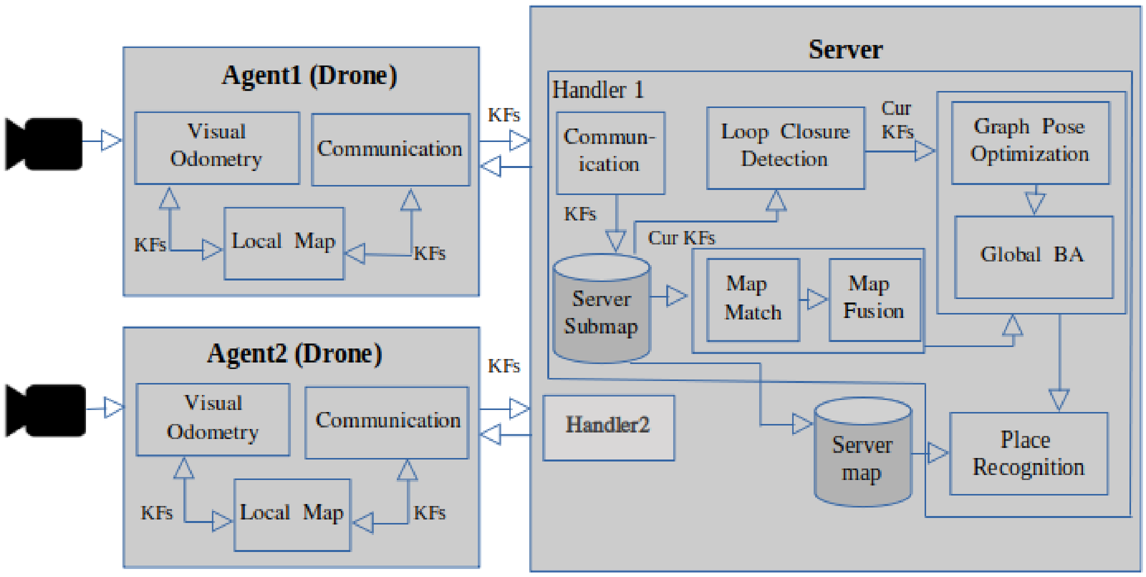


Fig. 2. Overview of Collaborative Visual SLAM Framework

Key-frame based Visual-Odometry

It is possible to use visual-odometry (VO) predicated on keyframe as the front end for an agent to analyze new frames. Each agent has three activities running in parallel: tracking, mapping, and communication. A tracking thread predicts the location of sequence landmarks on the agents map by using the cameras frame-to-frame rotation. Local bundle adjustment is carried out by the mapping thread using tracking data. Create a local map, starting from the first keyframes map point. As fresh keyframe information is received, the local map is updated.
Bundle adjustment can improve mapping by eliminating keyframe and map feature re-projection mistakes. KF and MP updates are sent to the server via an agents communication module, which is responsible for delivering them to the server. They can converse with each other in either direction. Multi-client applications connect to the server using the communication module.

Agent Handler

Every one of the agents does have its own handler on the server-side. Using an agent handler, a server and agent can communicate. Additionally, this handler includes a serial communication that controls data exchange with the agent and detects a loop in their map on the server-side. A separate thread is used by each agent handler to perform loop closure detecting and communication. This allows many agents handlers to work independently and concurrently. The handler is in charge of connecting all of the server components together. Sim(3)-transformation data is kept in the agents handler and used to convert information stored locally on the agent to a corresponding location on the server.

Communication

The communication module is used by multi-client to communicate with the server. Creates a server sub-map for each agent, inserts it into the server map, and assigns a server sub-map manager to it. The sub-maps manager can detect a loop inside a the agents and the server can communicate in both directions. The communication module is used by multi-client to communicate with the server. Sub map find similar areas across several sub-maps, and perform global optimization within or across sub-maps.

Map fusion and map matching

The map fusion module mixes two identical KFs from two distinct maps. All server maps are replaced with one that contains information from both of the relevant agents server maps, and then an optimization algorithm process is carried out. When the maps of agents 0 and 1 are combined, a dedicated server map is created that includes the combined maps of the two agents, and the original server maps are deleted. It is now possible to add new information from both agents to the server map which has been produced.

Optimization

To engage global optimization, you must close a loop. Server Maps pose graphs are optimized using bundle adjustment in our stack of Maps. In order to increase the efficiency of a map, BA reduces the re-projection error including all keyframes and map points that are considered throughout the optimization process. Afterward, we undertake an optimal solution for the posture graph to optimize the graph and minimize scale drift, taking into account all the information on the map. To save time and resources, global BA is only conducted on a server.

Place Recognition

The place recognition method involves the recognition of a previously visited location. This method uses feature matching to find similar scenes and calculate relative poses.

EXPERIMENTS AND RESULTS

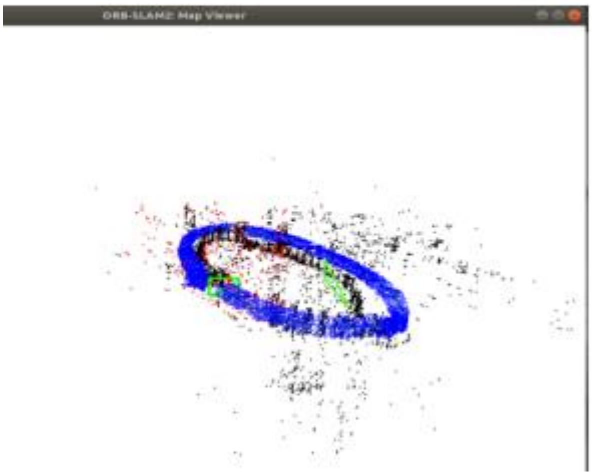
The system collaborative monocular SLAM infrastructure is analyze on two different scenarios. In experiment I, the DJI Tello was the small UAV used in the experiment and data collection. Although In experiment II, we apply EUROC dataset. As a result, ORB-SLAM2 is used for estimation camera position and location recognition. A ground station was assigned

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

to PC. Both experiments run ROS melodic under Ubuntu 18.04 , 4-bit system, Intel(R) Core i5-2320 CPU@3.00GHz, 500G hard drive, 16G RAM and Nvidia GeForce GTX 480 graphics card as a server to test our method.

In experiment I, In a real environment two small UAVs flying with lower computation capabilities. Navigate across an indoor environment, each with a single front-facing camera.

(a)



(b)

Fig. 3. (a) In real-time indoor corridor area feature extraction. (b) Point cloud and trajectory.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

We used the university library as the indoor environment and flown both agents in that library. Which each of the sequences represents a specific area within a relatively small space. Data from all small UAVs is collected at the same time and processed online. Collaborative VSLAM drone lives steaming in the library. Where agent 1 fly on the second floor and agent 2 fly on the first floor. In Fig. 3 The green squares in the picture show key features. Agent 1 shows (blue) and agent 2 shows (black), showing the agents keyframes and map points.
In experiment II, We use the EuRoC dataset, which is publicly available, as well as our own dataset, with the main essential properties described in Table 1. Both give precise ground station position data. The EuRoC dataset contains video sequences. Flight duration is 5:32 minutes, with a total trajectory length of 150 meters. This is a well-textured, industrial area setting. Repetitive, slow visits to the same site are encouraged. see Fig. 4

TABLE I
DATA-SETS UTILIZED IN THE MULTI - ROBOT EVALUATION

Dataset	Trajectory Length	Flight Time	Camera View	Environment	Note
MH 01	80	3 min	Front	Industrial , Indoor	good quality, bright scene
MH 02	70	2:30 min	Front	Industrial , Indoor	good quality, bright scene
CSD 01	58	1:44 min	Front	Educational , Indoor	good quality
CSD 02	36	1:12 min	Front	Educational , Indoor	good quality

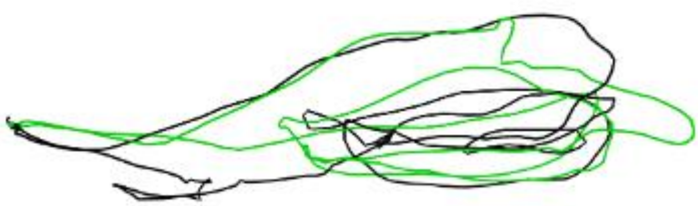


Fig. 4. Euroc dataset Collaborative trajectory estimation for MH1 and 2 (view color different agents).

Our dataset was utilized to fly a mini UAV over an inside space with a front-facing camera. The sequences from this dataset described also in Table 1 are used here. Each sequence is

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

executed in parallel while communicates with the server. Experiment were conduct in surround by corridor on all four sides having two glass doors, the roof top is elevated covered with green fiber sheet. Our method has the lowest error in small and medium indoor scenarios.

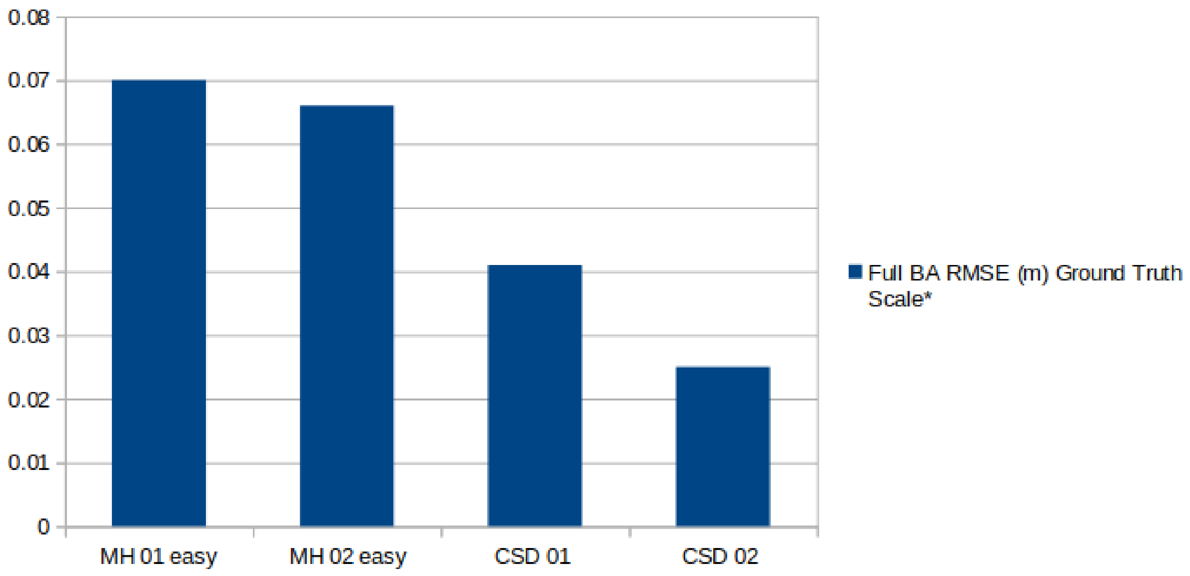


Fig. 5. Data set trajectory comparison Error

Our method has the lowest error in an indoor scenarios. We go through each environment's sessions one at a time. A global alignment may be used to compute RMSE for all trajectories in the same environment, as each has the same ground truth of the same world reference.

CONCLUSION

In terms of inaccuracy and mapping time, we showed that employing a collaborative SLAM technique has advantages over single-agent SLAM strategies. We employed Collaborative SLAM for a central server and a large number of agents, such as small robots with monocular cameras and processing units. Indoor surroundings are constrained in terms of physical space, and because to the lack of GPS, there is an issue with localization. As a result, the goal of this study is to pave the way for the use of drones in the indoor environment. To look into the placement of drones in this workspace.

REFERENCES

[1] Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping: part I. *IEEE robotics & automation magazine*, *13*(2), 99-110.
[2] Davison, A. J. (2003, October). Real-time simultaneous localisation and mapping with a single camera. In *Computer Vision, IEEE International Conference on* (Vol. 3, pp. 1403-1403). IEEE Computer Society.
[3] Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. (2007). MonoSLAM: Real-time single camera SLAM. *IEEE transactions on pattern analysis and machine intelligence*, *29*(6), 1052-1067.
[4] Chong, T. J., Tang, X. J., Leng, C. H., Yogeswaran, M., Ng, O. E., & Chong, Y. Z. (2015). Sensor technologies and simultaneous localization and mapping (SLAM). *Procedia Computer Science*, *76*, 174-179.
[5] Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, *31*(5), 1147-1163.
[6] 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.
[7] Zou, D., & Tan, P. (2012). Coslam: Collaborative visual slam in dynamic environments. *IEEE transactions on pattern analysis and machine intelligence*, *35*(2), 354-366.
[8] Forster, C., Lynen, S., Kneip, L., & Scaramuzza, D. (2013, November). Collaborative monocular slam with multiple micro aerial vehicles. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3962-3970). IEEE.
[9] Schmuck, P., & Chli, M. (2017, May). Multi-uav collaborative monocular slam. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3863-3870). IEEE.
[10] Li, F., Yang, S., Yi, X., & Yang, X. (2017, December). Corb-slam: a collaborative visual slam system for multiple robots. In *International Conference on Collaborative Computing: Networking, Applications and Worksharing* (pp. 480-490). Springer, Cham.
[11] Vemprala, S., & Saripalli, S. (2018, June). Monocular vision based collaborative localization for micro aerial vehicle swarms. In *2018 International Conference on Unmanned Aircraft Systems (ICUAS)* (pp. 315-323). IEEE.
[12] Schmuck, P., & Chli, M. (2019). CCM-SLAM: Robust and efficient centralized collaborative monocular simultaneous localization and mapping for robotic teams. *Journal of Field Robotics*, *36*(4), 763-781.
[13] Liu, J., Liu, R., Chen, K., Zhang, J., & Guo, D. (2021, May). Collaborative Visual Inertial SLAM for Multiple Smart Phones. In *2021 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 11553-11559). IEEE.
[14] Jiménez, A. C., García-Díaz, V., González-Crespo, R., & Bolaños, S. (2018). Decentralized online simultaneous localization and mapping for multi-agent systems. *Sensors*, *18*(8), 2612.
[15] Cieslewski, T., Choudhary, S., & Scaramuzza, D. (2018, May). Data-efficient decentralized visual SLAM. In *2018 IEEE international conference on robotics and automation (ICRA)* (pp. 2466-2473). IEEE.
[16] Yu, J., Gao, F., Cao, J., Yu, C., Zhang, Z., Huang, Z., ... & Yang, H. (2020, May). CNN-based Monocular Decentralized SLAM on embedded FPGA. In *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)* (pp. 66-73). IEEE.
[17] Dubois, R., Eudes, A., & Frémont, V. (2022). Sharing visual-inertial data for collaborative decentralized simultaneous localization and mapping. *Robotics and Autonomous Systems*, *148*, 103933.
[18] Egodagamage, R., & Tuceryan, M. (2017). Distributed monocular SLAM for indoor map building. *Journal of Sensors*, *2017*.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-02 | June-2022

[19] Chen, X., Lu, H., Xiao, J., & Zhang, H. (2018, July). Distributed monocular multi-robot slam. In *2018 IEEE 8th annual international conference on CYBER technology in automation, control, and intelligent systems (CYBER)* (pp. 73-78). IEEE.
[20] Zhang, H., Chen, X., Lu, H., & Xiao, J. (2018). Distributed and collaborative monocular simultaneous localization and mapping for multi-robot systems in large-scale environments. *International Journal of Advanced Robotic Systems*, *15*(3), 1729881418780178.