International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

Recent Advancements in Visual-Inertial Simultaneous Localization and Mapping (VIO-SLAM) for Autonomous Vehicles: A Review

Md. Syamul Bashar, A K M Ashikuzzaman, Mostafa Rafid, Saad Been Mosharof
Mechanical Engineering Department
Shahjalal University of Science and Technology
Sylhet, Bangladesh

***Abstract*— VIO-SLAM is a complex robotics challenge that includes calculating a vehicle's trajectory while also mapping the surroundings using a combination of visual and inertial sensors. VIO-SLAM has drawn a lot of attention recently because of the potential uses it could have in a number of industries, including mapping, surveillance, and search and rescue. In this work, current developments in VIO-SLAM for autonomous aerial vehicles are reviewed. The review covers the latest research and developments in VIO-SLAM, including optimization algorithms, and deep learning-based approaches used in VIO-SLAM systems. The paper also discusses the challenges and limitations of VIO-SLAM and concludes by highlighting the potential future directions of research in VIO-SLAM, including the integration of other sensors such as lidar and radar, real-time implementation, and robustness to various environmental conditions. The article offers a thorough overview of recent developments in VIO-SLAM and their prospective use in autonomous aerial vehicles. It also outlines the present difficulties and potential future research initiatives in this area.**

Keywords—VIO-SLAM; Autonomous Vehicles;

I. INTRODUCTION

One of the most important technologies for autonomous systems, such as drones, self-driving automobiles, and robotics, is VIO-SLAM. VIO-SLAM makes use of both inertial measurements from Inertial Measurement Units (IMUs) and visual data from cameras to estimate the system's state, including the position and orientation of the camera or robot in relation to its surroundings.

The accuracy and robustness of the calculated state have significantly improved as a result of recent developments in VIO-SLAM. This is accomplished by better integrating visual and inertial measurements using sophisticated optimization approaches, such as non-linear optimization and machine learning. Furthermore, the performance of VIO-SLAM has been enhanced by the use of cutting-edge sensors like event cameras and stereo cameras.

Highlighting recent developments in VIO-SLAM for autonomous systems is the goal of this paper. We will go over the various algorithms, sensors, and methods created to increase the precision and reliability of the estimated state. We will also go through the drawbacks and restrictions of VIO-SLAM as well as possible future lines of inquiry.

II. AN OVERVIEW OF DIFFERENT SLAM ALGORITHMS

A crucial issue in robotics and autonomous systems is simultaneous localization and mapping (SLAM). It involves mapping the environment while simultaneously estimating the position of a robot or camera. In the literature, a variety of SLAM algorithms have been presented, each with unique advantages and disadvantages.

Visual Slam: Visual SLAM is a technique that uses visual information from cameras to map the environment and estimate the location and orientation of a robot or camera. This technology is especially helpful in circumstances when other sensors, like GPS or LiDAR, might not be readily available or trustworthy. The majority of the time, visual SLAM algorithms track elements across many camera frames and estimate their 3D coordinates in relation to the camera. The ORB-SLAM, PTAM, and LSD-SLAM visual SLAM algorithms are a few of the well-liked ones.

VIO SLAM: In order to estimate the location and orientation of a robot or camera and create a map of the environment, VIO SLAM combines visual data from cameras with inertial measurements from IMUs. In comparison to using either modality alone, VIO-SLAM algorithms can increase the precision and robustness of the estimated state by combining visual and inertial data. In order to combine visual and inertial measurements and predict the system's state, advanced optimization techniques like non-linear optimization and machine learning are frequently used in VIO-SLAM algorithms. OKVIS, ROVIO, and VINS-Mono are a few well-known VIO-SLAM algorithms.

OKVIS: A monocular camera and an inertial measuring unit (IMU) are used in the cutting-edge Visual-Inertial SLAM system known as OKVIS to determine the location and orientation of a moving platform, such as a drone or a robot. Using a powerful feature detector and descriptor, the system initially finds and tracks feature points in the image sequence. The mobility of the platform over time is then estimated using the feature tracks and the IMU measurements. In order to jointly optimize the estimated trajectory and the 3D locations of the feature points in the environment, OKVIS employs an optimization-based technique. In order to rectify any drift in the predicted trajectory brought on by cumulative inaccuracy over time, the system additionally includes loop-closure detection.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

ROVIO: Using a monocular camera and an IMU, the ROVIO Visual-Inertial SLAM system establishes the position and orientation of a moving platform, such as a robot or a drone. The system uses a powerful feature detector and descriptor to initially identify and track feature points in the image sequence. The platform's mobility between image frames is then predicted using the IMU readings, allowing the feature points to be tracked over time. The camera posture, IMU biases, and 3D locations of the feature points in the environment are all collaboratively estimated by ROVIO using an optimization-based approach. Additionally, loop-closure detection is incorporated into the system to correct any drift in the estimated trajectory brought on by accumulated error over time.

VINS-mono: Robotics and autonomous vehicles use the visual-inertial state estimation algorithm known as VINS-Mono. The program can determine the position and orientation of the vehicle with the use of just a single monocular camera and an IMU. The system tracks the vehicle's pose over time by employing a sliding window technique. The camera's motion between frames and the properties of the environment that the camera is viewing are estimated during the slam process. To produce a precise estimate of the vehicle's position and orientation, the algorithm merges the information from the camera and IMU. When GPS is unavailable or unreliable, such as when GPS signals are obstructed or when the environment is indoors, VINS-Mono is especially helpful.

III. LITERATURE REVIEW

The capacity of visual-inertial simultaneous localization and mapping (SLAM) algorithms to precisely predict the posture and position of a robot in real time has garnered a lot of interest recently. In general, the goal of recent research in this area is to create efficient, reliable algorithms that can operate in difficult situations and increase the precision of SLAM estimates. Accurately fusing the data from the visual and inertial sensors is one of the main issues in visual-inertial SLAM. Researchers are investigating various approaches, including deep learning, graph optimization, and probabilistic filtering, to overcome this difficulty. Additionally, work is being done to create portable, low-power hardware that can be applied to visual-inertial SLAM tasks in robots, and drones.

TABLE I: RECENT ADVANCEMENTS IN VIO-SLAM ALGORITHMS: A SUMMARY OF KEY WORKS

Ref.	*Objectives/ Purpose*	*Method/ Approach*	*Key Findings*
[1]	The monocular visual-inertial state estimator VINS-Mono, which is reliable and adaptable, is proposed in this study. The goal is to deliver a full and dependable system that can be used for a variety of localization-intensive applications.	To produce very precise visual-inertial odometry, the strategy begins with a reliable initialization procedure and then uses a tightly coupled, nonlinear optimization-based method. Relocalization is possible with minimal compute thanks to a loop detection module. The suggested approach can reuse a map by efficiently saving and loading it. To ensure global consistency, 4-DOF pose graph optimization is also carried out.	The suggested system is examined against other cutting-edge algorithms and validated using open datasets and practical tests. The system performs better than other open-source implementations. Future research areas include developing dense maps based on monocular VINS results, online techniques to evaluate the observability characteristics of monocular VINS, and online calibration of practically all sensor intrinsic and extrinsic parameters.
[2]	This study examines the accuracy, latency, and processing demands of publicly accessible visual-inertial odometry (VIO) techniques for a flying robot with onboard state estimation. Which VIO algorithms work well under these restrictions? is the research question. What is the right balance among precision, response time, and computational power for a flying robot?	Six VIO pipelines (MSCKF, OKVIS, ROVIO, VINS-Mono, SVO+MSF, and SVO+GTSAM) are examined by the authors on various hardware setups, including a number of single-board computers frequently used in flying robots. The EuRoC datasets, which feature 6DoF trajectories characteristic of flying robots, are processed for the evaluation, which takes into account the pose estimation accuracy, per-frame processing time, and CPU and memory load.	With more computing, VIO algorithms' accuracy and resilience can be increased, but for systems with constrained resources, striking the correct balance between competing needs can be difficult. The findings can aid researchers in selecting suitable trade-offs for their flying robot systems.
[3]	In this study, the system architecture for a swarm of quadrotors that uses monocular visual-inertial odometry for all onboard estimation is presented. These are the research queries: Can a group of quadrotors work together without the use of outside localization systems? Is the system expandable to accommodate more robots? How well does the system function both inside and outside environments?	The system architecture, estimation, planning, and control for the multi-robot system are presented by the authors. They don't rely on an external motion capture device or GPS, instead performing all estimations using monocular visual inertial odometry onboard. Up to 12 quadrotors are used to evaluate the approach's robustness and scalability in both indoor and outdoor settings.	The authors showed what they claim to be the biggest quadrotor swarm that doesn't rely on motion capture or GPS data. The system may be deployed just about any place because it is reliable in a wide range of indoor and outdoor conditions. The system is scaleable to bigger robot swarms and only minimally depends on wireless communication. Future research will concentrate on swarm size expansion, complex formation behaviors, loop closure detection, and shared global feature mapping.
[4]	In this research, a unique closely linked monocular visual-inertial SLAM system is presented that, when run at high frame rates on a conventional CPU, delivers reliable and accurate motion tracking. The goal is to increase the EKF VIO estimator's motion tracking accuracy while ensuring quick response to highly dynamic robot motion.	The research builds a globally consistent map by tracking motion with a visual-inertial extended Kalman filter (EKF).	With a per-frame processing time close to that of the filter-based technique, the proposed algorithm offers motion tracking that is accurate and comparable to optimization-based methods. The tests validate the suggested algorithm's advantage in providing precise and reliable motion tracking.
[5]	The goal of this research is to create a deep	The study presents an algorithmic and	The developed VIO algorithm is appropriate for

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

	learning-trained visual-inertial odometry system. The issues of rigorous calibration, computing limitations, and system accuracy improvement are all addressed in the research questions.	computational architecture that calculates the velocity between two frames using deep learning methods. Stereoscopic image cameras and an affordable inertial system are used to create the system. The method shows how training a neural network based on optical flow can avoid the procedure of estimating optical and scaling parameters.	systems with light computational requirements and real-time processing. According to the study, LSTM outperforms KF-based drift correction in terms of performance. Future applications of deep neural networks for system identification are also noted. According to the baseline experiments, an unaided visual inertial system without any external helping sensors would have mean velocity errors of around 1.7 m/s.
[6]	This study presents a wheel-speed anomaly detection wheel-speed estimation algorithm for monocular inertial simultaneous localization and mapping (SLAM). The least squares problem in the algorithm incorporates wheel speed data and uses a nonlinear optimization technique to predict the ideal state. The study focuses on how speed measurement errors affect the accuracy of posture prediction, and it suggests using a torque-based Mecanum wheel chassis-control approach.	In this paper, a torque control-based control algorithm for a Mecanum wheel-moving chassis is presented. In addition to the robot's pose, speed, visual feature point depth, and IMU zero bias, the system also calculates other variables. It detects aberrant movement of the Mecanum wheel using sensor data from monocular cameras, an IMU, and mobile chassis speed measurements. In order to increase positioning accuracy as the robot moves on the ground plane, the study also adds a plane constraint element.	The proposed algorithm successfully detects abnormal chassis movement and isolates wrong wheel speed measurements during slip processes, enabling accurate pose estimation and robustness in cases of severe wheel slip due to collisions. The proposed algorithm achieved a low cumulative position error rate of 0.28% after the robot walked 812 meters, which is lower than the VINS-Mono algorithm's 4.09%, and the error rate further decreased to 0.04% after enabling the loopback-detection function.
[7]	The goal of the work is to create a semi-direct SLAM system with loop closure detection that can run quickly and attain accuracy on par with feature-based techniques. The approach monitors non-keyframes directly without the need for feature extraction and matching by using Oriented FAST and Rotated BRIEF characteristics in a keyframe.	The system monitors non-keyframes directly, without extracting and matching features, utilizing ORB features, which are multiscale FAST corners with a 256-bit descriptor. It includes a tracking thread for camera position localization, keyframe insertion, local mapping with local BA, loop closing for drift error correction through loop detection, and drift error correction through loop detection.	The SVL combines direct and feature-based techniques to address long-term navigation drift and provide real-time performance with precision on par with cutting-edge techniques. Long-term navigation and localization on light platforms are made possible by the inclusion of optical and inertial fusion in SVL-VI. This provides the groundwork for multi-sensor platform applications.
[8]	The paper presents a method for visual-inertial localization with a prior LiDAR map. It combines sparse visual feature measurements with NDT-registered visual semidense map to provide real-time posture estimates. The method uses a previous map of a different sensory modality to bound localization errors, holding potential for use in lightweight platforms.	The proposed approach tightly fuses global registrations of visual semi-dense clouds with a prior LiDAR map and sparse visual feature measurements to correct drift, addressing a cross-modality constraint between visual and LiDAR pointclouds. The method is validated through Monte Carlo simulations and field tests. The visual processing module employs stereo pairs for semi-dense visual to LiDAR map registration and sparse feature tracking.	The proposed approach achieved better orientation and position estimates compared to the conventional MSCKF, even in situations where preceding maps had low quality. The method was robust to prior maps of lower quality and outperformed VINSMono with loop closures. The proposed lightweight global localization system operates at a lower frequency with only a secondary thread for semi-dense reconstruction and NDT, making it a promising technology for real-world uses like autonomous driving.
[9]	This study investigates the use of a tightly coupled stereo camera and IMU system for accurately estimating the position of mobile robots in indoor environments without occlusions.	The forward-backward optical flow method is suggested, and STCM, a feature management technique, is also introduced. The research also introduces a visual-inertial SLAM approach based on STCM and a one-circle feature matching method to enhance pose estimation accuracy. Image tiling is used to extract features, with the image divided into 25×25 image blocks and a FAST feature extracted for each block, followed by optical flow tracking between frames.	The STCM-SLAM system is found to have the best performance in terms of relative pose error, with the lowest values for mean error, median error, minimum value, RMSE, and STD across 11 datasets. Additionally, it runs at the highest frequency, averaging 36.070 Hz. The system achieves accurate trajectory estimation for mobile robots in indoor settings and outperforms other SLAM systems in terms of RMSE, mean error, and STD.
[10]	The research introduces SD-VIS, a novel semi-direct visual-inertial SLAM framework that offers both speed and accuracy by combining feature-based and direct methods. It uses direct method for non-keyframe tracking, ensuring faster computation without loss of accuracy. Additionally, it can detect loop closures and resolve the problem of drift in long-term operation, which is not possible with the direct method.	IMU measurements are pre-integrated and used as constraints between successive images when SD-VIS is employed, and it employs sensor data from a monocular camera and IMU. For back-end optimization and loop closure detection, non-keyframes are tracked using the direct technique, while keyframes are tracked using the feature-based method.	Due to the outstanding effectiveness of the direct technique, SD-VIS shows higher accuracy than VINS-Mono and VINS-Fusion while moving quickly in low-texture settings. Further accuracy improvements result from the algorithm's keyframe selection strategy, which generates more keyframes during fast motion. In contrast to direct techniques, SD-VIS can detect loop closures and addresses the long-term drift problem. Furthermore, feature matching and descriptor computation are not required when using the KLT sparse optical flow algorithm to track keyframes.
[11]	In this paper, a new set of data for multi-robot stereo-visual and inertial SLAM is introduced, and a benchmark for assessing the performance of a single robot in coordinated settings is included.	The dataset was generated by deploying ground and aerial robots to record five indoor multi-robot situations in a former French Air Museum. The illustrations were created to highlight particular advantages and disadvantages of collaborative SLAM. Ground-truth trajectories were generated using Structure-from-Motion algorithms and fixed	Modern monocular, stereo, and visual-inertial SLAM algorithms were compared against one another on the dataset in order to establish a baseline for future single-robot performance enhancements in collaborative frameworks. The authors intend to add measurements obtained by the Intel RealsenseTM D435i sensor to the dataset, which is currently made available to the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

		AprilTag markers positioned as beacons. The robots had mounted AprilTag markers for explicit direct contacts.	public.
[12]	To establish a strong basis for metric-semantic SLAM and perception research by offering an open-source toolkit for real-time metric-semantic visual-inertial SLAM that supports mesh reconstruction and semantic labeling in 3D.	A VIO module, a reliable pose graph optimizer, a lightweight 3D mesher module, as well as a 3D metric-semantic reconstruction module makeup Kimera's four main parts. The modules' flexibility in execution enables running them singly or in tandem.	Kimera creates 3D metric-semantic mesh from semantically annotated photos instantaneously. It offers cutting-edge applications of mesh reconstruction, VIO, robust pose graph optimization, and 3D semantic labeling. Kimera wants to give academics from many communities an easy-to-use infrastructure.
[13]	In order to enhance SLAM's performance, the goal of this work is to incorporate object-level semantic information into the system.	The suggested technique, called OrcVIO, combines VIO with tracking and optimization over structured object models. OrcVIO performs batch optimization over the pose and shape of objects by differentiating based on semantic features and bounding box reprojection errors.	OrcVIO is tested with actual data, demonstrating its capacity for precise trajectory estimate and extensive object-level mapping. Future work by the authors will concentrate on object-level data association for loop closure, multiple object categories, and more general models of object shape. They also intend to expand the object categories that will be mapped.

IV. CONCLUSION

In conclusion, research into Visual-Inertial Simultaneous Localization and Mapping (VIO-SLAM) has shown promise for use in autonomous aircraft. The challenge of state estimation can be effectively solved by combining visual and inertial sensors, enabling precise localization and mapping even in situations where GPS signals are nonexistent or inaccurate. An overview of recent developments in VIO-SLAM, including several strategies for optimization, and mapping, has been provided in this paper. We've also given some insight into the fundamental problems still facing this sector. The trade-off between accuracy and computing economy is one of the main issues with VIO-SLAM, as was previously mentioned. Even though there has been a lot of development in creating computationally efficient algorithms, more study is still required. To allow fair comparisons across various algorithms, consistent datasets, and evaluation metrics are also required. Despite these difficulties, VIO-SLAM has already demonstrated considerable promise in a variety of applications, including autonomous navigation in GPS-denied environments and 3D mapping of both indoor and outdoor environments. VIO-SLAM is anticipated to play a bigger role in a variety of airborne applications, such as search and rescue, environmental monitoring, and infrastructure inspection, as technology progresses.

Research on the visual-inertial slamming algorithm is likely to go in the direction of enhancing performance in difficult situations including low light, dynamic scenes, and GPS-denied locations. Additionally, researchers may concentrate on creating stronger, more effective algorithms that can be used in real-time applications like autonomous vehicles and drones. Exploring the application of machine learning approaches to enhance the accuracy and dependability of visual-inertial slam algorithms is another potential research area. Additionally, the incorporation of other sensor modalities like radar and LiDAR may improve the performance of visual-inertial slam algorithms in difficult terrain.

Overall, VIO-SLAM appears to have a promising future for unmanned aerial vehicles. The next generation of intelligent and autonomous aerial systems will only be possible with the continuation of research and development in this field.

ACKNOWLEDGMENT

We would like to express our gratitude to the authors of the research we have examined for their essential contributions to the area. We owe a debt of gratitude to them for their devotion and hard work in making our analysis possible.

REFERENCES

[1] Qin, Tong; Li, Peiliang; Shen, Shaojie (2018). VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. IEEE Transactions on Robotics, (), 1–17. doi:10.1109/tro.2018.2853729

[2] Delmerico, Jeffrey; Scaramuzza, Davide (2018). [IEEE 2018 IEEE International Conference on Robotics and Automation (ICRA) - Brisbane, Australia (2018.5.21-2018.5.25)] 2018 IEEE International Conference on Robotics and Automation (ICRA) - A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots. , (), 2502–2509. doi:10.1109/ICRA.2018.8460664

[3] Weinstein, Aaron; Cho, Adam; Loianno, Giuseppe; Kumar, Vijay (2018). VIO-Swarm: A swarm of vision based quadrotors. IEEE Robotics and Automation Letters, (), 1–1. doi:10.1109/LRA.2018.2800119 [4] Quan, Meixiang; Piao, Songhao; Tan, Minglang; Huang, Shi-Sheng (2019). Accurate Monocular Visual-Inertial SLAM Using a Map-Assisted EKF Approach. IEEE Access, 7(), 34289–34300. doi:10.1109/ACCESS.2019.2904512

[5] Lee, Hongyun; McCrink, Matthew; Gregory, James W. (). [American Institute of Aeronautics and Astronautics AIAA Scitech 2019 Forum - San Diego, California ()] AIAA Scitech 2019 Forum - Visual-Inertial Odometry for Unmanned Aerial Vehicle using Deep Learning. , (), –. doi:10.2514/6.2019-1410 [6] Gang, Peng; Zezao, Lu; Shanliang, Chen; Dingxin, He; Xinde, Li (2020). Pose Estimation Based on Wheel Speed Anomaly Detection in Monocular Visual-Inertial SLAM. IEEE Sensors Journal, (), 1–1. doi:10.1109/JSEN.2020.3011945

[7] Li, Shao-peng; Zhang, Tao; Gao, Xiang; Wang, Duo; Xian, Yong (2018). Semi-direct monocular visual and visual-inertial SLAM with loop closure detection. Robotics and Autonomous Systems, (), S0921889018301374–. doi:10.1016/j.robot.2018.11.009

[8] Zuo, Xingxing; Geneva, Patrick; Yang, Yulin; Ye, Wenlong; Liu, Yong; Huang, Guoquan (2019). Visual-Inertial Localization With Prior LiDAR Map Constraints. IEEE Robotics and Automation Letters, 4(4), 3394–3401. doi:10.1109/LRA.2019.2927123

[9] Chen, Chang; Zhu, Hua; Wang, Lei; Liu, Yu (2019). A Stereo Visual-Inertial SLAM Approach for Indoor Mobile Robots in Unknown Environments Without Occlusions. IEEE Access, 7(), 185408–185421. doi:10.1109/ACCESS.2019.2961266

[10] Liu, Quanpan; Wang, Zhengjie; Wang, Huan (2020). SD-VIS: A Fast and Accurate Semi-Direct Monocular Visual-Inertial Simultaneous Localization and Mapping (SLAM). Sensors, 20(5), 1511–. doi:10.3390/s20051511

[11] Rodolphe Dubois;Alexandre Eudes;Vincent Fremont; (2020). AirMuseum: a heterogeneous multi-robot dataset for stereo-visual and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 05, May-2023

inertial Simultaneous Localization And Mapping . 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), (), –. doi:10.1109/MFI49285.2020.9235257

[12] Rosinol, Antoni; Abate, Marcus; Chang, Yun; Carlone, Luca (2020). [IEEE 2020 IEEE International Conference on Robotics and Automation (ICRA) - Paris, France (2020.5.31-2020.8.31)] 2020 IEEE International Conference on Robotics and Automation (ICRA) - Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping. , (), 1689–1696. doi:10.1109/ICRA40945.2020.9196885

[13] Mo Shan;Qiaojun Feng;Nikolay Atanasov; (2020). OrcVIO: Object residual constrained Visual-Inertial Odometry . 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), (), –. doi:10.1109/iros45743.2020.9341660