A REVIEW OF USING VSLAM METHODS IN AUTONOMOUS UAV NAVIGATION IN GPS-DENIED ENVIRONMENT

Hussam Rostum[1], ***József Vásárhelyi***[2]
[1]PhD Student, [2]Associate Professor
[1,2]*Institute of Automation and Info-Communication,*
Faculty of Mechanical Engineeing and Informatics
[1]hussam.rostum@uni-miskolc.hu, [2]jozsef.vasarhelyi@uni-miskolc.hu

ABSTRACT

This review paper centers on strategies employed for location determination in regions lacking GPS signals. It primarily explores a range of VSLAM-based methods. The article delves into Visual Simultaneous Localization and Mapping (VSLAM) and delves into advanced approaches that involve the use of combined VO/IMU solutions and the approaches that involve the use of extended Kalman filters solutions. It compares the efficacy and practical applications of frequently utilized VSLAM methods while also checking the efficiency of previous research by reapplying the algorithms to new data and comparing the results.

1. INTRODUCTION

According to [1], autonomous navigation systems must possess the capability to estimate, perceive, and comprehend their surroundings in order to accomplish tasks such as path tracking, motion planning, obstacle avoidance, and target detection. GNSS, or Global Navigation Satellite Systems, furnish dependable environmental data and instantaneous positioning for self-governing vehicles such as Autonomous Vehicles (AVs) and Unmanned Aerial Vehicles (UAVs). In scenarios where Unmanned Aerial Vehicles (UAVs) are operating in congested and complex environments, GNSS signals may encounter issues like fading, multipath effects, jamming, and spoofing, which can result in signal loss. Hence, GNSS may not be a reliable solution for UAVs flying at lower altitudes in terrains marked by numerous obstructions, such as forests, cities, and canyons/mountains [2]. To tackle this problem, over the recent years, scientists have devised various localization approaches specifically for UAVs, including vision-based and LIDAR-based techniques. Incorporating multiple sensors in robotic localization studies can be an expensive undertaking, and it may also add extra weight and power consumption to the device. The challenges mentioned above have impeded the widespread implementation of several robot localization algorithms [3]. As an alternative, visual localization, a computer vision-based technology, has emerged as an appealing option. The operating principle of visual localization is based on capturing images of the surrounding environment using a visual camera, followed by determining the position and orientation of the area around it, thereby producing a map of the unknown territory. One of the major advantages of this technology is that the camera's cost is relatively low, and it has the ability to capture extensive environmental data, including visual aspects such as color and texture. Nonetheless, visual localization demands high computational power, and images require a significant amount of storage space, and software development for visual localization is relatively challenging. Moreover, the visual system is susceptible to lighting conditions, and it may not function effectively in poorly lit environments. Visual Simultaneous Localization

and Mapping (VSLAM), on the other hand, can effectively reduce drift in state estimation by revisiting the previously mapped regions since it generates a globally consistent map estimate. Absolute Visual Localization (AVL), in contrast, mainly depends on offline map registration. Consequently, AVL typically does not have position and state estimation drift issues and has a known error range and estimation accuracy. Hence, Absolute Visual Localization (AVL) can determine the absolute position by matching the edge features of the environmental image with the geo-referenced image [12]. This literature review summarizes the current state of research and challenges facing VSLAM technologies in environments where GPS is unavailable. The primary contributions are summarized in the following areas:

- The paper classifies and summarizes the latest research progress in visual SLAM algorithms used for localization, while describing their respective characteristics.
- The paper summarizes the main challenges in the development of localization technologies under GPS-denied conditions and proposes potential solutions.
- The research prospects in this field are also introduced, which can help researchers to focus on the main research problems and bottlenecks.

The structure of this review paper is as follows. In Section 2 it introduces VSLAM methods and optimization techniques. Section 3 concludes the application of the VSLAM. Section 4 concludes the conclusion of the review article.

2. VSLAM-BASED METHODS

Despite the fact that monocular Visual Odometry (VO)-based techniques can be quite effective, they are unable to accurately determine distances, which may restrict their use in certain tasks. Rather than VO, SLAM-based methods are able to generate exact metric maps using their sophisticated SLAM algorithms. In comparison, VO has a tendency to forget the world structure that it has experienced, while VSLAM is able to retain a lasting map of the environment. Visual-SLAM (VSLAM) relies on one or more cameras as its primary means of perception, which is generally more cost-effective than SLAM systems that use LiDAR yet can produce superior localization results on multiple datasets [14]. Without GPS, which provides exact positioning, Visual Odometry/Visual Simultaneous Localization and Mapping (VO/VSLAM) can be used as a viable alternative to the wheel odometer/inertial odometer. Using a combination of two data streams, Visual-inertial SLAM (VI-SLAM) has the potential to increase the precision, dependability, and rate of state estimation when compared to VO/VSLAM which only utilizes visual sensors. Fig. 1. illustrates the whole SLAM system which integrates sensor data, front-end visual odometry, backend nonlinear optimization, loop detection, and mapping as outlined in [13].

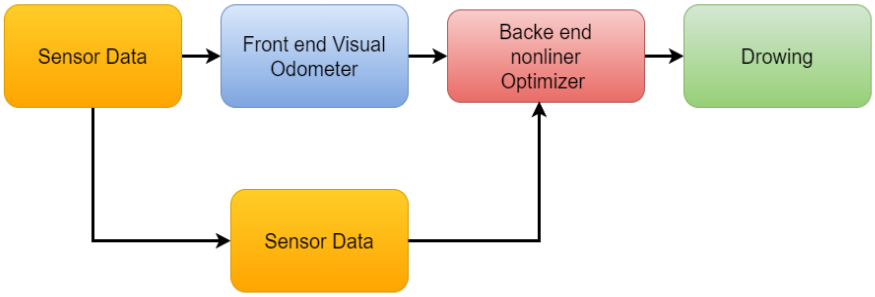


Fig. 1.
Schematic diagram of the visual SLAM framework

Fusing the data from a Visual Odometry (VO) system and an Inertial Measurement Unit (IMU) is a popular approach to navigation when it comes to visual simultaneous localization and mapping (VSLAM). This method combines the speed of the optical flow localization system with the inertial navigation system, both of which work independently. The variance between the velocity data from the optical localization approach and the positioning data from the inertial navigation system can be used as an input for the system. The speed miscalculation of the inertial navigation system can be estimated with the help of the Kalman filter and the mistakes of the inertial navigation system can be adjusted accordingly. It has been proposed that the state vector can be subdivided based on if image feature information is incorporated or not, falling into either the loosely coupled or tightly coupled categories [15]. Rather than estimating the states of the visual localization system and the IMU system separately, loose coupling determines the estimation results; while tight coupling assesses the camera pose and the state determined by the IMU together. Generally, the outcome of tight coupling tends to be more effective. A schematic diagram demonstrating the differences between loose and tight coupling in an integrated navigation system is presented in Fig. 2.

The most current integration of visual odometry (VO) and inertial measurement unit (IMU) approaches (which incorporate a close link between the two) can be broken down into either optimization-based approaches or those based on an extended Kalman filter.

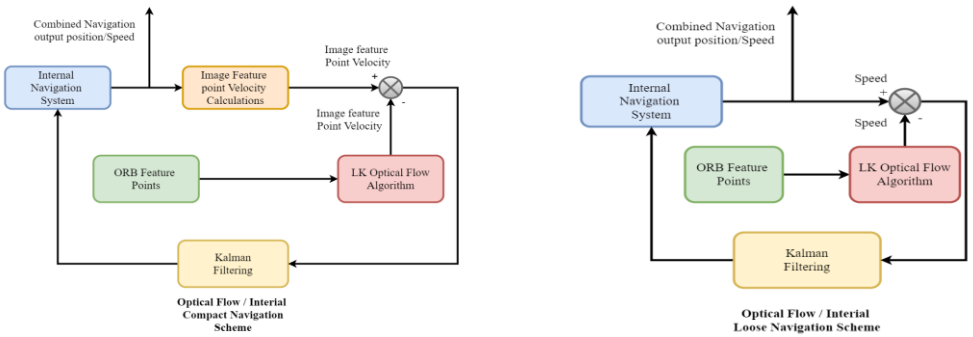


Fig. 2.
Schematic diagram of loose and tight coupling in an integrated navigation scheme

Optimization-based approaches that involve the use combined VO/IMU solutions

- VINS-Mono [7] takes a tightly integrated approach to visual and inertial joint state estimation, utilizing a nonlinear least squares optimization technique over a rolling window of measurements to not only determine the external parameters and latency of the sensor, but also to guarantee the high level of accuracy of the Visual Odometry (VO).
- OKVIS [6] uses a closely integrated Visual-Inertial Odometry (VIO) technique, which employs Harris corners and BRISK feature point detection to establish temporal and stereo matches. It also involves the use of all visible measurements from relevant keyframes to calculate motion states, and IMU sensors to anticipate state values. Finally, it optimizes a cost function with various parameters to perform nonlinear optimization.
- ORB-SLAM [3] is a single camera SLAM system which operates solely on sparse feature points, using ORB as its main feature. It incorporates a complicated sensor fusion method and the detection and loop closing technique to enhance the accuracy of its estimation of the state and error. The algorithm can adjust the assembled sparse feature map while simultaneously permitting the camera position to be moved within the sparse feature map without any pre-existing positional data. ORB-SLAM2 [4] has been augmented to include binocular and RGB-D cameras, utilizing the Bundle adjustment process to identify camera poses and coordinates, as well as producing a reconstruction of the surrounding environment objects. ORB-SLAM3 [5] utilizes a feature-based tight coupling approach, relying solely on maximum posteriori estimation, thereby enhancing localization precision in various contexts. The algorithm implements the Atlas system, which is beneficial for maintaining the tracking process. When the tracking is interrupted, the system can automatically re-align the current frame with the pre-constructed map, thereby improving the robustness of localization.

Optimization-based approaches that involve the use extended Kalman filters solutions

- MSCKF [9] is an example of a visual-inertial algorithm that uses an extended Kalman filter, by relying on IMU data to forecast the filter's state and visual feature measurements to modify the state vector. This algorithm is more robust than the more general VO algorithm, and is especially suitable for environments with sparse visual texture, as well as being more amenable to platforms with limited computing power.
- ROVIO [8] combined the data from vision and inertial sensors through a Kalman filter algorithm. They used QR decomposition to lower the complexity of the least squares problem. This was accomplished by tracking characteristics of multiple multi-level image blocks while further increasing the solution speed.

- SVO [10][11] is an approach to visual odometry that amalgamates the feature point approach and the direct method. Different from ORB-SLAM, SVO utilizes a gray value matching method rather than a feature-based one. To further advance the performance of SVO, Foster et al. proposed SVO_PRO as an improvement. This algorithm is equipped with the ability to handle monocular, binocular, and array cameras. Moreover, it makes use of OKVIS to refine the pose and the feature points, which contributes to enhancing the algorithm's effectiveness. Fig. 3 shows a schematic diagram of the SVO algorithm.



Fig. 3.
Schematic diagram of the SVO algorithm

3. VSLAM APPLICATIONS

Cheng et al. [16] enhanced the original ORB-SLAM system by taking advantage of optical flow to distinguish and eliminate dynamic features, thus avoiding the effect of dynamic objects on the performance of VSLAM. Zheng and his colleagues [17] enhanced the existing VSLAM technique based on point features and endeavored to incorporate line features in the computation to battle the fragility of point features in scenarios with low illumination and texture changes. They offer a novel Trifo-VIO based VSLAM method, which tightly binds filtering and line features to strengthen the system's robustness in hard conditions. Building on previous investigations into line characteristics, Lim et al. [18] devised a technique to identify deficient lines which can appear when line features are used. Moreover, they came up with a line tracking method based on optical flow to enhance the stability of line matching. The EuRoC dataset was used to carry out experiments to validate the results of the advanced monocular SLAM system. It was demonstrated that this system was able to generate more precise localization and mapping results than VINS-Mono. Li et al. [19] conducted analogous research and put forth a rapid line feature description and matching approach based on a midpoint and compared three-line detection algorithms: LSD (line segment detector), FLD (fast line detector), and EDLines (edge drawn lines). Moreover, they merged the above technique with VINS-Mono and generated FPL-VIO. In comparison to similar line feature-based approaches, FPL-VIO enhances line processing speed by a factor of 3-4 times while keeping the same precision [19]. It is evident that the line features have been enhanced in comparison to VSLAM techniques, as evidenced by several studies [20][21]. Lin et al. offered a multi-sensor combination method that capitalizes on ORB-SLAM, IMU, and wheel odometer. IMU and wheel odometer offer accurate approximations of the robot's movement distance in the world coordinate system in order to adjust the issue of mapping scale deviation, which might be present in ORB-SLAM during localization. Studies have demonstrated that it has a remarkable capacity to pinpoint its location in places where GPS is not available [22]. Shan and his colleagues suggested the VINS-RGBD system, which is founded on

VINS-Mono and combines IMU and RGBD camera information to use depth data during both the initialization phase and the VIO stage. Moreover, they integrated a mapping system based on condensed depth data and octree filtering to enable real-time mapping. It is evident that the VINS-RGBD system produces better results than VINS-Mono and ORB-SLAM when evaluated on hand-held, wheeled, and tracked robots [23]. Xu et al. [24] developed an RTLS (Real-Time Locating System) that can be used in a GPS-deprived space. This system combines a sparse feature map of ORB2-SLAM (RGBD camera) and an Occupancy Grid Map to reduce memory usage in vast areas, while still allowing user interaction and path planning. The concept of their RTLS system and RTLS do not necessitate any outside tools or be contingent on any existing human framework, making its deployment quick and easy. Sun et al. [25] opted for the MSCKF algorithm as the basis for their stereo VIO solution, S-MSCKF, because it is more reliable in various settings and scenarios. Contrary to the popular notion that stereo vision-based estimation would be more computationally expensive than monocular methods, S-MSCKF was revealed to be equally efficient or even faster than OKVIS and VINS-MONO. Previous investigations [25] and [26][27] have explored similar topics. Building on the MSCKF, Zhang et al. employed a variational Bayesian adaptive algorithm to further enhance the results of the algorithm. They modeled the noise as observation noise and the covariance as an inverse Wishart distribution with harmonic mean, a departure from the standard Gaussian noise with fixed mean and covariance used in the initial algorithm [28]. Patrick and his associates were able to make considerable enhancements to the algorithm's performance by reworking the Schmidt-Kalman formulation within the MSCKF structure. [29] Similarly, in order to address the situation where VIO is impeded by additional directions that are beyond observation during special motions (such as local acceleration), Ma et al. created ACK-MSCKF. This algorithm combines the Ackermann error state measurement with the tightly coupled filter-based mechanism of S-MSCKF, using the additional constraints of the Ackermann measurement to advance the preciseness of the pose estimation. The accuracy of S-MSCKF for pose estimation is substantially improved by the use of ACK-MSCKF in certain kinds of autonomous vehicle maneuvers, and this holds true across a wide range of vehicle operations and environmental conditions [30]. Additional studies on the MSCKF have been undertaken recently, such as those in [32] shows a summary of the VSLAM method.

4. CONCLUSION

This review examines the VSLAM localizing methods without GPS. The research was divided into Optimization-based approaches that involve the use of combined VO/IMU and approaches that involve the use of extended Kalman filter solutions. VSLAM is a powerful technique that enables devices to autonomously navigate and create maps of unknown environments by using visual information, opening up possibilities for various applications in robotics and computer vision.

Table 1**.**
Provides a concise overview of the VSLAM technique

Research	Method	characteristic
Cheng et al. [16] Lin et al. [22] Xu et al. [23]	ORB-SLAM	- Exploring the use of OF to identify and remove feature points - Adjusting the inclination of the VSLAM with IMU. - The joint utilization of an Optimized Graphical Model (OGM) and a Sparse Feature Map is being implemented.
FPL-VIO [19] VINS-RGBD [23]	VINS-Mono	- Exploring ways of enhancing the performance of Monocular SLAM through the utilization of line features. - The integration of IMU and RGBD camera data in VSLAM
S-MSCKF [25] Zhang [28] Patrick [29] ACK-MSCKF [30]	MSCKF	- Utilizing the Binocular Camera in the MSCKF - Enhancing Effectiveness via Variational Bayesian Adaptive Tech - Revising the Schmidt-Kalman equation within the construction to boost effectiveness. - Applying further limitations from Ackermann tests

There are different approaches to VSLAM, including feature-based methods that rely on extracting and matching visual features, and direct methods that directly optimize the alignment of image pixels between frames. Selecting the most suitable approach involves considering trade-offs between accuracy, computational complexity, real-time performance, and adaptability to different environmental conditions. Researchers and practitioners often evaluate these methods based on benchmark datasets, comparing performance across different metrics to determine which approach suits their needs best. But the feature-based methods (ORB-SLAM and SIF/SURF based methods) still the most effective approach.

REFERENCES

[1] AGUILAR, Wilbert G., Verónica P. CASALIGLLA and José L. PÓLIT: **Obstacle avoidance based-visual navigation for micro aerial vehicles**. Electronics 6.1, 2017, p. 10

[2] CHAO, Haiyang, et al.: **A comparative study of optical flow and traditional sensors in UAV navigation**. 2013 American Control Conference. IEEE, 2013.

[3] MUR-ARTAL, Raul, Jose Maria Martinez MONTIEL and Juan D. TARDOS: **ORB-SLAM: a versatile and accurate monocular SLAM system**. IEEE transactions on robotics 31.5 (2015): 1147-1163.

[4] MUR-ARTAL, Raul and Juan D. TARDÓS: **Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras**. IEEE transactions on robotics 33.5 (2017): 1255-1262.

[5] CAMPOS, Carlos, et al.: **Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam**. IEEE Transactions on Robotics 37.6 (2021): 1874-1890.
[6] LEUTENEGGER, Stefan, et al.: **Keyframe-based visual–inertial odometry using nonlinear optimization**. The International Journal of Robotics Research 34.3, 2015, pp. 314-334.
[7] QIN, Tong, Peiliang LI and Shaojie SHEN: **Vins-mono: A robust and versatile monocular visual-inertial state estimator**. IEEE Transactions on Robotics 34.4, 2018, pp. 1004-1020.
[8] BLOESCH, Michael, et al.: **Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback**. The International Journal of Robotics Research 36.10 (2017): 1053-1072.
[9] MOURIKIS, Anastasios I. and Stergios I. ROUMELIOTIS: **A multi-state constraint Kalman filter for vision-aided inertial navigation**. Proceedings 2007 IEEE international conference on robotics and automation. IEEE, 2007.
[10] FORSTER, Christian, Matia PIZZOLI and Davide SCARAMUZZA: **SVO: Fast semi-direct monocular visual odometry**. 2014 IEEE international conference on robotics and automation (ICRA). IEEE, 2014.
[11] FORSTER, Christian, et al.: **SVO: Semidirect visual odometry for monocular and multicamera systems**. IEEE Transactions on Robotics 33.2, 2016, pp. 249-265.
[12] LAN, Huang and Song JIANMEI: **Research of autonomous vision-based absolute navigation for unmanned aerial vehicle**. 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV). IEEE, 2016.
[13] WEI, Wenle, et al.: **A survey of uav visual navigation based on monocular slam**. 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2018.
[14] GEIGER, Andreas, Philip LENZ and Raquel URTASUN: **Are we ready for autonomous driving? the kitti vision benchmark suite**. 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.
[15] CHEN, Di, et al.: **Research on Navigation Algorithm of UAV Optical Flow/Inertia Multi-Information Fusion**. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Vol. 1. IEEE, 2020.
[16] CHENG, Jiyu, Yuxiang SUN and Max Q-H. MENG: **Improving monocular visual SLAM in dynamic environments: An optical-flow-based approach**. Advanced Robotics 33.12, 2019, pp. 576-589.
[17] ZHENG, Feng, et al.: **Trifo-VIO: Robust and efficient stereo visual inertial odometry using points and lines**. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018.
[18] LIM, Hyunjun, et al.: **Avoiding degeneracy for monocular visual SLAM with point and line features**. 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021.

[19] LI, Wen-Kuan, et al.: **A fast vision-inertial odometer based on line midpoint descriptor**. International Journal of Automation and Computing 18.4, 2021, pp. 667-679.
[20] ZHANG, Peizhi, et al.: **Vins-pl-vehicle: Points and lines-based monocular vins combined with vehicle kinematics for indoor garage**. 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2020.
[21] WEN, Huanyu, Jindong TIAN and Dong LI: **PLS-VIO: Stereo vision-inertial odometry based on point and line features**. 2020 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS). IEEE, 2020.
[22] LIN, Jian, et al.: **ORB-SLAM, IMU and Wheel Odometry Fusion for Indoor Mobile Robot Localization and Navigation**. Acad. J. Comput. Inf. Sci 3 (2020).
[23] SHAN, Zeyong, Ruijian LI and Sören SCHWERTFEGER: **RGBD-inertial trajectory estimation and mapping for ground robots**. Sensors 19.10, 2019, p. 2251.
[24] XU, Lichao, et al.: **An occupancy grid mapping enhanced visual SLAM for real-time locating applications in indoor GPS-denied environments**. Automation in Construction 104, 2019, pp. 230-245.
[25] SUN, Ke, et al.: **Robust stereo visual inertial odometry for fast autonomous flight**. IEEE Robotics and Automation Letters 3.2, 2018, pp. 965-972.
[26] RAMEZANI, Milad and Kourosh KHOSHELHAM. "Vehicle positioning in GNSS-deprived urban areas by stereo visual-inertial odometry." IEEE Transactions on Intelligent Vehicles 3.2 (2018): 208-217.
[27] RAMEZANI, Milad, Kourosh KHOSHELHAM and Clive FRASER: **Pose estimation by omnidirectional visual-inertial odometry**. Robotics and Autonomous Systems 105, 2018, pp. 26-37.
[28] ZHANG, Zhuqing, et al.: **Improving S-MSCKF with variational Bayesian adaptive nonlinear filter**. IEEE Sensors Journal 20.16, 2020, pp. 9437-9448.
[29] GENEVA, Patrick, Kevin ECKENHOFF and Guoquan HUANG: **A linear-complexity EKF for visual-inertial navigation with loop closures**. 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.
[30] MA, Fangwu, et al.: **ACK-MSCKF: Tightly-coupled Ackermann multi-state constraint Kalman filter for autonomous vehicle localization**. Sensors 19.21, 2019, p. 4816.
[31] HEO, Sejong, Jaehyuck CHA and Chan Gook PARK: **EKF-based visual inertial navigation using sliding window nonlinear optimization**. IEEE Transactions on Intelligent Transportation Systems 20.7, 2018, pp. 2470-2479.
[32] ECKENHOFF, Kevin, et al.: **Multi-camera visual-inertial navigation with online intrinsic and extrinsic calibration**. 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019.