LGVINS: LiDAR-GPS-Visual and Inertial System Based Multi-Sensor Fusion for Smooth and Reliable UAV State Estimation

Mahammad Irfan, *Graduate Student Member, IEEE*, Sagar Dalai, *Graduate Student Member, IEEE*, Petar Trslic, Matheus C. Santos, *Member, IEEE*, James Riordan, and Gerard Dooly, *Member, IEEE*

***Abstract*—With the development of Autonomous Unmanned Aerial Vehicle's (UAV's), Precise state estimation is a fundamental aspect of autonomous flight and plays a critical role in enabling robots specially in GPS denied environment to operate safely, reliably, and effectively across a wide range of applications and operational scenarios. In this paper, we propose a tightly-coupled multi-sensor filtering framework for robust UAV/UGV state estimation, which integrates data from an Inertial Measurement Unit (IMU), a stereo camera, GPS, and 3D range measurements from two Light Detection and Ranging (LiDAR) sensors. The proposed LGVINS system significantly improves the accuracy and robustness of state estimation in both structured and unstructured outdoor environments, such as bridge inspections, open fields, urban city and areas near buildings. It also improves positioning accuracy in scenarios with or without GPS signals. The goal is to exploit the fact that these sensor modalities have mutually exclusive strengths, the visual, inertial and the Lidar sensor techniques are implemented to compensate for the robots state estimate errors in multiple outdoor challenging environment. It effectively reduces long-term trajectory drift and ensures smooth, continuous state estimation, regardless of GPS satellite availability. We demonstrate and evaluate the LGVINS approach on public dataset as well as our own dataset collected from the proposed hardware integration on UAV, deployed on computationally-constrained systems. This demonstrates that the proposed system achieves higher accuracy and robustness in state estimation across various environments compared to currently available methods.**

***Index Terms*—Intelligent UAV/UGV, LiDAR-visual-inertial odometry, multi-sensor fusion, ROS, state-estimation.**

Received 30 June 2024; revised 8 September 2024; accepted 18 September 2024. Date of publication 27 September 2024; date of current version 17 October 2025. This work was supported in part by the European Commission's Horizon 2020 Project RAPID under Grant 861211, and in part by the Enterprise Ireland's Disruptive Technologies Innovation Fund (DTIF) Project GUARD under Grant DT2020 0286B. *(Corresponding author: Mahammad Irfan.)*

Mahammad Irfan and Sagar Dalai are with the CRIS Research Group, Department of Electronic and Computer Engineering, University of Limerick, V94 T9PX Limerick, Ireland (e-mail: mahammad.irfan@ul.ie).

Petar Trslic, Matheus C. Santos, and Gerard Dooly are with the CRIS Research Group and Department of School of Engineering, University of Limerick, V94 T9PX Limerick, Ireland.

James Riordan is with the ALMADA Research Centre, School of Computing, Engineering and Physical Science, University of the West of Scotland, G72 0LH Glasgow, U.K..

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIV.2024.3469551.

Digital Object Identifier 10.1109/TIV.2024.3469551

I. Introduction

THE field of Autonomous and Intelligent Unmanned Aerial Vehicle (UAV) has grown tremendously over the past few years, as they have potential to explore unknown environments and perform challenging tasks thereby reducing the need for human intervention [1]. UAV flight control is primarily dependent on GPS and some use-cases where GPS robustness is low, there are safety concerns. For example, in the application of autonomous bridge inspection, GPS signal under or close to the bridge suffers from low signal strength and multi-path interference. An accurate localization and stable odometry position can be crucial in these GPS-denied use cases to provide robust autonomous UAV navigation and control. Navigation and control are difficult in close quarter surveys close to infrastructure, however other perception payloads such as LiDAR and stereo-camera can aid with localizing position and automatically building up a surrounding map. These systems provide scene perception for vehicle understanding, learning, and ultimately interacting with the surrounding. For autonomous drones, 3D perception of the environment with fusing data from several sensors is a crucial task, which can be used to achieve an improved localization in an obstacle-rich environment [2].

In the autonomous UAV navigation, localization and state estimation task, an accurate GPS-RTK position is always necessary during outdoor applications. One of the challenges is the safe navigation in law GPS environment which requires a robust UAV pose state estimation to rely on highly dynamic responses. However, an autonomous UAV localization and pose estimation in a complex scenario using a vision-based SLAM operation cannot easily be solved due to error and uncertainties in onboard sensors, resulting in inaccurate localization [3]. These sensors when used individually are not accurate enough to produce precision drone position estimation. In order to solve these challenges, an optimized multi sensor fusion approach is implemented with SLAM which ultimately minimizes single sensor errors but also maximizes the UAV pose and localization by fusing the information from multiple sensors and pass it through a Kalman filter [4]. This will provide the level of integrity and redundancy that a multi sensor fusion SLAM system requires during autonomous UAV operation. Recent research, however, has focused on integrating camera-LiDAR multi-sensor fusion into SLAM systems [5]. The sensors such as LiDAR and stereo

TABLE I
SUMMARY OF EXISTING MULTI SENSOR FUSION METHODS

Algorithms	Fusion Type	Optimization	Evaluated Platform	Sensor Suite
LGVINS (Proposed)	Tightly Coupled	Filter	UAV, UGV	Imu=1, Camera+Lidar+GPS=Any No.
Wu et al. (2022) [18]	Loosely Coupled	Filter	UGV	Imu+GPS+Lidar=1, Camera=2
Meng et al. (2017) [19]	Tightly Coupled	Filter	UGV	Imu+GPS+Lidar=1
Hausman et al. (2016) [20]	-	Filter	UAV	Imu+GPS+Camera=1
Kubelka et al. (2015) [21]	-	Filter	UGV	Imu+Lidar+Camera=1
VINS Fusion (2019) [6]	Tightly	Graph	UGV	Imu+Camera=2, GPS= Any No.
LVIO-Fusion (2021) [22]	Tightly	Graph	UGV	Imu+Lidar+GPS=1, Camera=2
GR-Fusion (2021) [2]	Tightly	Graph	UGV	Imu+Lidar+GPS=1, Camera=2
VIRAL Fusion (2022) [23]	Tightly	Graph	UAV	Imu=1, Lidar+Camera=Any No.

cameras provide limited information by themselves, but when combined creates a system resembling an eye, differentiating colors, features and an accurate drone position and orientation.

The key contribution of this proposed paper lies in addressing the challenges associated with UAV/UGV state estimation. Based on our knowledge, this study is among the first to establish a consistent and tightly-integrated multi sensor fusion, designed to efficiently fuse commonly used sensors by addressing challenges like processing complexity, sensor synchronization issues, and intra-sensor calibration. Although this article builds on the ideas from previous studies [6], [7] and [8], it introduces several distinct contributions and novelty as discussed in Table I. Specifically, LGVINS utilizes the tightly coupled filtering framework, allowing for the seamless high-frequency fusion of multiple sensors regardless of their timing or delay within a unified continuous-time structure. Benefiting from this, multiple sensors can be easily fused into the framework as new factors. Furthermore, this paper provides a novel fusion technique which automatically includes/excludes GPS position to accurately fuse true GPS position into state estimation as per the available satellite signal that is robust to various challenging scenario. Proposed LGVINS has added a new Lidar fusion technique, embedding it within the existing VINS framework to facilitate multi-LiDAR fusion. The system adopts the direct alignment method [9] in place of other like LOAM feature extraction, enhancing both the efficiency and consistency of updates.

The proposed method builds upon and enhances the VINS and VINSFUSION [10] [4] models. We demonstrated a fusion system which quickly and accurately determines the orientation and position of the vehicle in order to give better state estimation during close infrastructure inspection such as bridge survey operation, near the high-rise buildings or cluttered environment. Our proposed approach architecture is represented in the Figs. 2 and 1 depicts the proposed LGVINS with its trajectory plotted on google map, indicating a strong alignment with RTK ground truth data obtained using our handheld UAV integrated with proposed multiple sensors at the University of Limerick campus living bridge area. The Fig. 1 also shows the GPS signal drifts near the high-rise buildings, however LGVINS trajectory remains reliable and smooth. Hence, the developed system can allow to the successful execution of unmanned flight close to target, allowing the drone to navigate in a GPS obstructed or GPS denied environment.

Fig. 1. Top view of Google maps plot of the proposed LGVINS global estimated result with trajectory aligned well with the ground truth obtained at UL living bridge area.

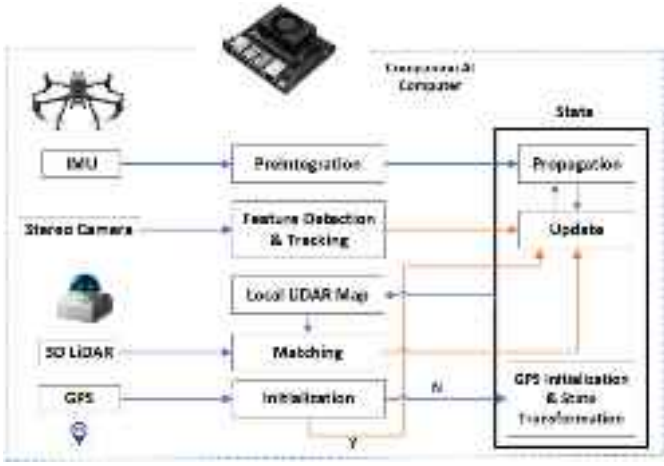

Fig. 2. Depicts the proposed LGVINS architecture.

This paper excels in addressing state estimation challenges. The algorithms presented have the capability to autonomously choose relevant sensor data, thereby improving positioning accuracy in situations with or without GPS signals. Furthermore, in the absence of GPS, it can leverage Stereo-inertial or Lidar-intertial odometry outcomes to ensure global positioning.

The key contributions of our work can be outlined as follows:

- Proposed method leverages a tightly integrated multi-sensor fusion framework comprising a VGA stereo camera, two 3D LiDAR sensors, a 9 DOF IMU, and GPS-RTK network optimization to achieve highly accurate trajectory state estimation validated and tested both in UAV and UGV platform.
- Acquisition of challenging datasets using commercial UAV with in-house integrated and calibrated sensor hardware platform.
- The performance and effectiveness of the proposed Stereo-Visual-LiDAR fusion framework are validated through experimental evaluations which achieve high efficiency, robustness, consistency and accuracy in many challenging scenarios.
- Validation demonstrates, drift-free 6-DoF global estimations in environments such as beneath bridges and other areas where GNSS signals are frequently lost for extended periods without loop closure. The developed algorithm automatically includes/excludes GPS position as per the available satellite signal.

This paper is organized as follows Section II reviews the relevant work. Section III describes the LGVINS Methodology. Section IV describes Hardware and Experimental Setup. Section V describes the Experiments and Results. Finally, Section VI reports the conclusions and outlines future work.

II. Related Work

Over the past few years, several UAV SLAM navigation techniques in challenging environment have been proposed [11]. 3D LiDAR based and Visual SLAM algorithms have their own benefits and disadvantages. Visual SLAM systems often use a monocular or stereo camera in order to determine the camera motion. Open-source vision-inertial slam, are sensor cost efficient and perform well in many scenarios, but the vision-based approaches become unreliable in poor light and sensitive to illumination change which results prone to errors and an inaccurate odometry localization. However, SLAM when used with RGB-D based cameras are sensitive to sunlight as it has an inbuilt IR sensor. Therefore, these sensors are only useful in indoor applications. Furthermore, the performance of these SLAM algorithms is often challenging, especially in long-range large-scale SLAM applications like a bridge inspection or outdoor open field operation during a sunny and dark day time [12].

During the past few years, different open-source SLAM algorithms has been introduced for the robotics research community based on sensors used such as monocular, stereo, event cameras, RGB-D, and when integrated them with an internal/external IMU measurements known as visual-inertial odometry (VIO) based SLAM approach. These techniques are mostly feature based indirect and photometric or direct methods, However VIO methods can be categorized into semi dense, dense, sparse etc. Nowadays, LiDAR and RADAR sensors-based SLAM approaches are mostly implemented in autonomous cars [13]. Other research approaches have been conducted in aerial vehicles with 3D/2D LiDAR based SLAM such as FASTLIO2, LIO-SAM, LeGO-LOAM etc. [14]. Multi-sensor fusion technique with combination of stereo and Lidar odometry is a new widely researched topic in the robotics community. Recent research shows low-cost consumers grade sensors may also be viable in autonomous vehicles for multi-sensor fusion and internet of vehicle communication etc. [15]. An ArUco marker-based SLAM algorithm has also been introduced in an indoor UAV navigation for path detection, guidance and an improved odometry fusion using a stereo camera [16]. An improved fusion odometry results in robust localization which help the UAVs to navigate in GPS denied environments and help to achieve an Optimus control, obstacle avoidance and path planning [17].

The implementation of a sensor-fusion approach for state estimation has proven its effectiveness and robustness, with abundant literature supporting this assertion. Although there is a rich literature on the proposed topic however, our focus lies in the integration of compact sensors like cameras, IMUs, Lidar and GPS receivers, aiming to achieve precise estimation of state in challenging environments. While systems equipped with any number of sensor modalities boast advantages in terms of robustness, accuracy, and versatility, their development has been limited. This is largely due to the complexities involved in processing large-scale computations and managing sensor data that may be asynchronous, delayed, or both. Effective management of asynchronous and delayed sensor measurements is crucial in multi-sensor systems. This functionality allows the system to flexibly integrate in any number of sensors, whether they are homogeneous or heterogeneous, accurately model the measurements, and retain all available data without unnecessary loss. Table I provides a summary of the relevant literature. The fusion of VI system with a tightly coupled, is classified by graph/optimization-based methods, learning-based approaches, or filter-based methodologies.

A. Graph Based Approaches

Optimization methods are classified into two parts: the back/front end. The front end handles map construction through image processing, while the back end aim into pose optimization. Graph-based formulations excel at handling delayed measurements because they can maintain a complete or partially obtained measurements. This capability enables delayed states to be seamlessly incorporated into the optimization problem. The latest VINS-Fusion [4] uses a pose graph to loosely combine VINS which tightly integrates IMU and camera data with other global sensors like GNSS, magnetometers, and barometers, under the assumption that all sensors are perfectly synchronized. OKVIS [24], an open keyframe-based visual inertial SLAM system, introduced bundle adjustment along with keyframe-based sliding window nonlinear optimization, marking a significant advancement in VO technology

Lvio-fusion [22] tightly integrates IMU, camera, and LiDAR for trajectory estimation, while incorporating GNSS data in a pose graph with a looser coupling. VINS-Mono [7] incorporates loop-closure constraints into both local sliding window and global pose graph optimization. It also incorporates efficient IMU pre-integration with bias correction, online extrinsic calibration, and the capability for relocalization. The

VINS-Mono [7] framework integrates loop closure technique into both pose graph and sliding window method. It also includes highly desired preinegrated IMU with correct biases, live calibrated offset parameters, and the capability for relocalization.

Furthermore, VIRAL SLAM [23]merges IMU and stereo camera data with multiple UWBs and LiDARs by aligning all sensor measurements with the base LiDAR measurement time. This process entails interpolating IMU and UWB data, aligning non-base LiDAR point clouds with the base LiDAR measurement time, and choosing camera measurements that are closest in time while disregarding the rest. Through synchronizing sensors via independent state creation regardless of sensor measurement timings, VIRAL-Fusion [23] uses interpolating states to accommodate UWB measurements, and producing synthetic measurements at the state timestamp using measurement interpolation. GR-FUSION [2] tightly integrates visual, inertial, LiDAR, and gps data to find the true odometry position of a robot. It leverages LiDAR data and scenes from object to extract depth information of visual features and improve their quality. Moreover, it chooses both stereo-point cloud from lidar-camera for tracking and integrates all the sensors in a coupled manner to optimize the robot's current state.

B. Learning Based Approaches

In recent times, there has been a growing research in using deep neural networks to learn and predict robots states. Learning-based methods offer the advantage of automatically learning features rather than relying on hand-crafted ones. These attributes empower deep learning-based approaches to achieve greater robustness, particularly in scenarios where images exhibit variations in lighting or textures. These methodologies comprise probabilistic methods, statistical approaches, knowledge-based techniques (including fuzzy logic and possibility theory), interval analysis methods, and evidential reasoning methods. LO-Net [25] explores learning LIDAR odometry by utilizing geometric evaluation based motion identification and integrating pose correction similar to deep simultaneous localization and mapping (SLAM) techniques. Its accuracy comparable to traditional approaches has been demonstrated through experimentation.

Lvio-Fusion [22] is a tightly integrated SLAM framework that utilizes 3D LiDAR, stereo camera, IMU, and GPS. It employs factor graph optimization to fuse sensor data and incorporates a lightweight deep reinforcement learning algorithm for adaptive adjustment of sensor model weights. Posenet [26] is the pioneering study to employ convolutional neural networks (CNNs) for the regression of pose from 2D images. Additional improvements to PoseNet technique by integrating convolutional neural networks (CNNs) with long short-term memories (LSTMs) to enhance feature correlation [27]. MS-Transformer [28] represents a recent advancement in relocalization utilizing a transformer architecture, yielding sota results. Nonetheless, deep learning approaches suffer from the drawback of requiring extensive training data and significant computational resources. Hence, traditional fusion approaches are deemed more practical for drone applications.

C. Filtered Based Approaches

The tightly-coupled approach utilizes raw input data from the sensors to collectively establish states. The MSCKF [29] is an impressive filter-based state estimator that leverages the camera poses and geometric constraints to effectively utilize and optimize a robot's state. The state estimation relies on multiple past camera poses, while maintaining tracking of the geometric parameter of camera features across multiple images. Filter-based techniques continue to be ideal over graph-based systems due to their effectiveness, which is essential for handling computationally intensive multi-sensor fusion systems. Moreover, in order to maintain asynchronous and delayed sensor measurements is challenging, as filters usually maintain only the most information and a brief historical record of stochastic poses. Recent studies show that it can be feasible to study odometry estimation from imu data using recurrent neural networks, thereby enabling advanced deep VIO estimation techniques. DeepVIO [30] recently introduced a fusion network specifically designed to integrate vo and vio features. This function is trained using a specialized function.

Furthermore, Robust Visual-Inertial Odometry (ROVIO) [31] utilizes patch-based direct photometry use to updates via an iterated ekf. Some initially introduced filtering approaches, such as those by [21]and [32], which fused cameras, IMU, wheel, and LiDAR data using an EKF. [19], which combined GPS, IMU, wheel, and LiDAR data with an unscented Kalman filter (UKF), and [33], which integrated IMU, camera, GNSS, wheel, and symbolic road map data with a particle filter (PF), sensor measurements were assumed to be synchronized and time-ordered. This assumption can introduce unmodeled errors into the estimator when it does not hold. MSF-EKF [34] is one of the pioneering works that addresses asynchronous and delayed sensor measurements by integrating generic relative and absolute values from inertial sensor, visual parameter, GPS, and pressure sensor within an EKF framework. MSF-EKF addresses asynchronicity by forward-predicting the current state from the queried state when a delayed measurement is received, hence requires subsequent measurement updates to be re-executed. Similarly [20] introduced approaches dealing with asynchronous and delayed sensor measurements, which fused IMU, camera, GNSS, and UWB in iterated EKF.

D. Different Sensor Modalities Uses VINS Framework

A GPS sensor offers direct absolute position data, though its accuracy heavily relies on environmental conditions and access to external correction data. Consequently, many researchers have explored integrating GPS with VINS to achieve locally precise and globally accurate localization. Numerous studies have examined the fusion of GPS with VINS through graph formulations, employing both loosely-coupled [4], [35] and tightly-coupled [36], [36], [37] approaches. Yet, GPS measurements frequently occur intermittently and asynchronously with respect to other sensors, complicating the design of the graph construction process. Recent studies have concentrated on integrating lower-level GNSS measurements, such as GNSS

TABLE II
SENSOR CONFIGURATION

Sensor H/W	Type	Specification	Frequency
GPS/RTK	DJI M300	GPS+GLONASS+BeiDou+Galileo	50 Hz
LiDAR	Livox mid360	Laser Wavelength:905 nm FOV: Horizontal: 360°, Vertical: -7°∼52°, Pointrate: 200,000 points/s Detection Range: 40m=10%reflectivity,70m=80% reflectivity	20 Hz
IMU	DJI M300	6 axis Mechanism	400 Hz
Camera	DJI M300	Grayscale image in VGA (640x480) resolution at 10 fps.	15 Hz

satellite signals, with VINS, as demonstrated by [6], [18], [38], [39]and [40].

When we talk about Lidar only case,these efforts showcase enhanced robustness in scenarios with limited field of view (FOV) and improved localization accuracy. LiDAR is commonly integrated with VINS because its measurements remain unaffected by lighting conditions and it offers direct depth information. This capability can address the limitations of cameras in various scenarios, such as navigating in a sunny day, dark caves or encountering featureless walls. Loosely coupled systems have aimed to resolve the degeneracy problems inherent in independent camera and LiDAR odometry algorithms. For instance, LiDAR (combined with IMU) odometry is integrated with VINS in a graph formulation, treating them as separate sub-systems to enhance robustness [41]. Numerous research has employed filtering frameworks to develop tightly coupled systems, resulting in improved estimate consistency and accuracy. MSCKF-based designs, pioneered by [29], have gained popularity and have been successfully implemented in various studies, including those by [18], [40] and [14]. Likewise, the error state iterated Kalman filter (ESIKF) is utilized while fusing LiDAR without extracting the features (a.k.a. direct method) was also studied by [42].

III. METHODOLOGY

In this section, we discuss an in-depth explanation of the LGVINS multi sensor fusion framework and demonstrate how the each sensor modules are integrated. The system is based on modified version of VINS system [7], [8].

A. Data-Preprocessing

The architecture of the proposed method is shown in Fig. 2. The measurement frequency for each sensor is described in Table II. The sampling rates of different sensors differ, like a GPS at 50 Hz, camera at 15 Hz and a LiDAR at 20 Hz, making their measurements generally asynchronous without hardware synchronization. Hence, this leads difficulties in constructing their measurement models, as the corresponding IMU pose states are unavailable at the exact measurement times can be seen in (1). A basic strategy is to integrate IMU poses into the state at the sampling time of each sensor, but this would increase the state size and lead to a substantial computational burden. In comparison, We use linear interpolation method given by [40] of the bounding clones to estimate the IMU pose at the measurement time to calculate 3D position increments of the robot, which significantly reduces computational overhead compared to continuously adding clones to the state with each incoming LiDAR measurement. We use the 1st-order linear interpolation to handle asynchronous and delayed measurements. For the time synchronization, We use ROS timestamps subscribe from dji-osdk and integrated Livox Lidar sensor nodes. In addition, The extrinsic parameters of stereo cameras and camera-IMU, Lidar-IMU are calibrated offline by using the Kalibr tools [43] and [44]. The extrinsic calibrated sensor parameters for public dataset Kaist is obtained from [45].

B. Lgvins State Estimation

In this section, we reframe the state estimation issue within a probabilistic filter graph methodology. Considering INS as the pivotal component of the system, the proposed LGVINS propagates UAV state estimation using IMU as primary measurements and efficient filtering techniques.

It integrates multi-modal measurements from various adopted sensors like cameras, LiDARs, IMU and GPS receivers for accurate odometry updates, which is shown in the Fig. 2, With reference to the VINS based MSCKF approach [8], [29]. The integrated imu navigation state estimation vector x_{I_k} and x_{k_k} is the sum of all the obtained IMU poses in the global frame G:

$$x_k = (x_{Ik},\ x_{N_k}) \tag{1}$$

$$x_{I_K} = ({}^{I_k}_{E}R, {}^{E}p_{I_k}, {}^{E}v_{I_k}, F_g, F_a) \tag{2}$$

$$x_{N_k} = ({}^{I_{k-1}}_{E}R, {}^{E}p_{I_{k-1}}, \ldots, {}^{I_{k-n}}_{E}R\ {}^{E}p_{I_{k-n}}) \tag{3}$$

Where ${}^{I_k}_{E}R$ is known as unit quaternion in JPL for rotation matrix ${}^{D}_{C}R$ corresponding to C,D. ${}^{D}_{C}p$ and ${}^{D}_{C}v$ is position, linear velocity corresponding to C and D respectively. F_g is gyroscope bias and F_a is accelerometer bias. Further we have taken $x = \widetilde{x} \boxplus \bar{x}$, Here $\widetilde{x}$ is state estimation, $\bar{x}$ known as obtained error state, and x is known as true state.

Now, we will discuss each sensor modality and the proposed LGVINS approach for updating and propagating the state estimation corresponding to their measurements.

C. IMU Kinematic Model

The state estimation evolves over time by incorporating high-rate IMU measurements, which typically denote angular velocity ω_v and local linear acceleration a_v:

$$\omega_{v_k} = \omega_k + F_g + n_g \tag{4}$$

$$a_{v_k} = a_k + {}^{I_k}_{E}R\,{}^{E}g + F_a + n_a \tag{5}$$

Here, ω_k and a_k is the true local angular velocity and linear acceleration at time t_k, respectively. F_g and F_a denote the bias for the accelerometer and gyroscope.

$E_g \simeq [0\ 0\ 9.81]^T$ represents the gravity as global, while n_a and n_g denote zero-mean Gaussian noises. The resultant measurements are utilized to propagate the state estimation and covariance from time t_k to t_{k+1}, following the standard inertial kinematic model with zero noise assumption [46]:

$$\hat{x}_{I_{k+1|k}} = f(\hat{x}_{I_{k|k}}\, a_{v_k}, \omega_{v_k}, 0, 0) \tag{6}$$

$$P_{I_{k+1|k}} = \vartheta_I(t_{k+1}, t_k)\, P_{I_{k|k}} \vartheta_I(t_{k+1}, t_k)^{\intercal} + J_k Q_d J_k^{\intercal} \tag{7}$$

where, $I_{k+1|k}$=$I_{C|D}$

Furthermore, $\hat{x}_{I_{C|D}}$ is the estimate at time t_C which is obtained by the processing its measurements in time t_D.ϑ, J and Q are the state transition matrix, Jacobian's of f(.) w.r.t noise vector and discrete noise covariance respectively which is found from [47].

D. Camera Measurement Model

Let's consider a scenario where a 3D feature is identified in a random camera image at time t_k. Its uv measurement, representing the pixel coordinates on the image plane, is obtained from [8]:

$$z_C = h_C(x_k) + n_C \tag{8}$$

$$= h_d\left(h_\rho\left(h_t({}^{E}p_F, {}^{C_k}_{E}R\,{}^{E}p_{C_k})\right),\ x_{CI}\right) + n_C \tag{9}$$

Here, x_{CI} denotes the camera's intrinsic parameters encompassing focal length and distortion parameters, n_C represents the zero-mean white Gaussian noise, z_n denoted by undistorted uv measured in normalized form, ${}^{E}p_F$ is called as feature pose in global, $({}^{C_k}_{E}R\,{}^{E}p_{C_k})$ is the latest stereo visual pose obtained the global frame. In above expression, we break down the measurement function into several concatenated functions, each representing distinct operations that transform the states into raw UV measurements on the image plane. Each parameters (i.e. h_d, h_ρ, h_t) is elaborated on next:

Projection Coefficients (h_p): The standard pinhole camera model is utilized to project a 3D point, ${}^{C_K}p_F = [{}^{C_k}_{x}\ {}^{C_k}_{y}\ {}^{C_k}_{z}]^T$ onto the normalized image plane assuming a unit depth:

$$z_n = h_\rho({}^{C_K}p_F) = \begin{bmatrix} {}^{C_k}x/{}^{C_k}z \\ {}^{C_k}y/{}^{C_k}z \end{bmatrix} \tag{10}$$

Distortion Coefficient (h_d): To obtain the normalized coordinates $z_n = [x_n\ y_n]^T$ of a 3D feature projected onto the image plane, a distortion model tailored to the specific camera lens type is employed. Our LGVINS equipped drone uses VGA 640x480 resolution results grey scale images via ROS which we map to the normalized coordinates into the raw pixel coordinates. In regards with above circumstances, we chose h_d with the camera model:

$$\begin{bmatrix} u \\ v \end{bmatrix} := z_C = h_d(z_n, x_{CI}) = \begin{bmatrix} f_x x + c_x \\ f_y y + c_y \end{bmatrix} \tag{11}$$

Where,

$$x = x_n\ \left(1 + k_1 r^2 + k_2 r^4\right) + 2p_1 x_n y_n + p_2(r^2 + 2x_n^2) \tag{12}$$

$$y = y_n\ \left(1 + k_1 r^2 + k_2 r^4\right) + p_1\left(r^2 + 2y_n^2\right) + 2p_2 x_n y_n \tag{13}$$

$$r^2 = x_n^2 + y_n^2 \tag{14}$$

$$x_{CI} = (f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2) \tag{15}$$

intrinsic parameter x_{CI} is obtained from [48]

Transformation of Euclidean (h_t): To convert the 3D feature position from the frame E coordinate system to the current camera frame c_k, we apply a 6DOF rigid-body Euclidean transformation, which accounts for the current camera pose:

$${}^{C_k}\boldsymbol{p}_F = \boldsymbol{h_t}\left({}^{E}\boldsymbol{p}_F, {}^{C_k}_{E}\boldsymbol{R}, {}^{E}\boldsymbol{p}_{C_K}\right) \tag{16}$$

$$= {}^{C_k}_{E}R({}^{E}p_F - {}^{E}p_{C_k}) \tag{17}$$

The camera pose $({}^{C_k}_{E}R,\ {}^{E}p_{C_K})$ is described using the camera extrinsic $x_{CE} = ({}^{C}_{I}R, {}^{C}p_I)$ and the IMU pose $({}^{I_k}_{E}R, {}^{E}p_{I_K})$:

$${}^{C_k}_{E}R = {}^{C}_{I}R\,{}^{I_k}_{E}R \tag{18}$$

$${}^{E}p_{C_K} = {}^{E}p_{I_K} + {}^{E}_{I_k}R\,{}^{I}p_C \tag{19}$$

MSCKF Update: For the MSCKF update [29], we linearize the measurement function from the (4) at current state estimate $\hat{x}_k and {}^{E}\hat{p}_F$ to obtain the following residual:

$$\tilde{z}_C := z_C - \boldsymbol{h}_C\left(\hat{\boldsymbol{x}}_k, {}^{E}\hat{\boldsymbol{p}}_F\right) \tag{20}$$

$$\approx H_C\tilde{\boldsymbol{x}}_k + \boldsymbol{H}_F\,{}^{E}\tilde{\boldsymbol{p}}_F + \boldsymbol{n}_C \tag{21}$$

Here ${}^{H_C}k$ and H_F represent the Jacobian matrix of the function $h_c(\cdot)$ corresponding to the current state estimate $\hat{x}_k$ and other is estimated position for feature ${}^{E}\hat{p}_F$ respectively.

$$\widetilde{z\prime}_C = H\prime_C\widetilde{x}_k + n\prime_C \tag{22}$$

Next, we apply measurement compression following the method outlined by [49]. This step significantly reduces computational complexity before the EKF update.

E. LiDAR

LGVINS utilizes an adaptation of the direct alignment method from [9] rather than extracting other like LOAM features. This

approach enhances the efficiency and consistency of the update process and introduces a mapping functionality. LiDAR generates 3D point clouds representing the navigation environment. At time t_k, a new point cloud is captured in the LiDAR frame L.

Each point ${}^{L_k}p_F$ is transformed to the local map frame M, where we then identify several neighboring points, ${}^{M}p_{n_i} = \begin{bmatrix} x_{ni} & y_{ni} & z_{ni} \end{bmatrix}^T, i \in 1, ..., m$. We calculate the plane ${}^{M}\prod = \begin{bmatrix} a & b & c \end{bmatrix}^T$, where the neighboring points are located as follows:

$$A_{m\times 3}\begin{Bmatrix} x_{n_1} & y_{n_1} & z_{n_1} \\ \vdots & \vdots & \vdots \\ x_{n_m} & y_{n_m} & z_{n_m} \end{Bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = b_{m\times 1}\begin{Bmatrix} 1 \\ \vdots \\ 1 \end{Bmatrix} \tag{23}$$

The linear least-square solution is determined by ${}^{M}\prod = (A^\intercal A)^{-1}A^\intercal b$. Upon finding the plane...${}^{M}\prod$, we define the following measurement for all planar points, encompassing ${}^{L_k}p_F$ and ${}^{M}p_{n_i}$:

$$z_L := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} = h_L(x_k) = \begin{bmatrix} {}^{M}\Pi^\top p_{n_1} - 1 \\ \vdots \\ {}^{M}\Pi^\top p_{n_m} - 1 \\ {}^{M}\Pi^\top p_F - 1 \end{bmatrix} \tag{24}$$

The computation of ${}^{M}p_F$ involves transforming ${}^{L_k}p_F$ into the map frame utilizing the IMU-LiDAR extrinsic calibration $x_L = ({}^{L}_{I}R,\ {}^{L}p_I)$, the IMU pose $({}^{I_k}_{E}R, {}^{E}p_{I_k})$, and the map pose $({}^{M}_{E}R,\ {}^{E}p_M)$:

$${}^{M}p_F = {}^{M}_{E}R\left({}^{E}p_{I_k} + {}^{E}_{I_k}R\left({}^{I}p_L + {}^{I}_{L}R{}^{L_k}p_F\right) - {}^{E}p_M\right) \tag{25}$$

For the EKF update, we linearize (24) and obtain the subsequent residual:

$$\tilde{z}_L = \begin{bmatrix} -{}^{M}\hat{\Pi}^\top\hat{p}_{n_1} + 1 \\ \vdots \\ -{}^{M}\hat{\Pi}^\top\hat{p}_{n_m} + 1 \\ -{}^{M}\hat{\Pi}^\top\hat{p}_F + 1 \end{bmatrix} \tag{26}$$

$$\approx H_L\tilde{x}_k + H_{\prod}\,{}^{M}\tilde{\prod} + n_L \tag{27}$$

where H_L and $H_{\prod}$ are the Jacobian matrix of $H_L(\cdot)$ in respect to the state x_k and the plane ${}^{M}\prod$; n_L is noise as zero-mean Gaussian. Now we possess subsequent measurement model, which relies solely on the state and is prepared for the EKF update:

$$\tilde{z}\prime_L = H\prime_L\tilde{x}_k + n\prime_L \tag{28}$$

Please note that we also apply measurement compression similarly to MSCKF-based VINS to enhance efficiency even further.

F. GPS

A GPS receiver provide UAV position in a geodetic coordinate frame. These information are commonly converted to a local ENU or NED frame E for GPS based robot navigation purposes. For instance, this conversion can be accomplished by designating the initial measurement location as the reference point or by employing a base station. At time t_k, a GNSS measurement Z_G represents the global position of the receiver ${}^{E}p_{G_k}$. This aspect can be represented using the IMU pose $({}^{I_k}_{E}R, {}^{E}p_{I_k})$ and the extrinsic calibration ${}^{I}p_G$:

$$z_G = h_G(x_k) = {}^{E}p_{G_k} + n_G \tag{29}$$

$$= {}^{E}p_{I_k} + {}_{I_k}^{E}R{}^{I}p_G + n_G \tag{30}$$

where n_G represents white Gaussian noise. Next, we linearize this measurement for the EKF update:

$$\tilde{z}_G := z_G\ -\ h_G(\hat{x}_k) \approx H_G\hat{x}_k + n_G \tag{31}$$

Here H_G represent $h_G(\cdot)$ of the Jacobian matrix corresponding to x_k.

Considering that the measurement contains information about the pose of sensor, Therefore it can be written as $({}^{X}_{E}R, {}^{E}p_X)$, We represent the measurement by incorporating the IMU pose $({}^{I_k}_{E}R, {}^{E}p_{I_k})$ and the corresponding extrinsic calibration $({}^{X}_{I}R, {}^{X}p_I)$:

$$z_X := \begin{bmatrix} {}^{X_k}_{E}\theta \\ {}^{E}p_{X_k} \end{bmatrix} = h_X(x_k) = \begin{bmatrix} log({}^{X}_{I}R{}^{I_k}_{E}R) \\ {}^{E}p_{I_k} + {}_{I_k}^{E}R{}^{I}p_X \end{bmatrix} + n_X \tag{32}$$

Here n_x denotes the zero-mean Gaussian noise. Finally, we again linearize this global-pose measurement for the EKF update:

$$\tilde{z}_X := z_X - h_X(\hat{x}_k) \approx H_X\tilde{x}_k + n_X \tag{33}$$

Here H_x denotes the Jacobian matrix.

G. Coordinate Frames Navigation

The UAV's state estimate requires translation of GPS data by its body frame and correspond coordinate frame. The sensor frame is connected to the sensors, acting as the local reference frame where the sensor's readings are reported. The LGVINS typically includes sensor frame like IMU frame B, camera as C, GPS as G and Lidar frame as L. We designate the IMU frame as the target frame for estimation, referred to as the LGVINS body frame, as shown in Fig. 3. To integrate GPS coordinates, which are ECEF (Earth-Centered, Earth-Fixed) frame, and navigation solution needs to be changed into the NED (North-East-Down) coordinate system or ENU (East-North-Up) frames. ECEF (Earth-Centered, Earth-Fixed) one of the Cartesian coordinate system that remains fixed relative to the Earth, centered at its center of mass.

The x-y plane is aligned with the equatorial plane of earth, where x-axis points towards prime meridian. Furthermore, z-axis is perpendicular to the equatorial plane, directed towards geographic North Pole and y-axis is oriented to maintain in ECEF frame as right-hand coordinated. Here, we utilize the WGS84 realization from ECEF coordinated frame, depicted Fig. 3. To establish a connection between the local and global ECEF frame, and a semi-global called ENU (East-North-Up) is introduced. In this study, we utilized ENU coordinate systems, adhering to the

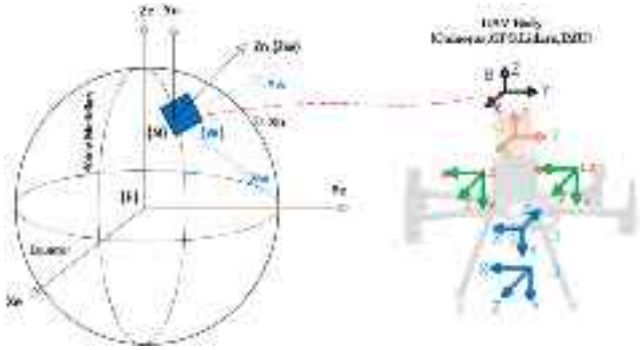

Fig. 3. An illustration of the local world, ECEF and ENU frames w.r.t UAV and sensor body frame. Here N denotes ENU frame, E is the ECEF frame, W for world frame, B is the IMU frame, G denotes the GPS frame, C for the camera frame and L for the lidar frame.

Fig. 4. LGVINS hardware integration setup on a UAV, SBC is a single board onboard computer based on Nvidia NX, front stereo camera which is VGA 640x480 with 15 Hz, IMU and GPS/RTK are dji onboard sensors and two Livox M360 3D LiDAR's.

ISO 8855 standard commonly used in the automotive industry. In the ENU reference frame, the x-axis is oriented east, the y-axis is oriented north, and the z-axis is oriented upward. An ENU frame with a unique origin can be created for any point in the ECEF frame. It is important to note that the alignment of the frame ENU by z-axis corresponds to the direction of gravity. In terms of time frames, GNSS data is documented in the GNSS time system (such as GPS time), whereas VI measurements are timestamped using the local time system. We proceed with the assumption that these two time systems are pre-synchronized and equivalent for the purposes of this research..

The following equation describes the conversion from the ECEF frame to the ENU frame, aligning the local coordinate system with the east, north, and up directions relative to the observer:

$$\begin{bmatrix} x_n \\ y_n \\ z_n \end{bmatrix} = \begin{bmatrix} cos\ \lambda & -sin\ \lambda & 0 \\ -sin\ \Phi\ sin\ \lambda & -sin\ \Phi\ cos\ \lambda & -cos\ \Phi \\ cos\ \Phi\ sin\ \lambda & cos\ \Phi\ cos\ \lambda & -sin\ \Phi \end{bmatrix} \begin{bmatrix} x_e - x_0 \\ y_e - y_0 \\ z_e - z_0 \end{bmatrix} \tag{34}$$

Where, λ is the longitude and Φ is called as the latitude origin of frame.

IV. Experimental Setup

To evaluate the performance of the proposed LGVINS fusion algorithm, we first tested with KAIST URBAN [45] public dataset as well as our own dataset collected using integrated UAV platform, as depicted in Fig. 4. The UAV is outfitted with a range of onboard sensors that have been calibrated before being used. The integrated sensors such as front facing stereo camera with dual VGA 640x480 resolution publishes at 15 Hz along with a GPS/RTK receiver which publishes at 50 Hz. This receiver connects to a differential station via Ethernet, providing both raw GPS observations and RTK ground truth data. Additionally, we integrated a 9-axis 400 Hz IMU module from a DJI drone onto the front facing stereo to align the vision by the ENU coordinate system of GPS. Two Livox M360 3D LiDAR's are connected in front and downward facing in order to acquire maximum number of 3D points. Each of the LiDAR's ROS pointcloud is obtained at 20Hz frequency. Table II. shows each hardware configuration used in the LGVINS UAV.

Moreover, to coordinate and multiple sensor synchronization, data collection, we utilize ROS (Robot Operating System). ROS is an open-source robotics solution tool [50] that facilitates message-passing between processes and provides synchronization utilities. All our datasets have been obtained using an onboard computer based on Nvidia Jetson with GPU support embedded in the proposed UAV platform can be seen in the Fig. 4. The corresponding dataset will be made available public for the community. Three different dataset in rosbag format, which can be accessed through https://github.com/mahammadirfan/LGVINSLGVINS. This repository provide significant insights into the system's performance and function as an important resource for ongoing analysis,updates and research.

The offline calibration of this system comprises three components: estimation of camera's extrinsic/intrinsic parameters and determination of IMU-camera exxtrinsic offset parameters, for which we utilize the well-known **Kalibr** calibration toolbox method [43] to estimate both the intrinsic and extrinsic parameters of the stereo camera. Lastly, 3D LiDAR-IMU offset calibration, we use latest popular toolbox **LI-Init** by [44] which serves as a real-time initialization method for LiDAR-inertial systems. This calibrates the temporal offset and extrinsic parameter between LiDARs and IMUs. To check the robustness of the algorithm under different circumstances, we collected multiple datasets in three scenario including both Handheld and in flight UAV dataset. The datasets are referred as *UL Car Bridge, UL living bridge (Handheld UAV), UL Car Parking*. Fig. 5. represents the experimental environment in different scenario during the data collection and the information of the dataset used is presented in Table III. The results presented in this paper were performed at the University of Limerick Campus of the CRIS Lab research group.

V. Results and Discussion

We perform and analyse the proposed LGVINS with public and our real-world experiment datasets to verify the performance. In this section, we conduct a comparison between LGVINS with the open-source algorithms such as FASTLIO2 [9], VINS-FUSION with GPS [6], VINS-FUSION

TABLE III
DATASET DETAILS

Dataset	Max. Distance (m)	Time (sec)	Max. Velocity (m/s)	GPS-RTK Satellite	Environment
KAIST URBAN [45]	11420	1866	NA	NA	Highly Complex Urban Environment
UL Car Bridge	672	505	2	37 max, 11 min	Structured,Agile Movements
Handheld UAV	533	538	1	38 max, 14 min	Structure-less, Unstable-features
UL Parking	1348	674	3	37 max, 32 min	Open-air, Tree-shadowed, Sunlight

Fig. 5. The experimental environment in different scenario during the data collection.

without GPS [7] and Lvio-Fusion [22] to demonstrate its effectiveness. Furthermore, we perform in detail comparison of LGVINS algorithm with its different sensor combination. We only fuse the GPS position when it is in good health, by subscribing to the DJI ROS SDK, we can monitor the number of available GPS satellites during the flight. Therefore, we have set a minimum threshold of 11 satellite signals to ensure good GPS health. LGVINS automatically use GPS position as per window limit in order to fuse only optimum true position from GPS.

All the experiments have been performed without loop closure scenario to validate the proposed LGVINS global consistency. All experiments in this section were performed using a laptop equipped with an Intel Core(TM) i7-10750H CPU @ 2.60 GHz processor running at 3.7 GHz and 32 GB of RAM. We conducted a thorough numerical analysis to showcase the consistency of proposed system and evaluated a detailed comparison with four state-of-the-art methods: VINS-FUSION (Stereo-inertial) [7] Fast-LIO2 [9], VINSFUSION (Stereo-inertial-gps) [7] and Lvio-Fusion [22]. Additionally, we opted exclusively for methods with open-source code availability and are compatible with our developed LGVINS hardware platform.

A. Evaluation Metrics

The mean error of the proposed multi-sensor fusion system is a crucial validation metric which will validate the LGVINS state estimation trajectory and poses to their actual positions as obtained from the ground truth data RTK. The metric is considered as the important indicator of the precision and accuracy of the system. The most common formula such as RMSE are used to calculate systems mean error. RMSE can be computed by taking the root mean square and squared mean difference between the estimated and ground truth values. This gives the metric which represents the average magnitude of the error correspond to the position estimate and the reference RTK position. The formula for RMSE calculation is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{35}$$

where;

- y_i is the actual value,
- n is the no. of observations,
- $\hat{y}_i$ is the predicted value,
- $\sum$ represents the summation over all observations.

We utilize Absolute Pose Error (APE) as our accuracy metric given by evo [51]. APE measures accuracy by directly comparing the corresponding poses of the robot state trajectory and the ground truth at identical timestamps. Then the whole trajectory is unified for analysis, which is helpful for assessing the global consistency of UAV trajectory.

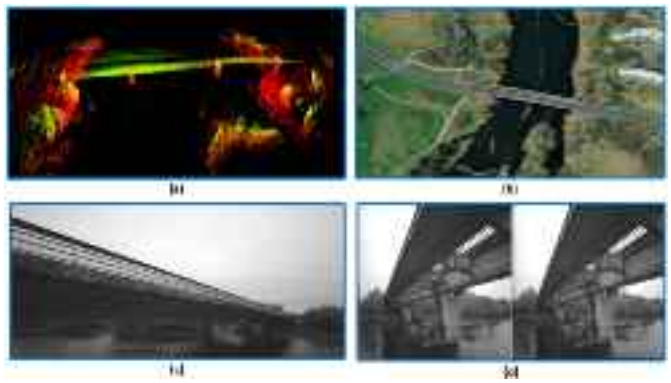

Fig. 6. The scenario of UL car bridge dataset. Panels (a) and (b) represent to the datasets 3D point cloud map and google map, while panels (c) and (d) are the bridge view from UAV.

B. UL car Bridge Dataset

The experiment takes place at the university of limerick car bridge, flying the drone under and above of the bridge environment is displayed in the Fig. 6. The reference for LGVINS is collected by RTK. The UAV flight was done manually by a trained pilot. We have performed this experiment without

TABLE IV
SUMMARY OF FINDINGS FROM ACCURACY EVALUATION ON UL CAR BRIDGE DATASET IN METERS

Method	Max	Mean	Median	Min	RMSE
LGVINS(L+I+S+G)	1.296882	0.384646	0.291391	0.012473	**0.473501**
LGVINS(L+I+G)	2.28875	0.590732	0.503384	0.148098	0.664017
LGVINS(L+I+S)	3.908775	1.270052	0.974482	0.092406	1.516019
LGVINS(L+I)	3.35426	0.967264	0.786011	0.176427	1.12622
FASTLIO (L+I)	2.485825	0.630901	0.438535	0.032907	0.805311
VINS FUSION (S+I+G)	2.284776	0.421503	0.253154	0.006963	0.683928
VINS FUSION (S+I)	22.006355	7.392873	5.851291	1.614134	8.892644

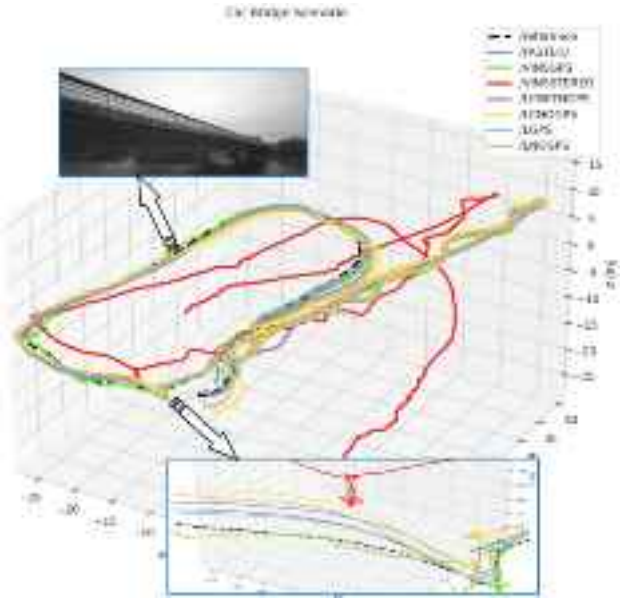

Fig. 7. Trajectory plots of various methods on UAV car bridge dataset.

loop closing condition to validate the global consistency of the LGVINS. The dataset environment involve acute ambient illumination changes, agile movements and GPS degraded, as depicted in Fig. 8, which are challenges for the VIO and LO methods, where proposed LGVINS fusion method performed well as compare with other SOTA methods. During the experiment, most satellites consistently maintain a strong signal lock, ensuring that the RTK status remains fixed throughout the entire route. However, while flying under the bridge GPS satellite number reached minimum to (11) satellite signal which resulted in unstable GPS signal in challenging environment.

The proposed LGVINS fuses GPS position only if it is in good health condition (enough satellite signal), otherwise LGVINS deselect GPS position while only fuses other sensor combination for state estimation. This experiment test the global consistency of proposed LGVINS, in a challenging unstable and noisy gps signal under the bridge environment, affecting the local UAV state and smoothness of the trajectory. Figs. 7 and 8 represents the LGVINS obtained trajectories along with other methods, Table IV lists the RMSE of each method. This experiment demonstrates that our system can achieve global consistency to eliminate single sensor drifts while maintaining local smoothness even under noisy GPS conditions.

Fig. 8 depicts the trajectory error plots and comparison generated by each methods. Trajectory error for vinsfusion [7] with stereo-inertial is much more as compare to others. Vinsfusion with no GPS [7] performs poorly because estimating UAV pose solely through visual odometry is difficult in challenging environment like bridge inspection case and the motion of the UAV also had an impact, as the drone occasionally underwent almost purely rotational movements which can be seen in the Fig. 7 Fastlio2 [9] was also resulting in significant trajectory drifts compare to ground truth as it only based with lidar-inertial odometry. In contrast, LGVINS consisting of all the sensors achieved the best performance.

When compare the LGVINS sensor combination, LGVINS (Lidar+Imu+Stereo+GPS) performs well in bridge case scenario. However, LGVINS (Lidar+Imu+GPS) and LGVINS (Lidar+Imu) comes in second and third place showing RMSE error of 0.66 and 1.12 respectively. Hence, LGVINS with GPS fusion perform the best in compare without GPS sensor combination. Table IV provides a detailed evaluation of the translational APE using the RMSE. The performance metric highlights the percentage improvement in APE compared to other methods. Table III. shows that proposed LGVINS outperforms all other in UL car bridge dataset. Figs. 9 and 10 represents the absolute position error of x,y,z and roll,pitch,yaw showing plots of various methods on UAV car bridge dataset. Fig. 11 is the estimated trajectory plot of the UAV car bridge dataset aligned well with Google map. Fig. 12 depicts Box plot of overall RPE of the seven strategies of UAV car bridge dataset showing LGVINS performed well compared with other SOTA methods.

C. *Handheld UAV Living Bridge Dataset*

In this experiment, we utilized the same UAV custom-designed sensor suite to showcase the capabilities of our framework. The RTK position is treated as ground truth while with good GPS signal recorded throughout the experiment. The dataset has been collected using handheld UAV method while walking around outdoor which encompass environments with image degradation and structure-less surroundings, dynamic targets and unstable features which is challenging for vision-based and lidar-based methods.

The experiment has been performed without loop closure scenario to validate the consistency of the proposed LGVINS approach and compared with RTK as reference data. Fig. 13 shows an overview of the environment. This experiment was conducted in handheld UAV mode to eliminate the noise generated during flight missions and validate the performance of the proposed LGVINS in such scenarios. We performed state estimation using LGVINS different sensor combinations as well

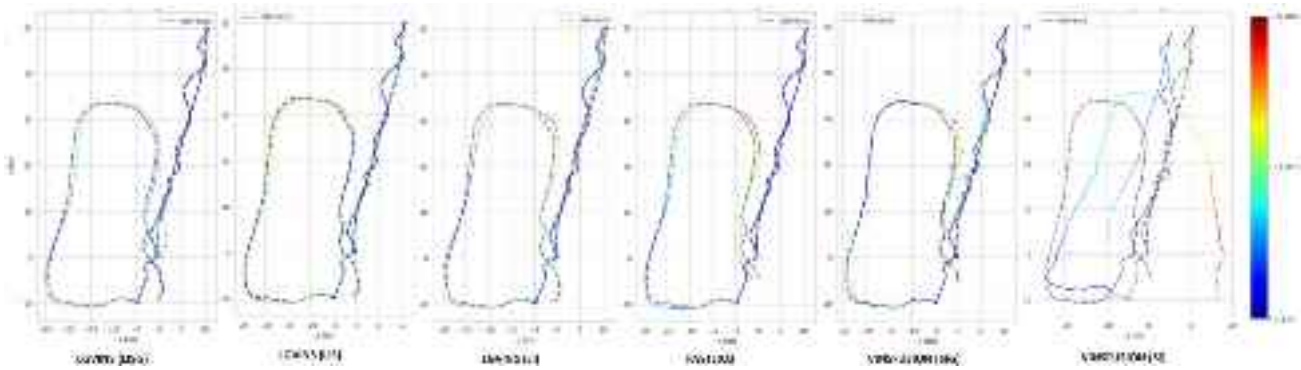

Fig. 8. Trajectory error plot for each method. Here, reference, L,I,S and G denotes RTK, lidar, IMU, stereo and GPS respectively.

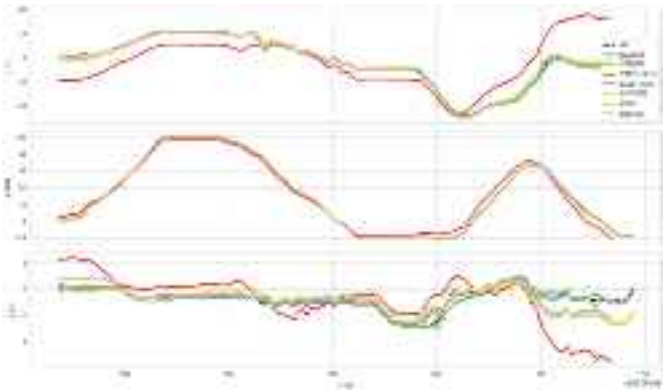

Fig. 9. Absolute position error of x,y,z axis showing plots of various methods on UAV car bridge dataset.

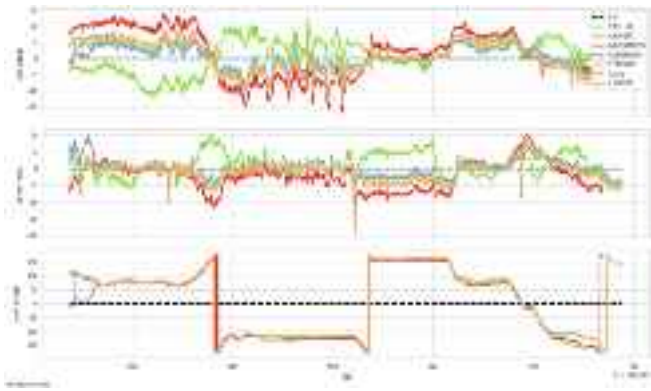

Fig. 10. Absolute position error of roll,yaw,pitch showing plots of various methods on UAV car bridge dataset.

Fig. 11. Estimated trajectory of the UAV car bridge dataset aligned well with Google map.

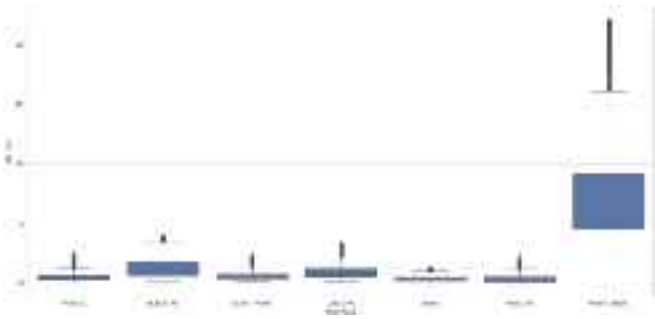

Fig. 12. Box plot of overall APE of the seven strategies of UAV car bridge dataset.

as other SOTA like VINSFUSION [7] and FASTLIO2 [9]. The plot shown in Fig. 14 trajectory obtained using different methods, and the trajectory error comparison of various methods is shown in Fig. 18. Table V illustrates the RMSE values for the experiment of each method. Figs. 15 and 16 represents the absolute position error of x,y,z and roll,pitch,yaw showing plots of various methods on handheld UAV living bridge dataset.

It can be observed from the dataset experiment, significant position drifts were observed in the stereo-imu only scenario. However, the accuracy improved substantially with the assistance of the Lidar or GPS or combination of all. VINS-Fusion's error plot continues to rise due to accumulated drift, whereas LGVINS maintains a constant and smooth trajectory approach. VINSFUSION [7] and Fast-LIO2 [9] do not accurately align with the ground truth, whereas LGVINS demonstrated strong performance across all three combinations.

TABLE V
SUMMARY OF FINDINGS FROM ACCURACY EVALUATION ON HANDHELD UAV LIVING BRIDGE DATASET IN METERS

Method	Max	Mean	Median	Min	RMSE
LGVINS(L+S+I+G)	3.482527	0.635567	0.550702	0.061366	**0.788274**
LGVINS(L+S+I)	5.858725	1.122269	0.858668	0.166064	1.532994
LGVINS(S+I+G)	2.998805	0.796859	0.699599	0.090758	0.921967
FASTLIO2 (L+I)	11.488647	6.466611	5.564047	2.445243	6.830505
VINS FUSION (S+I+G)	10.320782	6.479688	6.737983	1.811507	6.846302
VINS FUSION (S+I)	14.391085	6.772864	6.961714	2.853421	7.18024

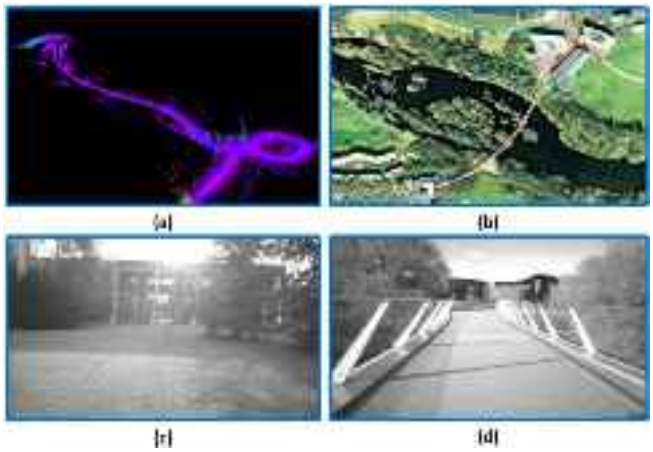

Fig. 13. The environments of the collected Handheld UAV dataset. Panels (a) and (b) represent the datasets 3D point cloud map and google map, while panels (c) and (d) are the drone view.

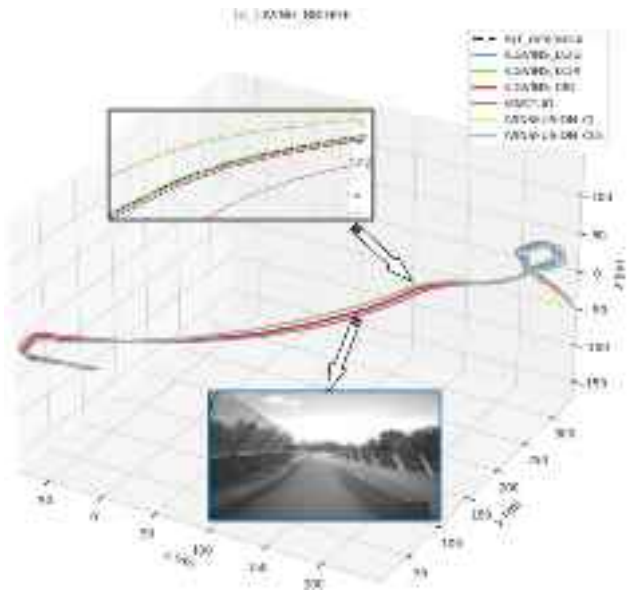

Fig. 14. Trajectory plots of various methods on handheld UAV living bridge dataset.

The experiment performed in varying illumination and structure-less environment, presenting challenges for tracking multiple visual features and matching geometric features. Nevertheless, LGVINS method continue to perform well, as they

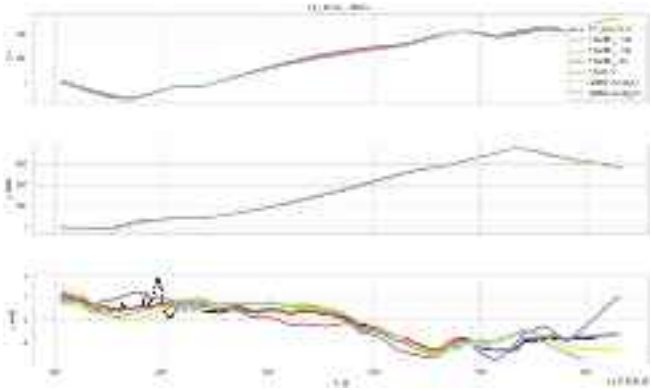

Fig. 15. Relative position error of x,y,z axis showing plots of various methods on handheld UAV living bridge dataset.

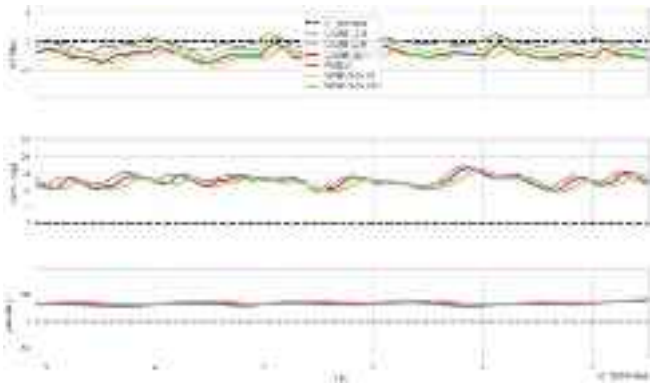

Fig. 16. Relative position error of roll,yaw,pitch showing plots of various methods on handheld UAV living bridge dataset.

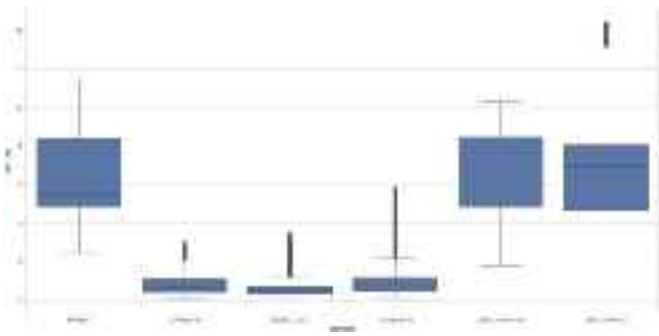

Fig. 17. Box plot of overall APE of the six strategies of handheld UAV living bridge dataset.

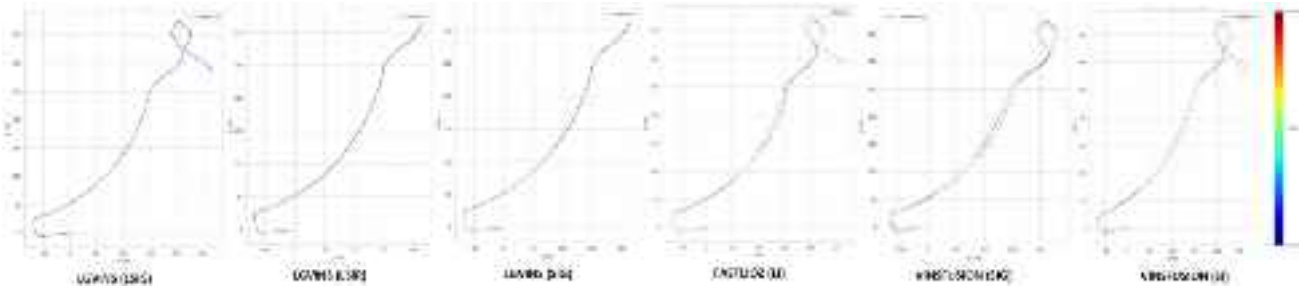

Fig. 18. Trajectory error of various methods on handheld UAV living bridge dataset.

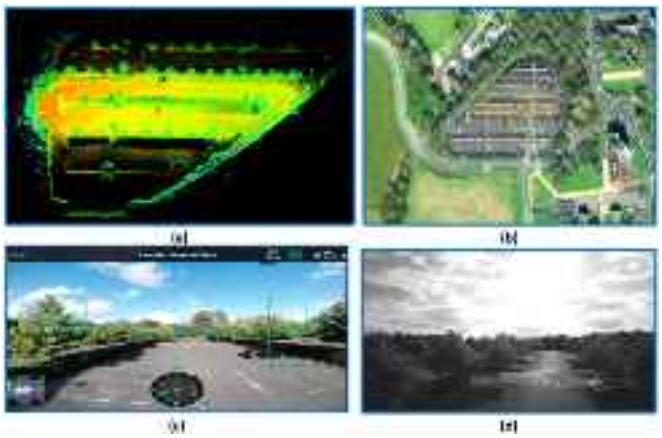

Fig. 19. The scenario of the UL car parking collected dataset. Panels (a) and (b) represent the 3D point cloud map and google map, while panels (c) and (d) are the drone view.

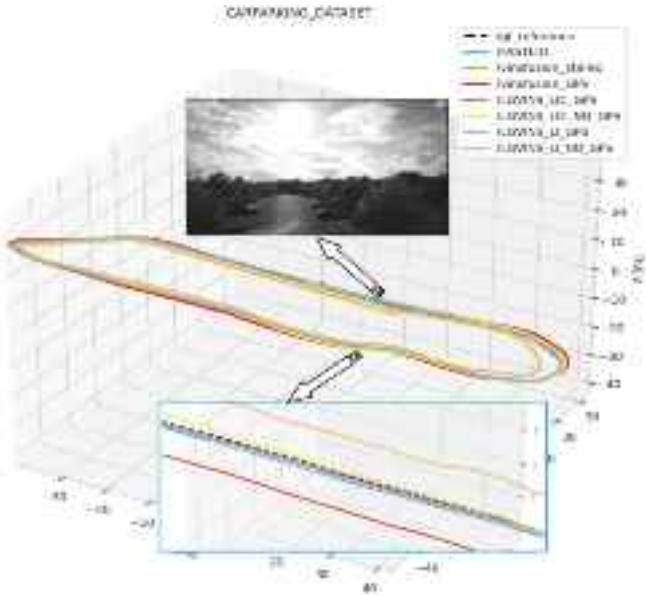

Fig. 20. Trajectory plots of various methods on car parking dataset.

effectively fuse data from camera, IMU, GPS and 3D LiDAR or combination of any. The RMSE values of each LGVINS combination are less compare to all the SOTA methods which is displayed in the table. V. The trajectories displayed in Figs. 14 and 18 illustrate that following dataset, the trajectory obtained by LGVINS(LSIG), LGVINS(SIG) and LGVINS(LSI) show a significant advantage over those produced by the VINSFUSION [7] and Fast-LIO2 [9] methods. These results suggest that the LGVINS method improves UAV state estimation accuracy compared to alternative algorithms and shows greater robustness due to its capabilities. Fig. 17 shows that proposed method outperforms all other in Handheld UAV living bridge dataset.

D. Car Parking Dataset

The car parking dataset environment includes open-air, tree-shadowed environment and acute ambient illumination changing scenario can be seen the Fig. 19. In this experiment, Evaluation of the LGVINS approach has been conducted in an large-scale outdoor, open-air environment. The experiment has been performed without loop closure condition to validate the consistency of proposed LGVINS. In this car parking dataset, the UAV navigate in a vast, open lawn with mostly few trees visible in the distance with very bright and sunny environment, which is challenging for stereo based approaches and most odometry fails and cannot be effectively addressed by vision-based or LiDAR-based methods individually. Furthermore, While flying above the parking area, Most of the features detected by the LiDAR sensor were located at ground level, leading to degraded motion estimation. Therefore, it can be observe from the obtained trajectory plot with respect to the ground truth RTK, FASTLO2 [9] produces a significant high errors due to Lidar degradation.

Similarly, VINSFUSION (stereo-inertial) performs worst as compare to all others whereas noticeable position drifts occurred in the VINSFUSION (stereo-imu-GPS) scenario. Additionally, we observed that in the car parking open lawn dataset, the features extracted from LiDAR were noticeably sparse, resulting in reduced performance of the Lidar based algorithms like FASTLIO2 [9]. Incontrast, proposed LGVINS uses stereo,imu,Lidars and GPS, hence producing an improved UAV state estimation and smooth trajectory in all challenging scenarios. In all four sensor combination LGVINS performs better and produces smooth trajectory state as compare to VINSFUSION [7] and FASTLIO2 [9] can be seen in the Fig. 20 and box plot in Fig. 23 and Table VI illustrate RMSE for each method. We have compared all the sensor combination for

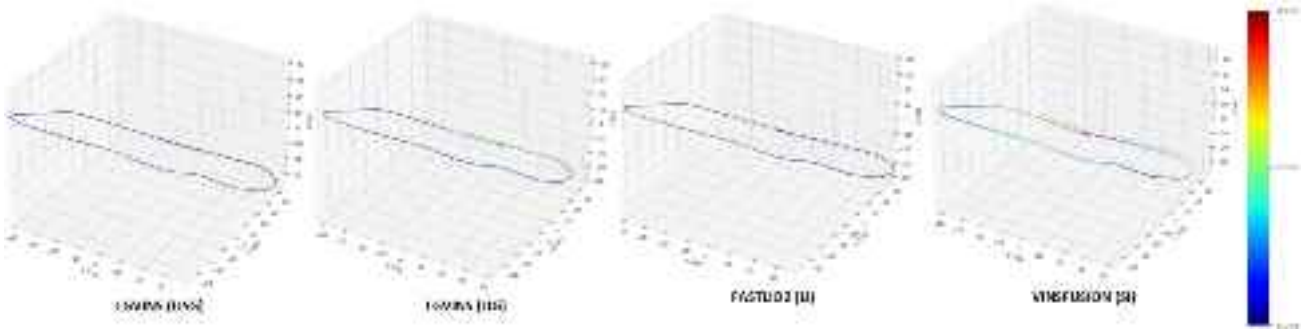

Fig. 21. Trajectory Error plots of various methods on car parking dataset.

TABLE VI
SUMMARY OF FINDINGS FROM ACCURACY EVALUATION ON UAV CAR PARKING DATASET IN METERS

Method	Max	Mean	Median	Min	RMSE
LGVINS(L+I+S+G)	0.07824	0.027466	0.027798	0.002795	**0.028609**
LGVINS(L+I+S)	0.461637	0.02703	0.027718	0.000572	0.028949
LGVINS(L+I+G)	2.74426	0.028032	0.027956	0.002519	0.037316
LGVINS(L+I)	1.471649	0.028056	0.028055	0.002295	0.030999
FASTLIO2 (L+I)	0.934567	0.173908	0.144395	0.000972	0.221818
VINS FUSION (S+I+G)	1.285993	0.193174	0.148126	0.002307	0.255256
VINS FUSION (S+I)	1.285993	0.193382	0.148126	0.002307	0.255535

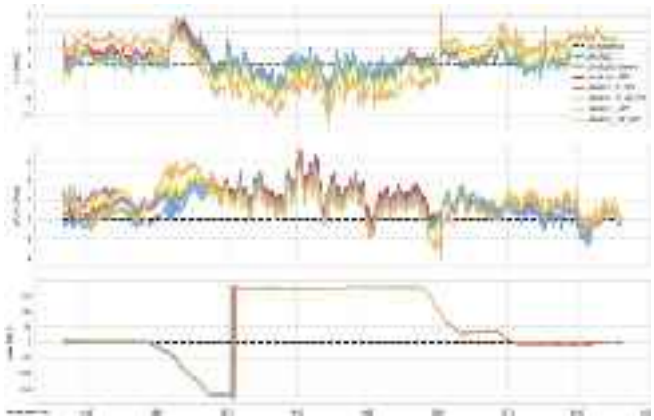

Fig. 22. Absolute position error of roll, yaw, pitch showing plots of various methods on car parking dataset.

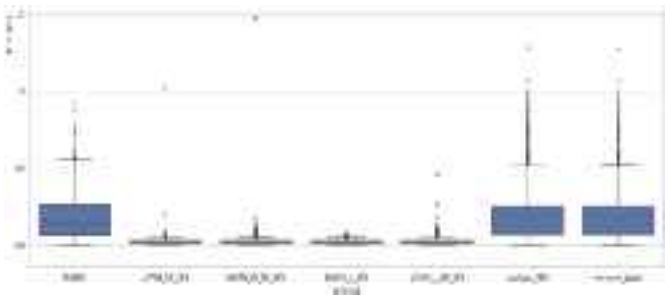

Fig. 23. Box plot of overall RPE of the seven strategies of car parking dataset.

LGVINS methods, and found that with the combination of fusing all sensor state, the accuracy improves in LGVINS hence the estimated global pose produces smooth and consistent UAV state estimation. Figs. 22 and 21 represents the absolute position error of roll,pitch,yaw and trajectory error showing plots of various methods on car parking dataset respectively. Fig. 23 shows that our method outperforms all other in UL car parking dataset.

TABLE VII
SUMMARY OF FINDINGS FROM ACCURACY EVALUATION ON KAIST URBAN DATASET IN METERS

Method	Mean	RMSE
LGVINS (L+I+S+G)	5.26	**5.79**
LVIO (L+I+S+G)	5.79	6.31
VINS FUSION (S+I+G)	13.28	14.16
VINS FUSION (S+I)	141.10	153.12
FASTLIO (L+I)	306.48	319.08

E. KAIST Urban Dataset

The dataset [45] has been collected using a ground vehicle in highly complex downtown environment including metropolis areas, complex buildings and residential areas with movable features. The presence of moving vehicles and pedestrians further complicates the state estimation, making multi-sensor fusion and global constraints even more critical. The device of kaist urban dataset sensor includes a 100 Hz Xsens MTi-300 IMU, two 10 Hz Pointgrey Flea3 stereo camera, two 10 Hz Velodyne VLP-16 channel LiDARs. Kaist Urban Sequence 39 which is the largest and most complex dataset (11 km) is used to evaluate the proposed LGVINS.

The results of APE is summarize in Table VII. The efficiency of the proposed approach is ultimately demonstrated by a numerical study, using mean and RMSE. Based on different combination of sensors are tested along with SOTA and the resulting error trajectories of each algorithm are shown in Fig. 24. Fig. 26 depicts the obtained APE trajectory of proposed LGVINS in kaist urban dataset. Overall, Fastlio2 (Lidars+IMU) and Vinsfusion (stereo+IMU) showed scale issue and combining additional

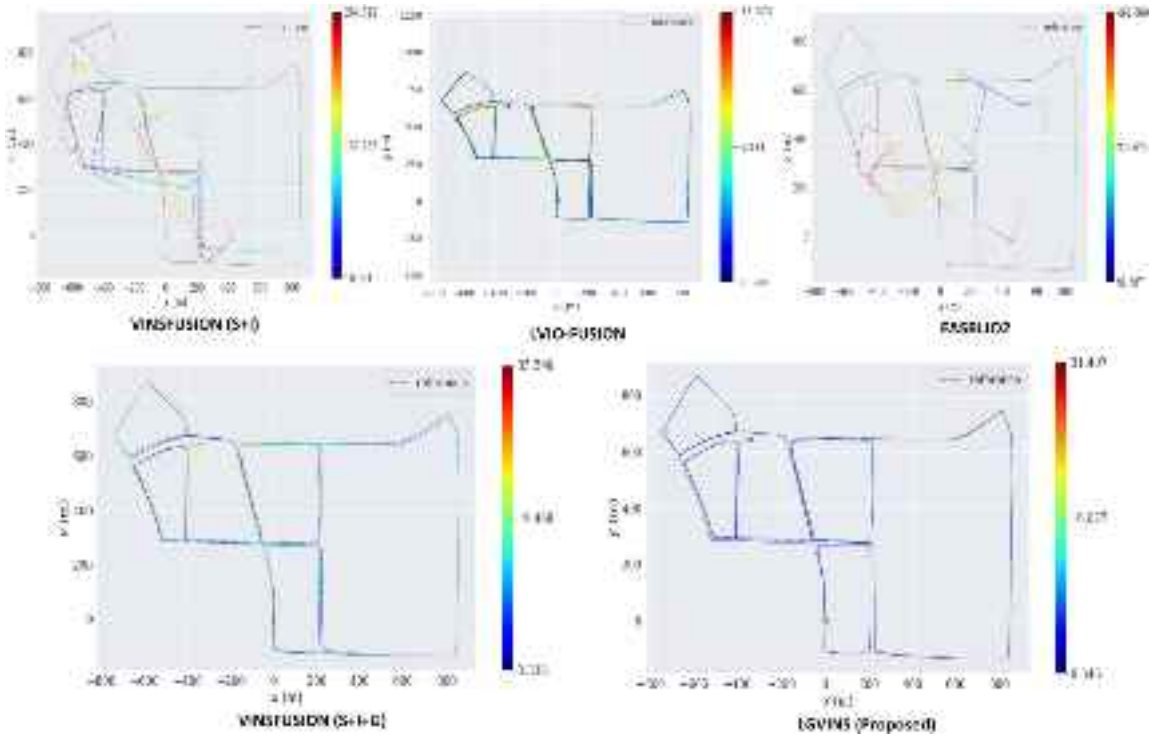

Fig. 24. Trajectory error plot for each method on KAIST urban dataset. Here L,I,S and G denotes Lidar, IMU, stereo and GPS respectively.

Fig. 25. The scenario of KAIST urban dataset. Panels (a) and (b) represent to the datasets platform and google map, while panels (c) and (d) are the view from KAIST sensor.

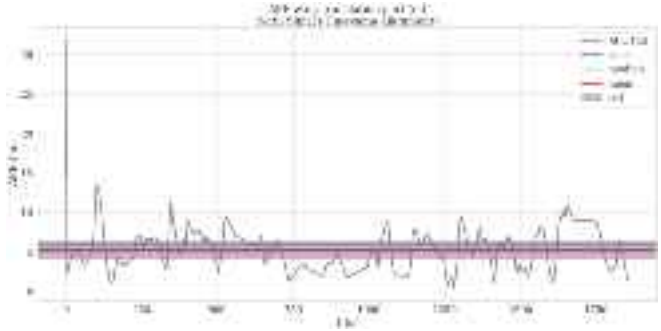

Fig. 26. Proposed LGVINS on KAIST urban dataset.

sensor was able to solve the problem. Note that Fastlio2 can fuse only one LiDAR sensor which showed poor performance on the KAIST dataset due to its limited overlapping points between the LiDAR scans. The key issue for poor performance is likely the lack of overlapping point clouds between the map and the new scan, resulting in a poorly defined ICP problem. The Vinsfusion methods using stereo-inertial only initialized successfully, but their performances were largely drifted by the dynamic objects present in Kaist seq 39. (see an exemplary case in Fig. 25. VINS-Fusion with adding a GPS sensor combination and Lvio-Fusion using all sensors failed to provide consistent estimates on kaist datasets, even when GPS global position was available, due to their loosely coupled approach. This occurs because the VIO sub-system of VINS-Fusion and the LIO sub-system of Lvio-Fusion (stereo+Lidar+IMU+GPS) struggled with initialization and dynamic objects present in kaist dataset, which led to inconsistent odometry data being fed into the pose graph. This caused global estimation to become highly unreliable, ultimately resulting in worse performance than relying solely on GPS data. In contrast, proposed LGVINS-Fusion integrated GPS data with other sensors in a tightly coupled manner, enabling the state estimation to effectively limit its drift, resulting in globally accurate and locally precise performance.

VI. Conclusion

The multi sensor fusion algorithms to improve the UAV/UGV state estimation have been proposed in this work. The accuracy and effectiveness of the proposed LGVINS were validated through real-world experiments and public dataset, which included challenging dataset collection using our UAV equipped

with an integrated hardware platform across a range of scenarios. Experiments have demonstrated that the proposed fusion technique can optimize UAV state estimation, resulting in an improved, robust and accurate state in GPS degraded environment under multiple use case. The aiding of an GPS along with onboard IMU fused with stereo/LiDAR odometry data can minimize growing GPS errors and can bridge even long GPS signal outages. Experiments conducted across diverse environments show that the proposed LGVINS method consistently outperforms other methods such as VINSFUSION (stereo-inertial, VINSFUSION (stereo-inertial-GPS) and FastLIO2, algorithms and achieve high state estimation accuracy and robustness in challenging environments, including those with bridge case, open air or poor lighting or no structure. Furthermore, Proposed LGVINS demonstrated and evaluated on KAIST Urban challenging dataset which outperform SOTA methods.

Future work could investigate using even larger, more diverse dataset adding sensors like radar sensor. However, the accuracy of IMU-based odometry can degrade over time due to the accumulation of errors. In contrast, radar sensors are emerging as a promising alternative for state estimation and SLAM (Simultaneous Localization and Mapping) because of their robustness against adverse ambient illumination condition and complex environmental structures. Despite this potential, the challenge of identifying reliable RADAR/LiDAR sensors remains an unresolved issue. Therefore, we would like to investigate LGVINS by adding more sensors in future.

REFERENCES

[1] C. Cadena et al., "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.

[2] T. Wang, Y. Su, S. Shao, C. Yao, and Z. Wang, "GR-Fusion: Multi-sensor fusion slam for ground robots with high robustness and low drift," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 5440–5447.

[3] S. Dorafshan and M. Maguire, "Bridge inspection: Human performance, unmanned aerial systems and automation," *J. Civil Struct. Health Monit.*, vol. 8, pp. 443–476, 2018.

[4] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," 2019, *arXiv:1901.03638*.

[5] M. Rossi, P. Trslić, S. Sivčev, J. Riordan, D. Toal, and G. Dooly, "Real-time underwater stereofusion," *Sensors*, vol. 18, no. 11, 2018, Art. no. 3936.

[6] S. Cao, X. Lu, and S. Shen, "GVINS: Tightly coupled GNSS-visual-inertial fusion for smooth and consistent state estimation," *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2004–2021, Aug. 2022.

[7] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.

[8] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A research platform for visual-inertial estimation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 4666–4672.

[9] W. Xu, Y. Cai, D. He, J. Lin, and F. Zhang, "FAST-LIO2: Fast direct LiDAR-inertial odometry," *IEEE Trans. Robot.*, vol. 38, no. 4, pp. 2053–2073, Aug. 2022.

[10] G. Huang, "Visual-inertial navigation: A concise review," in *Proc. Int. Conf. Robot. Automat.*, 2019, pp. 9572–9582.

[11] K. Ebadi et al., "Present and future of SLAM in extreme environments: The DARPA SubT challenge," *IEEE Trans. Robot.*, vol. 40, pp. 936–959, 2024.

[12] A. R. Sahili et al., "A survey of visual SLAM methods," *IEEE Access*, vol. 11, pp. 139643–139677, 2023.

[13] N. J. Abu-Alrub and N. A. Rawashdeh, "Radar odometry for autonomous ground vehicles: A survey of methods and datasets," *IEEE Trans. Intell. Veh.*, vol. 9, no. 3, pp. 4275–4291, Mar. 2024.

[14] D. Lee, M. Jung, W. Yang, and A. Kim, "LiDAR odometry survey: Recent advancements and remaining challenges," *Intell. Serv. Robot.*, vol. 17, pp. 95–118, 2024.

[15] M. Irfan, K. Kishore, and V. A. Chhabra, "Smart vehicle management system using Internet of Vehicles (IoV)," in *Proc. Int. Conf. Adv. Comput. Appl.*, 2022, pp. 73–83.

[16] M. Irfan, S. Dalai, K. Kishore, S. Singh, and S. A. Akbar, "Vision-based guidance and navigation for autonomous MAV in indoor environment," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol.*, 2020, pp. 1–5.

[17] S. Singh et al., "CACLA-based local path planner for drones navigating unknown indoor corridors," *IEEE Intell. Syst.*, vol. 37, no. 5, pp. 32–41, Sep./Oct. 2022.

[18] W. Lee, P. Geneva, Y. Yang, and G. Huang, "Tightly-coupled GNSS-aided visual-inertial localization," in *Proc. Int. Conf. Robot. Automat.*, 2022, pp. 9484–9491.

[19] X. Meng, H. Wang, and B. Liu, "A robust vehicle localization approach based on GNSS/IMU/DMI/LiDAR sensor fusion for autonomous vehicles," *Sensors*, vol. 17, no. 9, 2017, Art. no. 2140.

[20] K. Hausman, S. Weiss, R. Brockers, L. Matthies, and G. S. Sukhatme, "Self-calibrating multi-sensor fusion with probabilistic measurement validation for seamless sensor switching on a UAV," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2016, pp. 4289–4296.

[21] V. Kubelka, L. Oswald, F. Pomerleau, F. Colas, T. Svoboda, and M. Reinstein, "Robust data fusion of multimodal sensory information for mobile robots," *J. Field Robot.*, vol. 32, no. 4, pp. 447–473, 2015.

[22] Y. Jia et al., "Lvio-fusion: A self-adaptive multi-sensor fusion SLAM framework using actor-critic method," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 286–293.

[23] T.-M. Nguyen, S. Yuan, M. Cao, T. H. Nguyen, and L. Xie, "Viral SLAM: Tightly coupled camera-IMU-UWB-LiDAR SLAM," 2021, *arXiv:2105.03296*.

[24] S. Leutenegger, A. Forster, P. Furgale, P. Gohl, and S. Lynen, "OKVIS: Open keyframe-based visual-inertial SLAM (RoS version)," 2016.

[25] Q. Li et al., "LO-Net: Deep real-time LiDAR odometry," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8473–8482.

[26] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-DoF camera relocalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2938–2946.

[27] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, 2020, Art. no. 132306.

[28] Q. Jia, G. Luo, Q. Yuan, J. Li, C. Shao, and Z. Chen, "MS-transformer: Masked and sparse transformer for point cloud registration," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2023, pp. 1375–1381.

[29] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. 2007 IEEE Int. Conf. Robot. Automat.*, 2007, pp. 3565–3572.

[30] L. Han, Y. Lin, G. Du, and S. Lian, "Deepvio: Self-supervised deep learning of monocular visual inertial odometry using 3D geometric constraints," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2019, pp. 6906–6913.

[31] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *Proc. IEEE/RSJ Int. Conf. Intell. robots Syst.*, 2015, pp. 298–304.

[32] J. Simanek, V. Kubelka, and M. Reinstein, "Improving multi-modal data fusion by anomaly detection," *Auton. Robots*, vol. 39, no. 2, pp. 139–154, 2015.

[33] J. K. Suhr, J. Jang, D. Min, and H. G. Jung, "Sensor fusion-based low-cost vehicle localization system for complex urban environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1078–1086, May 2017.

[34] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to MAV navigation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013, pp. 3923–3929.

[35] C. Merfels and C. Stachniss, "Pose fusion with chain pose graphs for automated driving," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 3116–3123.

[36] R. Mascaro, L. Teixeira, T. Hinzmann, R. Siegwart, and M. Chli, "GOMSF: Graph-optimization based multi-sensor fusion for robust UAV pose estimation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 1421–1428.

[37] S. Han, F. Deng, T. Li, and H. Pei, "Tightly coupled optimization-based GPS-visual-inertial odometry with online calibration and initialization," 2022, *arXiv:2203.02677*.

[38] B. Dong and K. Zhang, "A tightly coupled visual-inertial GNSS state estimator based on point-line feature," *Sensors*, vol. 22, no. 9, 2022, Art. no. 3391.

[39] J. Liu, W. Gao, and Z. Hu, "Optimization-based visual-inertial SLAM tightly coupled with raw GNSS measurements," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2021, pp. 11612–11618.
[40] W. Lee, P. Geneva, C. Chen, and G. Huang, "Mins: Efficient and robust multisensor-aided inertial navigation system," 2023, *arXiv:2309.15390*.
[41] S. Zhao, H. Zhang, P. Wang, L. Nogueira, and S. Scherer, "Super odometry: IMU-centric LiDAR-visual-inertial estimator for challenging environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 8729–8736.
[42] J. Lin and F. Zhang, "R 3 live: A robust, real-time, RGB-colored, LiDAR-inertial-visual tightly-coupled state estimation and mapping package," in *Proc. Int. Conf. Robot. Automat.*, 2022, pp. 10672–10678.
[43] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending Kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2016, pp. 4304–4311.
[44] F. Zhu, Y. Ren, and F. Zhang, "Robust real-time LiDAR-inertial initialization," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 3948–3955.
[45] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse urban environments," *Int. J. Robot. Res.*, vol. 38, no. 6, pp. 642–657, 2019.
[46] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," Dept. Comput. Sci. & Eng., Univ. Minnesota, Minneapolis, MN, USA, Tech. Rep. 2005-002, 2005.
[47] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Observability-constrained vision-aided inertial navigation," Dept. of Comput. Sci. & Eng., MARS Lab, Univ. of Minnesota, Minneapolis, MN, USA, Tech. Rep. 2012-001, 2012.
[48] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1335–1340, Aug. 2006.
[49] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.
[50] A. Koubâa et al. *Robot Operating System (ROS)*, vol. 1. Berlin, Germany: Springer, 2017.
[51] M. Grupp, "EVO: Python package for the evaluation of odometry and SLAM," 2017. [Online]. Available: https://github.com/MichaelGrupp/evo

Mahammad Irfan (Graduate Student Member, IEEE) received Bachelor of Technology degree in electrical and electronics engineering from S'O'A University, Bhubaneswar, Odisha, in 2015, India, and the Master of Technology degree in electronics engineering from the Punjab Technical University, Punjab, India, in 2018. He is currently working toward Ph.D. degree in electronics and computer engineering with the University of Limerick, Limerick, Ireland. He is also a Member with the Centre for Robotics and Intelligent Systems (CRIS). His research interests include SLAM system, UAV State Estimation, Sensor Fusion, and autonomous UAV navigation and perceptiom.

Sagar Dalai (Graduate Student Member, IEEE) is currently working toward the Ph.D. degree with the Centre of Robotics and Intelligent Systems (CRIS) Group, University of Limerick, Limerick, Ireland. He received the Bachelor of Technology degree in electronics and telecommunication from the Biju Patnaik University of Technology (BPUT), Odisha, India. His research interests include hyperspectral image processing, simultaneous localization and mapping (SLAM), and path planning algorithms, leveraging advanced imaging, AI techniques to enhance environmental monitoring, and autonomous systems.

Petar Trslic received the B.Sc. and M.Sc. degrees in mechanical engineering from the University of Zagreb, Zagreb, Croatia, and the Ph.D. degree from the University of Limerick (UL), Limerick, Ireland, in March 2020. He is currently a Postdoctoral Researcher with CRIS, UL. During the Ph.D. degree, he was working on the development of advanced monitoring, control systems and onboard robotics for the automation of MRE inspection, repair, and maintenance.

Matheus C. Santos (Member, IEEE) received the bachelor's degree in computer engineering and the master's degree in electrical engineering from the Federal University of Sergipe (UFS), State of Sergipe, Brazil, in 2019 and 2021, respectively. He is currently working toward the Ph.D. degree in electrical engineering with the University of Limerick. He is also a member of the Centre of Robotics and Intelligent Systems (CRIS) and the Robotics Research Group (GPR), UFS.

James Riordan received the B.Eng. degree in electronic engineering specialising in aircraft simulation systems, and the Ph.D. degree in real-time processing of acoustic signals from the University of Limerick, Limerick, Ireland, respectively. He is currently a Full Professor with the University of the West of Scotland, Glasgow, U.K., where he is also the Director with Drone Systems Laboratory. He is a Principal Investigator of multiple research projects funded by the European Commission and U.K. Research and Innovation. His research interests include artificial intelligence, computer vision, and sensing methods to extend the safe and sustainable application of autonomous vehicles in land, air, and sea.

Gerard Dooly (Member, IEEE) received the B.Eng. degree from the Electronic and Computer Engineering Department, University of Limerick (UL), Limerick, Ireland, in 2003, and the Ph.D. degree from the Optical Fibre Sensors Research Centre, UL, in 2008, based on the topic "An Optical Fibre Sensor for the Measurement of Hazardous Emissions from Land Transport Vehicles,". He has worked extensively in field robotics with UL for more than ten years. His research interests include real-time 3-D reconstruction, machine vision, machine learning, optical fibre sensors, subsea structural health monitoring, teleoperation, and automated docking and intervention. He is involved in robotics for harsh environments in offshore setting and is developing systems to address beyond visual line of sight operations for UAS. He is also focused on the design and development of robotics and has engaged in numerous field operations and survey missions both in Ireland and on the continent. Some of his recent offshore operations involved environmental sensing, anti-mine countermeasure OPS, remote UAS for incident response, archaeological survey, and hybrid long range UAS technologies.