https://doi.org/10.1590/jatm.v18.1413 ORIGINAL PAPER

OPsCV: A Robust Framework for Aerial Navigation under Global Positioning System Denied Conditions

Nathan Augusto Zacarias Xavier[1,2,*], Elcio Hideiti Shiguemori[2], Marcos Ricardo Omena de Albuquerque Maximo[3]

1. Universidade Federal de Minas Gerais – Colégio Técnico – Educação Básica e Profissional – Belo Horizonte/MG – Brazil.
2. Departamento de Ciência e Tecnologia Aeroespacial – Instituto Tecnológico de Aeronáutica – Divisão de Comando, Controle, Comunicação, Computação, Inteligência, Vigilância e Reconhecimento – São José dos Campos/SP – Brazil.
3. Departamento de Ciência e Tecnologia Aeroespacial – Instituto de Estudos Avançados – Divisão de Ciência da Computação – São José dos Campos/SP – Brazil.

***Corresponding author:** nathanxavier@ufmg.br

ABSTRACT

Unmanned aerial vehicles (UAVs) rely heavily on the global navigation satellite system (GNSS) for accurate localization. However, GNSS signals are often unavailable or unreliable in contested or cluttered environments. This study presents the optimized pose prediction and cross-view (OPsCV), a robust and adaptable navigation framework that integrates deep inertial odometry with a simulated cross-view geolocalization module through an error-state Kalman filter. The system enables dynamic switching from GNSS-based positioning to a fused solution that combines inertial and vision-based estimates as GNSS signal quality degrades. The framework was evaluated using real UAV flight data under persistent GNSS denial, with results demonstrating reliable pose estimation and improved positioning accuracy compared to the UAV's internal navigation system. The OPsCV method maintained performance even with sparse cross-view updates, confirming its resilience under conservative operational conditions. These findings highlight the potential of fusing learned inertial measurements with statistical vision-based localization for autonomous aerial navigation in GNSS-denied environments.

Keywords: Unmanned Aerial Vehicle; Inertial-visual navigation; Air-ground collaboration; Deep inertial odometry; Cross-view geolocalization.

INTRODUCTION

The use of unmanned aerial vehicles (UAVs), also known as unoccupied aerial vehicles (Joyce *et al.* 2021), has grown rapidly over the past few decades due to their stability, flexibility, and broad applicability. The UAVs are widely used in defense, agriculture, mapping, and various civil and environmental applications (Ab Rahman *et al.* 2019; Afraimovich *et al.* 2000; Cobb *et al.* 1995; Li *et al.* 2025; Velusamy *et al.* 2021; Williams 2024; Zhou *et al.* 2024).

Many outdoor UAV applications require fully autonomous navigation, which depends on integrating multiple sensors and systems that must operate consistently, robustly, and harmoniously (Chang *et al.* 2023). Global navigation satellite systems (GNSS), including the Global Positioning System (GPS) (United States of America), GLONASS (Russia), Galileo (European Union), BeiDou (China), NavIC (India), and QZSS (Japan), are the primary real-time geospatial localization systems (Hegarty and Chatre 2008). However, GNSS signals are often vulnerable to jamming, spoofing, and other types of interference or denial (Afraimovich *et al.* 2000;

Received: Aug. 31, 2025 | **Accepted:** Oct. 26, 2025
Peer Review History: Single Blind Peer Review.
Section editor: Luiz Martins-Filho

Almeida *et al.* 2021; Allauddin *et al.* 2019; Chiella *et al.* 2019; Cobb *et al.* 1995; Torres *et al.* 2020; Xia *et al.* 2018; Xu *et al.* 2023; Yi *et al.* 2025; Yue *et al.* 2025). For UAVs, accurate localization is critical to mission success and autonomy, and this challenge can be addressed through the integration of inertial and visual navigation systems.

Inertial navigation, particularly when powered by tactical-grade inertial measurement units (IMUs) or high-precision sensors (Huang *et al.* 2023), offers a feasible solution for short-duration or small-scale trajectory estimation. However, consumer-grade IMUs often suffer from rapid drift in position estimates (Chen *et al.* 2018b; Li *et al.* 2012).

To mitigate this drift, deep learning techniques, particularly deep neural networks (DNNs), have demonstrated promising results in modeling inertial motion (Brossard *et al.* 2019; Chen *et al.* 2018b; Cortes *et al.* 2018; Esfahani *et al.* 2020a; T Wang *et al.* 2022; Zhou *et al.* 2022). Deep inertial odometry (DIO) models enable dead-reckoning-based pose estimation using only low-cost, off-the-shelf sensors. However, due to the cumulative nature of inertial drift, DIO solutions are generally limited to short-range or short-duration navigation tasks (Chen *et al.* 2018b; Esfahani *et al.* 2020b; Kim *et al.* 2021). Long-distance navigation remains challenging, as DIO models can reduce the drift amplitude but cannot eliminate it.

Another common approach for UAV navigation is vision-based localization, including techniques such as visual odometry (Bhowmick *et al.* 2021; Dantas 2023; Loianno *et al.* 2016; Romero *et al.* 2013; Roos 2018; Xie *et al.* 2021) and landmark recognition (Goltz *et al.* 2016; Nemra and Aouf 2009; Yang *et al.* 2022). These methods rely on images captured in flight to estimate global position. However, visual navigation methods can be constrained by payload, power, and computational requirements, limiting their use in smaller or resource-limited UAV platforms (Giubilato *et al.* 2020).

A recent advancement in visual navigation is the methodology based on cross-view geolocalization, which involves aligning images captured from different perspectives, typically ground-level and aerial views, that depict the same geographic location. This technique is often employed in collaborative multi-agent systems (Fervers *et al.* 2023a; b; Xia *et al.* 2024; S Zhu *et al.* 2022). The task presents significant challenges, such as extreme viewpoint variation, changes in lighting, occlusions, and seasonal differences, all of which must be addressed to reliably match ground-level images with satellite or aerial imagery (Ge *et al.* 2024; Liu and Li 2019; Zhu *et al.* 2021a). Cross-view geolocalization thus enables vision-based localization from a ground-based system to aid or enhance aerial navigation.

Despite advances in inertial and vision-based localization, no existing system robustly integrates the DIO and cross-view geolocalization into a unified framework for UAV navigation in GPS-denied environments. This gap motivates our proposed solution. In light of the growing interest in alternatives to simultaneous localization and mapping (SLAM) for outdoor UAV navigation, a robust sensor fusion framework that combines the DIO with cross-view geolocalization was proposed.

This work builds on our previous developments of both the DIO model and the cross-view geolocalization system, introduced and evaluated in earlier studies (Xavier *et al.* 2025a; b). The DIO model was trained on real UAV flight data using off-the-shelf IMU sensors. For the vision-based component, the statistical performance characteristics of a cross-view geolocalization model, including its mean localization error and uncertainty, which were previously derived from image-based experiments, are leveraged. In the proposed system, named optimized pose prediction and cross-view (OPsCV), the DIO module provides continuous motion estimation, while the statistical parameters of the cross-view model are used to inform drift correction. These two components are fused using a classic Kalman filter (KF), a well-established and interpretable method for multi-sensor integration.

The OPsCV approach is validated using post-processed IMU data from a real-world outdoor UAV flight conducted in collaboration with the Brazilian Air Force. A real-time kinematic (RTK) system provides ground truth for evaluation, while the cross-view error model informs the correction process during fusion. The statistical performance of the cross-view module to assess its impact on the fused system was analyzed. This evaluation builds upon earlier work, in which the DIO and cross-view systems were independently developed and validated. The present study confirms the robustness and practical value of the integrated framework for UAV navigation in GPS-denied scenarios.

In summary, the main contributions of this work are:

- To the best of the authors' knowledge, this is the first integration of a DIO model with a cross-view geolocalization-based error model, using statistical performance metrics to inform an error-state KF (ESKF) for outdoor UAV navigation in GPS-denied environments.

- The implementation of a dynamic switching mechanism that transitions planar position estimation from GNSS to fused DIO and cross-view updates when GPS signal quality degrades.
- A demonstration of how the complementary strengths of previously developed DIO and cross-view models can be effectively leveraged through integration.
- Validation of the proposed system using post-processed flight data from a real UAV mission conducted in collaboration with the Brazilian Air Force, confirming its performance and practical applicability.

DEEP INERTIAL ODOMETRY

The DIO has emerged as a promising solution to address the limitations of traditional inertial navigation systems, particularly the drift and error accumulation encountered in pedestrian dead reckoning (PDR). Before the adoption of DNNs, classical approaches such as strapdown inertial navigation algorithms (Savage 2007), KFs (Chui and Chen 2017; Huang and Dissanayake 2007; Kalman 1960), and particle filters (Berntorp *et al.* 2019) were commonly used to mitigate model uncertainty and sensor noise. These methods estimate position and velocity by integrating accelerometer and gyroscope data but are especially prone to drift over time, particularly when using low-cost sensors.

To improve robustness, hybrid approaches have emerged that integrate KFs with neural networks. In such systems, the network predicts or corrects sensor errors, enhancing navigation performance even with low-cost IMUs (Al-Sharman *et al.* 2020; Cohen and Klein 2024; Or and Klein 2022; Zou *et al.* 2020).

One of the earliest fully deep learning-based DIO models is the inertial odometry network (IONet) (Chen *et al.* 2018b), which specifically addressed drift in low-cost inertial sensors for pedestrian tracking. Since then, various architectures have been proposed, including long short-term memory (LSTM) networks (Chen *et al.* 2019; Esfahani *et al.* 2020a; Gong *et al.* 2021; Kim *et al.* 2021; Sun *et al.* 2021; Wagstaff *et al.* 2020), support vector machines (SVMs) (Yan *et al.* 2018), multilayer perceptrons (MLPs) (Wang *et al.* 2023), and other DNNs (Chen and Pan 2024). These models aim to improve odometry accuracy by enhancing sensor data processing, reducing noise, and correcting accumulated error.

Building on the Oxford Inertial Odometry Dataset (OxIOD) (Chen *et al.* 2018b), which includes data from smartphones carried in various ways (in hand, in a bag, on a trolley, or in a pocket), several extensions were developed, including MotionTransformer (Chen *et al.* 2019) and the Nine-Axis Extended IONet (Kim *et al.* 2021). Other research has focused on domain-specific assumptions such as zero-velocity updates (Brossard *et al.* 2019; Cortes *et al.* 2018; Wagstaff *et al.* 2020; Yu *et al.* 2019), heading estimation (Wang *et al.* 2019), sensor noise mitigation (Brossard *et al.* 2020; Wang 2021), and sensor fusion techniques (Gong *et al.* 2021; Sun *et al.* 2021; Wang *et al.* 2020).

While most DIO studies focus on pedestrian scenarios, some have extended these techniques to other platforms such as vehicles (Brossard *et al.* 2019; 2020; Tang *et al.* 2022), legged robots (Buchanan *et al.* 2021), and UAVs (Esfahani *et al.* 2020a; b; Zhang *et al.* 2021). Across these platforms, drift and related challenges persist, particularly when using low-cost, off-the-shelf sensors. The UAVs face additional challenges due to their operation in three-dimensional space, nonlinear and underactuated dynamics, and natural instability.

Moreover, most UAV datasets include only accelerometer and gyroscope data, often omitting the magnetometer, which is essential for accurate attitude estimation, particularly yaw, by resolving ambiguities that cannot be addressed by inertial sensors alone. In this study, three UAV-specific models based on LSTM architectures, trained on real-world outdoor flight data, were analyzed. These models were previously developed and evaluated in our earlier work (Xavier *et al.* 2025b), and this study builds upon that foundation as a natural continuation of previous research.

Cross-view geolocalization

Cross-view geolocalization is a rapidly advancing field with applications across various domains, including automotive (Hu and Lee 2019; Zhao *et al.* 2023; Zhou and Krahenbuhl 2022), aerospace (Cui *et al.* 2023; Shetty and Gao 2019), and robotics

(Ye *et al.* 2022), among others (Durgam *et al.* 2024; Wilson *et al.* 2023). The two predominant strategies for estimating global positions are based on convolutional neural networks (CNNs) and transformers.

The CNN-based approaches extract features from aerial and ground-view images and learn the correspondence between these two feature spaces (Cao *et al.* 2018; Shi *et al.* 2020; Workman *et al.* 2015; Xia *et al.* 2024). A notable advancement is the spatial-aware feature aggregation (SAFA) method (Shi *et al.* 2020), which introduced polar transformations to facilitate alignment between ground and aerial views. Several subsequent works have built upon this concept (Shi *et al.* 2020; 2022; Zhu *et al.* 2021a). Generative adversarial networks (GANs) have also been used to improve feature extraction through synthetic augmentation, increasing robustness and generalization (Regmi and Borji 2019; Regmi and Shah 2019; Toker *et al.* 2021; Wu *et al.* 2022).

In contrast, transformer-based models bypass the need for polar transformations and extensive data augmentation. The TransGeo (S Zhu *et al.* 2022) demonstrated that transformers can reduce computational demands, including GPU usage and inference time, while maintaining high accuracy. This finding has inspired the development of fully transformer-based models (Fervers *et al.* 2023a; b; Dai *et al.* 2022; Rodrigues and Tani 2023), as well as hybrid architectures combining transformers and CNNs (Shi *et al.* 2023; Y Wang *et al.* 2022; Zhao *et al.* 2023).

Cross-view geolocalization methods also benefit from principles of image retrieval, where localization is achieved by matching features based on image similarity (Xia *et al.* 2024). Recent methods such as GeoDTR (L Zhang *et al.* 2023) and the feature recombination module (FRM) (Zhang and Zhu 2024) enhance spatial alignment by leveraging improved feature representations. Alternatively, Sample4Geo (Deuser *et al.* 2023) focuses on optimizing the architectural pipeline for more robust performance.

Datasets such as CVUSA (Workman *et al.* 2015), CVACT (Liu and Li 2019), VIGOR (Zhu *et al.* 2021b), and the Brooklyn and Queens subsets (Workman *et al.* 2015), among others (Wilson *et al.* 2023), provide diverse ground-view images from arbitrarily distributed locations. These datasets pose challenges for models relying on spatial alignment, especially those using polar transformations (Xia *et al.* 2024; Y Zhu *et al.* 2021; S Zhu *et al.* 2022).

Combining semantic segmentation with cross-view geolocalization has emerged as an effective strategy to address variability in scene content and appearance. Segmentation can simplify neural network architectures, support parallelization, and enable real-time operation (Elhashash and Qin 2022; Zhang *et al.* 2024; Zhou and Krahenbuhl 2022. Semantic maps also allow for filtering out dynamic or transient elements such as vehicles, pedestrians, foliage, and seasonal color variations (Elhashash and Qin 2022; Y Zhu *et al.* 2022). Pseudo-segmentation techniques have been applied to align satellite images and bird's-eye view representations (Dai *et al.* 2022; Wang *et al.* 2023), while pseudo-labeled pose estimation has improved localization accuracy across heterogeneous data sources (Fervers *et al.* 2023a; Xia *et al.* 2022; 2024). Additionally, integrating sequential ground-view images and segmented satellite imagery into visual odometry frameworks has led to better pose estimation (Balaska *et al.* 2022). Most recently, the CVLocationTrans model (Yuan *et al.* 2024) introduced a fine-grained cross-view approach by combining self-attention and cross-attention layers. In this architecture, features are first extracted via ResNet50, then fused to establish spatial correspondences, followed by classification and regression heads for final position prediction.

Research applying true semantic segmentation to satellite images for cross-view geolocalization remains limited and was initially developed in previous work (Xavier *et al.* 2025a), specifically using aerial segmentation as ground truth. In that study, a method was proposed, in which semantic segmentation maps were used to guide location estimation, even when the street-view samples were not spatially aligned with the satellite imagery. The approach was designed to be adaptable to various cross-view datasets.

In this manuscript, the statistical performance results obtained in our previous work are expanded upon. As the current study relies on proprietary data provided by the Brazilian Air Force, access to the corresponding images required to implement the full cross-view geolocalization pipeline is not available. Therefore, the performance of the visual-based module by using GNSS-RTK measurements combined with the statistical parameters derived from our prior experiments was simulated (Xavier *et al.* 2025a; b). These estimated location outputs are then used as inputs to an integrated solution for UAV navigation in GPS-denied environments.

Air-ground collaboration

Collaborative robotics combines multiple intelligent agents to address challenges in complex environments, including physical constraints and multi-task execution (C Liu *et al.* 2022). Air-ground collaboration, in particular, is implemented to compensate for the individual limitations of UAVs, unmanned ground vehicles (UGVs), or both.

The UAVs are widely adopted due to their relatively low manufacturing cost and high operational flexibility. However, their payload capacity and battery life are limited, restricting the use of heavy equipment or computationally intensive systems (Giubilato *et al.* 2020; C Liu *et al.* 2022). UAVs also face sensor performance limitations caused by environmental factors such as proximity to buildings, signal interference, and poor coverage in certain areas (Bhowmick *et al.* 2021; Miller *et al.* 2022; Ran *et al.* 2021). In contrast, ground platforms offer larger payload capacity, enabling the integration of more powerful computers and complex operational systems for long-duration missions. However, they are constrained to a limited field of view and reduced onboard sensing capabilities.

By recognizing these complementary strengths and limitations, air-ground collaboration can support a range of strategies, including perspective sharing (D Liu *et al.* 2022; Miller *et al.* 2022; X Zhang *et al.* 2023), planning and decision-making (Duan *et al.* 2019; Korsah *et al.* 2013; Peng *et al.* 2019; Yan *et al.* 2013; Ulmer and Thomas 2018), and coordinated motion or goal achievement (Dong and Hu 2016; Ke *et al.* 2023; Mohiuddin *et al.* 2020; Parker 2009), among others (C Liu *et al.* 2022).

Effective air-ground collaboration often relies on sensor fusion techniques that integrate data from heterogeneous sources such as IMUs, cameras, and radar systems (Jin *et al.* 2024). These systems typically involve architectural considerations such as sensor availability, communication bandwidth, and time synchronization. Based on these factors, various algorithms, such as KFs (Carrillo-Arce *et al.* 2013; Ko *et al.* 2018; Rigatos 2012; Shen *et al.* 2017), graph-based optimization methods (Wen *et al.* 2020; Xu *et al.* 2022), and particle filters (Minetto *et al.* 2020; Sottile *et al.* 2011; Zocca *et al.* 2021), can be employed to enable reliable multimodal state estimation. In this study, a KF was adopted due to its robustness, interpretability, and real-time applicability for fusing motion- and vision-based localization estimates was adopted.

An air-ground collaborative scenario was conceptualized in which a ground-based system supports the UAV by handling computationally intensive geolocation tasks. This system captures and processes ground-level imagery to aid aerial localization via cross-view geolocalization. Building upon the previously described DIO and cross-view modules, a fused system integrating the DIO and a simulated cross-view correction through a KF was evaluated, forming a practical and continuous localization solution for UAV navigation in GPS-denied environments.

PROBLEM STATEMENT

In a UAV mission, GPS-denied scenarios can occur for various reasons, such as environmental obstructions or intentional signal jamming. Yet the mission must still be completed using alternative methods. The UAV is equipped with a Nine-Axis IMU, including accelerometers, gyroscopes, and magnetometers, operating at a fixed frequency of 100 Hz. This sensor exhibits noise and bias characteristics comparable to widely available commercial units.

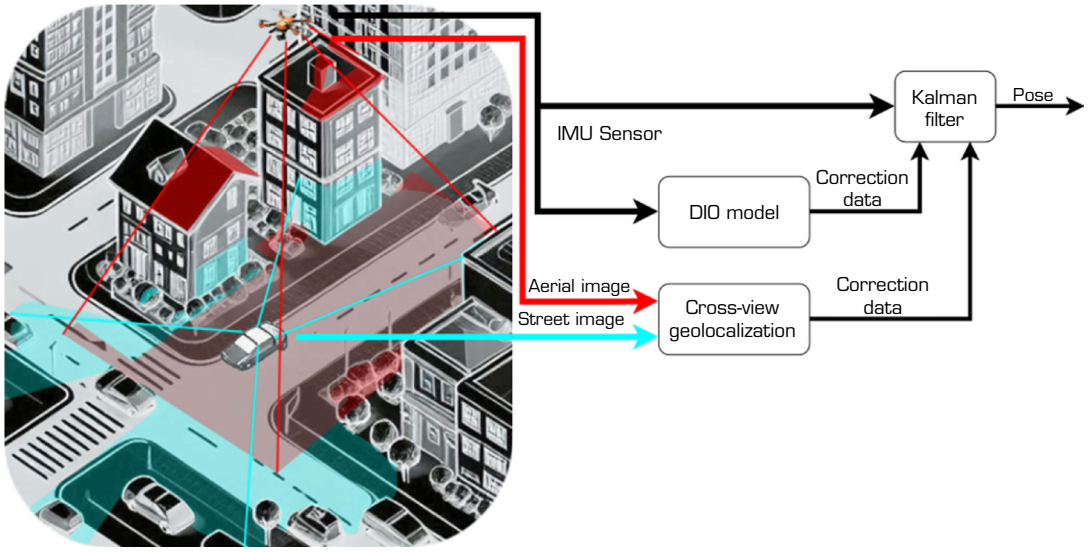
During the mission, the UAV captures nadir-oriented aerial images, with each image's center aligned to the UAV's latitude and longitude. The cross-view geolocalization module relies on ground-level imagery obtained in one of two ways: (i) real-time capture by a mobile ground vehicle equipped with an imaging system operating collaboratively with the UAV, or (ii) pre-captured and georeferenced images from existing databases covering the operational area. In both cases, this ground-level dataset is paired with aerial imagery to enable absolute position estimation without GNSS. When real-time ground capture is available, the processing can be offloaded to the ground system; otherwise, the UAV performs the processing onboard using its preloaded dataset.

The proposed solution combines the DIO, which estimates the UAV's pose from IMU data, with simulated cross-view geolocation data derived from one of the two acquisition modes described above. These are fused using a KF to provide robust pose estimation and enable autonomous aerial navigation in environments where GPS signals are unavailable.

Implementation details

Figure 1 presents an overview of the proposed method addressing the previously stated problem. This study focuses on UAV autonomous navigation in GPS-denied environments, integrating inertial odometry and cross-view geolocalization into a unified

framework. IMU data are processed using the DIO model, responsible for propagating the UAV's state in a KF and predicting its location.



Source: Elaborated by the authors.

Figure 1. Proposed UAV navigation framework for GPS-denied environments. The OPsCV integrates onboard inertial navigation via DIO processing IMU data with a simulated ground-based cross-view geolocalization module. A KF fuses IMU and location estimates to enhance navigation accuracy.

In parallel, a simulated cross-view geolocalization module has access to the UAV's nadir-oriented aerial images as well as geotagged street-view images that spatially overlap with the aerial images. It estimates the absolute position corresponding to the center of each nadir-oriented aerial image. This estimated position is then used by the KF as a measurement update, refining the predicted pose.

Due to the proprietary nature of the image dataset provided by the Brazilian Air Force, the cross-view module was implemented as a simulation using the statistical performance obtained from our previous work (Xavier *et al.* 2025a). The objective was to evaluate the feasibility and performance of this integration under realistic operational conditions, emphasizing practical implementation aspects in a GPS-denied scenario.

Unmanned aerial vehicle flight data

The UAV flight data used to validate this work was collected using a DJI Matrice 600 Pro hexacopter (DJI 2022), equipped with an A3 flight control system, multiple IMUs, GPS, and GNSS-RTK. The flight was conducted during the General Integrated Exercise for Emergency Response and Nuclear Physical Security (Exercício Geral Integrado de Resposta à Emergência e Segurança Física Nuclear) in 2023, in Angra dos Reis, Brazil, coordinated by the Brazilian Air Force (Cunha 2017). While a government press release (GSI 2023) provides general context for the exercise, this manuscript focuses on the technical aspects relevant the present study. Further details regarding the flight setup, model configurations, sensor settings, and related information can be found in our previous work (Xavier *et al.* 2025b).

Kalman filter

The KF, introduced by Rudolph Emil Kalman, in 1960 (Kalman 1960), is widely regarded as one of the most important tools for state estimation in dynamic systems. Based on statistical and probabilistic foundations, the algorithm seeks to provide optimal estimates of a system's states, even in the presence of uncertainty and noise (Chui and Chen 2017). Its significance is well documented and is considered a landmark theoretical development in the fields of estimation and control (Teixeira *et al.* 2010).

The filter operates by modeling linear dynamic systems subject to stochastic noise. Its classical formulation considers a state-space model described by Farrell and Farrell (2008):

$$x_k = F_k x_{k-1} + G_k u_k + w_k \tag{1}$$

$$y_k = H_k x_k + v_k \tag{2}$$

where $x_k \in R^n$, $y_k \in R^p$, and $u_k \in R^m$ represent the state, observation, and control input vectors, respectively. The term $w_k \in R^n$ denotes the process model noise, and $v_k \in R^p$ represents the sensor measurement noise. The matrices $F_k \in R^{n\times n}$ and $H_k \in R^{p\times n}$ define the state transition and observation matrices, respectively, while $G_k \in R^{n\times m}$ describes the control input model. All components are indexed by the discrete time step k.

The UAV's states are denoted as $x_k = [p^T, \theta^T]^T$, where p and θ represent the UAV's position and orientation, respectively, in the global frame. The KF also considers the process and measurement noise are assumed to be zero-mean Gaussian and mutually independent:

$$w_k \sim N(0, Q_k) \tag{3}$$

$$v_k \sim N(0, R_k) \tag{4}$$

where $Q_k \in R^{n\times n}$ is the process noise covariance matrix and $R_k \in R^{p\times p}$ is the measurement noise covariance matrix.

The KF operates in two main stages: prediction and update.

Prediction step: In the prediction phase, the state and expected output are estimated based on information available up to the previous time step. For linear systems, this step is defined as:

$$x_{k|k-1} = F_k x_{k-1|k-1} + G_k u_k \tag{5}$$

$$y_{k|k-1} = H_k x_{k|k-1} \tag{6}$$

The notation $[.]_{k|k-1}$ refers to a prior estimate (based on data from step $k-1$), while $[.]_{k|k}$ refers to a posterior estimate (after incorporating the current measurement).

The prediction of the state covariance matrix P_k is given by:

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \tag{7}$$

Update step: Given a new sensor measurement z_k, the innovation (or residual) r_k is computed as the difference between the actual and predicted outputs:

$$r_k = z_k - y_{k|k-1} = z_k - H_k x_{k|k-1} \tag{8}$$

In the proposed framework, the cross-view geolocalization module generates a discrete probability distribution (DPD) over candidate ground locations, with the peak value indicating the most likely position where the ground-level image was taken. Since each ground image is geotagged, it is possible to estimate the corresponding aerial view center point, which is treated as the predicted UAV position. This information serves as a global position measurement z_k and is used to correct the drift from the inertial prediction within the KF.

Additional sensor inputs, such as magnetometer data, can be incorporated into the measurement model to improve orientation updates and enhance overall pose estimation. By combining magnetometer and accelerometer readings, it is possible to estimate all attitude angles, with particular emphasis on yaw, as it is the most critical for navigation.

This residual is then used to correct the predicted state:

$$x_{k|k} = x_{k|k-1} + K_k r_k \tag{9}$$

where, K_k is the Kalman gain, which optimally weights the residual to minimize estimation error. It is computed as:

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \tag{10}$$

where S_k is the innovation covariance:

$$S_k = H_k P_{k|k-1} H_k^T S_k^{-1} + R_k \tag{11}$$

Finally, the state covariance matrix is updated:

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \tag{12}$$

The KF enables robust state estimation for IMU-based navigation systems, even in the presence of sensor noise and uncertainty. The following sections describe its real-world implementation, highlighting challenges such as initialization, calibration, and long-term accuracy degradation.

Deep inertial odometry model

Three DIO models were considered: IONet, AbolDeepIO, and Nine-Axis Extended IONet, all adapted to use Nine-Axis IMU sensor data as input and trained specifically for UAV systems. The performance of each model is presented in our previous work (Xavier *et al.* 2025b).

The original IONet model (Chen *et al.* 2018a) is a deep inertial solution that proposes a planar error reduction method for low-cost inertial sensors, originally applied to pedestrian tracking. This solution utilizes a single LSTM cell, which retains temporal information from 200 sequential data frames to track movement and heading variations.

The original AbolDeepIO (Esfahani *et al.* 2020a) is the first DIO solution developed for UAVs, using the indoor ASL EuRoC MAV dataset (Burri *et al.* 2016). Similar to IONet, this model relies solely on accelerometer and gyroscope sensors.

Finally, the Nine-Axis Extended IONet (Kim *et al.* 2021) incorporates magnetometer data to enhance attitude tracking. Unlike IONet and AbolDeepIO, this approach includes 3-axis magnetometer data as an additional input to the neural network, improving

orientation estimation. Similar to AbolDeepIO, the IMU data are first processed through two layers of one-dimensional 1D-CNNs, followed by a max-pooling layer to reduce computational complexity and extract relevant features. These features are then passed through two bi-directional LSTM (bi-LSTM) layers with dropout layers interspersed to capture temporal dependencies from both forward and backward sequences. The final dense layer estimates position and quaternion variations.

In this study, updated versions of these DIO models, as detailed in our previous work (Xavier *et al.* 2025b), were employed. All DIO implementations are responsible for estimating delta position and delta attitude at a frequency synchronized with the IMU and are implemented as dead-reckoning solutions.

Cross-view geolocalization model

The cross-view geolocalization performance presented in our previous work, which was implemented and evaluated across multiple datasets, was considered. The implemented model performs semantic segmentation of satellite imagery and calculates the DPD corresponding to a ground-view image taken within the same aerial region. This study emphasizes the statistical performance derived from these DPDs rather than relying on full image-to-image matching.

The evaluation metrics included multiple accuracy scenarios in a baseline method, categorized as "positive" and "semi-positive," based on the alignment of the street-view position with the aerial image. A ground view is classified as semi-positive if the corresponding aerial image covers only a portion of the scene. Specifically, a street-view position is deemed positive if it falls within the central region of size $L/2$ x $L/2$; otherwise, it is classified as semi-positive (Zhu *et al.* 2021b). These metrics informed our assessment of the transformer architectures described below.

The previous study also evaluated the top-ranked regions of the predicted discrete distribution by filtering probabilities below specific quantile thresholds: 68.27% (1σ), 95.45% (2σ), and 99.73% (3σ).

Two Vision Transformer (ViT) (Dosovitskiy *et al.* 2020) implementations were considered. The first, the Multi-Scale Transformer (MST), based on the Crossformer architecture (Wang *et al.* 2024), extracts embeddings from multiple scales via successive patch embedding stages and fuses the information with a fusion module. This method also captures global local features through sequential attention blocks at the end of each stage.

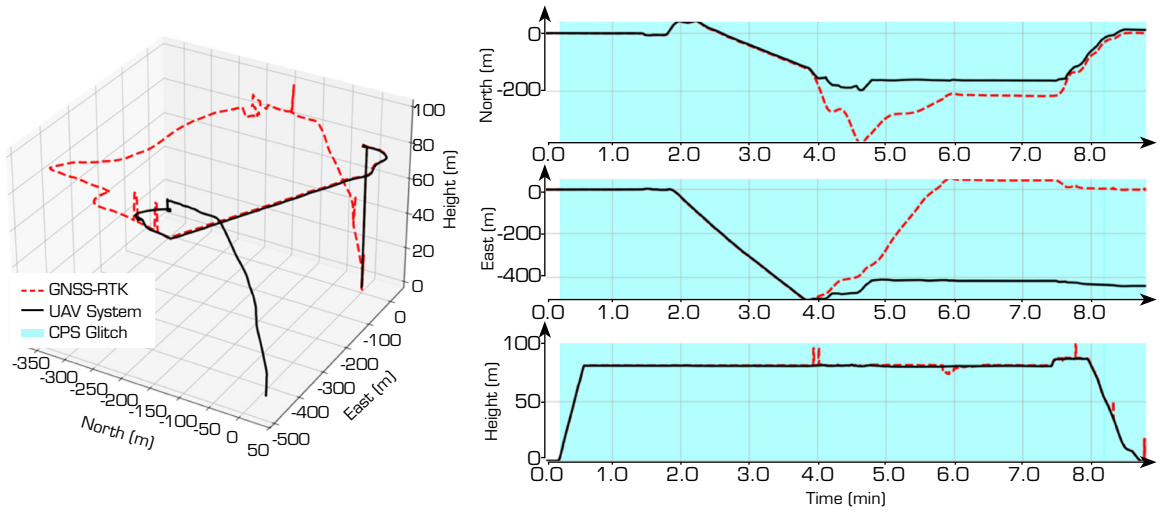
The second implementation, FeatUp, differs from previous approaches by aiming to produce a high-resolution feature output, maintaining consistency across multiple low-resolution feature maps (Fu *et al.* 2024). It employs joint bilateral upsampling (JBU) (Kopf *et al.* 2007) as an upsampling strategy. Results show that the downsampled features and the transformed original images are comparable, enabling the model to reconstruct a high-quality, high-resolution feature map.

To enable integration with the KF, the discrete probability output from the cross-view geolocalization model was simplified to a Gaussian-like representation. Specifically, the distribution was approximated by computing a weighted mean and covariance matrix over the predicted coordinates, restricted to regions within a selected confidence threshold, typically the 1σ quantile (68.27%) of the cumulative probability mass. This simplification allows the geolocation output to be treated as a pseudo-measurement with associated uncertainty, compatible with the update step of the KF. The resulting pseudo-observation includes the estimated position (mean) and a corresponding measurement covariance matrix, which reflects the spatial spread of the top-ranked predictions. This approach balances model expressiveness with computational tractability, enabling seamless fusion with the inertial navigation pipeline.

EXPERIMENTAL SETUP

The UAV platform used was the DJI Matrice 600 Pro, as detailed. The onboard IMU data was were processed using three updated DIO models, IONet, AbolDeepIO, and the Nine-Axis Extended IONet, all trained on UAV flight data. In the absence of reliable GPS signals, a KF was used to propagate the UAV state based on DIO predictions and to incorporate measurement updates from simulated geolocation inputs derived from the statistical model of cross-view geolocation performance. The proposed method was evaluated against the UAV's internal navigation system, which utilizes all available onboard data, including the degraded GPS signal.

Figure 2 presents a 3D comparison between the GNSS-RTK trajectory (red) and the trajectory estimated by the UAV's internal navigation system (black). The GPS availability was assessed based on dilution of precision indicators, namely position dilution of precision (PDOP) or geometric dilution of precision (GDOP) (Kartal *et al.* 2023; Liu *et al.* 2017). However, in the flight log used for this analysis, it was not possible to determine the exact moment or degree of signal degradation, as the information is recorded in a binary format that only indicates whether the GNSS signal was available or unavailable. As observed, GNSS signals were unavailable for most of the flight, remaining lost for approximately 9 min, with only a few intermittent peaks of valid signal. This behavior is illustrated in the north and east plots, where intermittent GNSS availability, referred to as GPS glitches, is highlighted in cyan.



Source: Elaborated by the authors.

Figure 2. Comparison of UAV flight trajectories under GNSS-denied conditions. The GNSS-RTK (red), UAV internal navigation (black), and GPS glitch intervals (cyan) were highlighted.

The flight dataset is proprietary to the Brazilian Air Force and cannot be publicly shared due to its sensitive nature, including restricted flight trajectories and operational details. A specific flight was selected in which continuous GPS denial was observed throughout the mission. Notably, no intentional GPS blocking or spoofing was configured; rather, the UAV naturally operated in an environment where GPS signals were persistently unavailable. As a contingency, the flight was manually controlled by a ground pilot and a copilot. The UAV's onboard log includes both internally predicted motion and GNSS-RTK data. However, the RTK information is not part of the UAV's internal navigation solution and was only partially recorded during the flight.

It is important to emphasize that this flight was not used during training or validation phases. Due to its distinctive characteristics, it was reserved exclusively for testing. All performance evaluations and analyses presented in this work are based on post-processed data collected independently of the model development stage.

For the proposed integration, an ESKF (Dai *et al.* 2022; Deilamsalehy and Havens 2016; Lupton and Sukkarieh 2009) based on IMU measurements was implemented. The state vector includes the UAV's position (p), velocity (v), and attitude represented by a quaternion (q). The ESKF formulation also models IMU biases, specifically the accelerometer bias (b_a) and gyroscope bias (b_ω). The complete state vector x at time t is defined as:

$$x_t = \left[p_t^T \; v_t^T \; q_t^T \; b_{\omega_t}^T \; b_{a_t}^T\right]^T \tag{13}$$

To simulate the geolocation updates, samples were drawn from a multivariate Gaussian distribution centered on the ground truth (GNSS-RTK) location, using the error mean and covariance obtained in prior work. Updates were performed periodically to emulate realistic cross-view estimation intervals.

In the implemented solution, the distance error associated with the geolocation measurements in planar coordinates was modeled using a normal distribution $N(\mu = 23.49, \sigma^2 = 100)$, where the mean corresponds to 23.49 m and the standard deviation to 10 m. These statistical parameters represent the typical localization error observed in cross-view matching evaluations and are used to define the measurement noise covariance in the filter.

The system was designed to perform a seamless switching transition in the event of GNSS signal loss, automatically replacing GNSS updates with predictions from the DIO and cross-view geolocation modules. The commutation process can be guided by GNSS quality indicators, such as the number of available satellites, PDOP, and GDOP (Kartal *et al.* 2023; Liu *et al.* 2017).

The ESKF manages this transition by maintaining continuous and robust state estimation, fusing dead-reckoning outputs with cross-view pseudo-measurements under degraded navigation conditions. The filter's process and measurement noise covariances were empirically tuned to reflect the expected characteristics of the IMU sensors and the statistical performance of the geolocation module. The DIO models provide corrections to the north-east-down (NED) velocity components, while the cross-view module, when available, updates the NED position.

Evaluation metrics

The system performance is evaluated using the mean position error and the frequency of cross-view updates as metrics. These metrics were selected to capture both the accuracy of the estimated position and the temporal reliability of geolocation updates, which are key factors for robust navigation in GPS-denied environments.

The mean position error is defined as the Euclidean distance between the estimated UAV position and the ground truth provided by GNSS-RTK, averaged over the entire flight. The update frequency is calculated relative to the IMU sampling rate of 0.01 s. In the experiments, cross-view corrections were allowed at intervals of up to 30 s, representing a conservative scenario where updates are relatively sparse. Such long intervals may occur in real operations due to limited visual overlap, restricted communications, or occlusions.

To assess robustness, the system was evaluated under different update intervals, analyzing how delayed corrections affect drift and overall accuracy. Infrequent updates increase reliance on dead reckoning, which may lead to accumulated error, especially in the presence of noisy or degraded measurements.

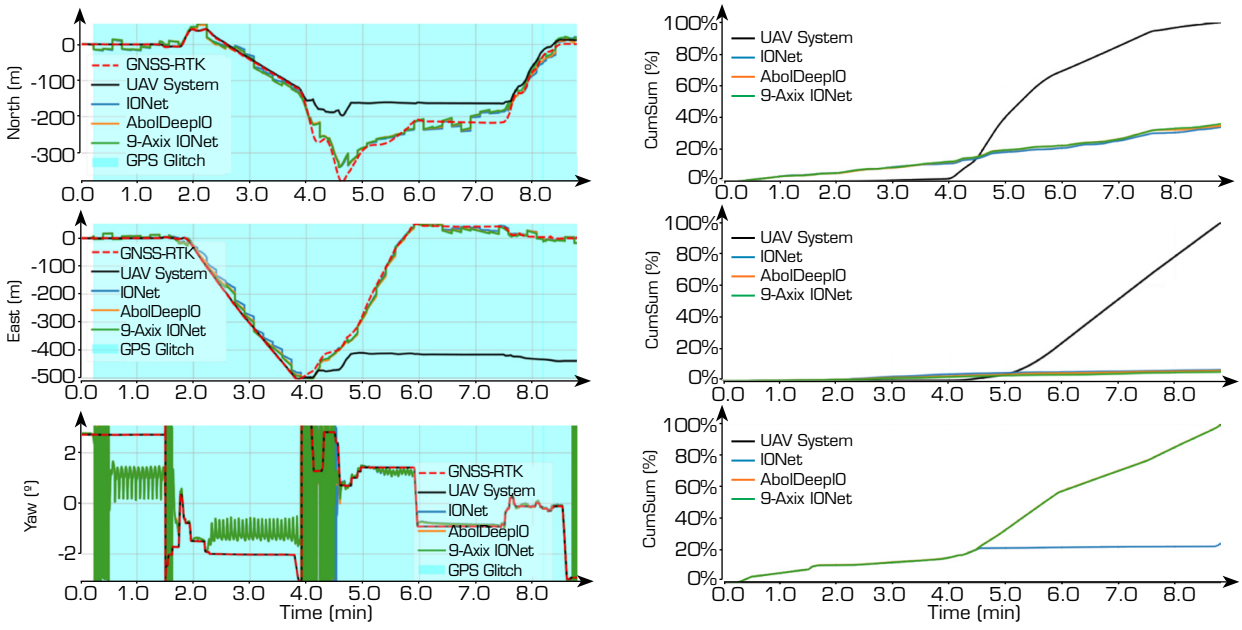
The performance of the proposed solution is also compared with the UAV's internal navigation system, extracted from the onboard log. The analysis focuses primarily on the horizontal components (north and east), since the altitude and attitude can typically be supported by auxiliary sensors such as barometers and magnetometers, which remain functional even in GPS-denied conditions.

RESULTS AND ANALYSIS

Figure 3 compares the GNSS-RTK reference trajectory (red) with the trajectory estimated by the UAV's internal navigation system (black) and by the integrated frameworks. The proposed approach is shown with cross-view corrections applied every 10 s, using updated DIO models: IONet (blue), AbolDeepIO (orange), and Nine-Axis IONet (green). The results are illustrated for the north and east coordinates, as well as for yaw orientation. Intervals of intermittent GNSS availability, referred to as GPS glitches, are highlighted in cyan. To further support the analysis, the cumulative sum error (CumSum), computed as the accumulated distance error values, is also presented for performance comparison.

As observed, GNSS signals were unavailable for most of the flight, except for a few seconds before takeoff when limited GPS data were still recorded. It was not possible to precisely determine the moment of complete signal loss, as the log continued to register occasional GPS readings even when PDOP and HDOP values were low. By integrating the DIO with simulated cross-view updates, the predicted trajectory showed a substantial improvement compared to the UAV's internal navigation system. All DIO-based solutions reduced the positional error by approximately 15–25 m, assuming a 10-second cross-view update interval.

In yaw estimation, the OPsCV performance remained considerably lower, consistent with limitations previously reported for DIO-based methods (Xavier *et al.* 2025b). Nonetheless, positional predictions outperformed the UAV's internal navigation system across all architectures. An improvement in yaw estimation was observed around the 5-minute mark of the flight; however, no specific measurement anomalies were identified or investigated that could explain the degraded yaw performance before this point.



Source: Elaborated by the authors.

Figure 3. Comparison of UAV flight trajectories under GNSS-denied conditions. GNSS-RTK (red), UAV internal navigation (black), and GPS glitch intervals (cyan) were highlighted. Proposed framework with cross-view updates every 10 s: Updated IONet (blue), Updated AbolDeepIO (orange), and Updated Nine-Axis IONet (green) were highlight. The plots also include the cumulative sum error (CumSum).

The cumulative errors of the integration techniques were similar throughout the flight. A crossing point between the predicted trajectories and the UAV's internal navigation system occurred around the 5-minute mark, coinciding with the observed improvement in yaw estimation. This suggests that, despite comparable overall positional performance among the integration approaches, temporal variations in orientation accuracy can influence the relative trajectory predictions.

As a qualitative analysis, Table 1 summarizes the median, mean, and maximum errors obtained throughout the UAV flight test for both position and orientation. As shown, all proposed DIO-based integration approaches significantly outperformed the UAV's internal navigation system in the horizontal plane. In the north axis, the integrated solutions reduced median errors from

Table 1. Comparison of median, mean, and maximum errors (in meters for position and in radians for yaw) over the entire UAV flight test.

Direction	Metrics	UAV system	Updated IONet	Updated AbolDeepIO	Updated Nine-Axis IONet
North (m)	Median	37.23	14.32	14.71	14.90
	Mean	12.83	9.75	9.35	11.06
	Max	184.16	65.17	58.84	58.85
East (m)	Median	188.52	25.54	20.18	19.39
	Mean	65.52	21.74	17.08	20.96
	Max	465.36	91.55	80.83	62.35
Height (m)	Median	0.67	6.46	6.23	7.50
	Mean	0.42	4.00	3.82	3.33
	Max	19.76	52.17	52.40	56.18
Yaw (rad)	Median	0.000	0.547	0.994	0.995
	Mean	0.000	3.136	3.140	3.141
	Max	0.001	3.136	3.140	3.141

Note. For each metric, the best performance is highlighted in bold, and the second best is underlined.
Source: Elaborated by the authors.

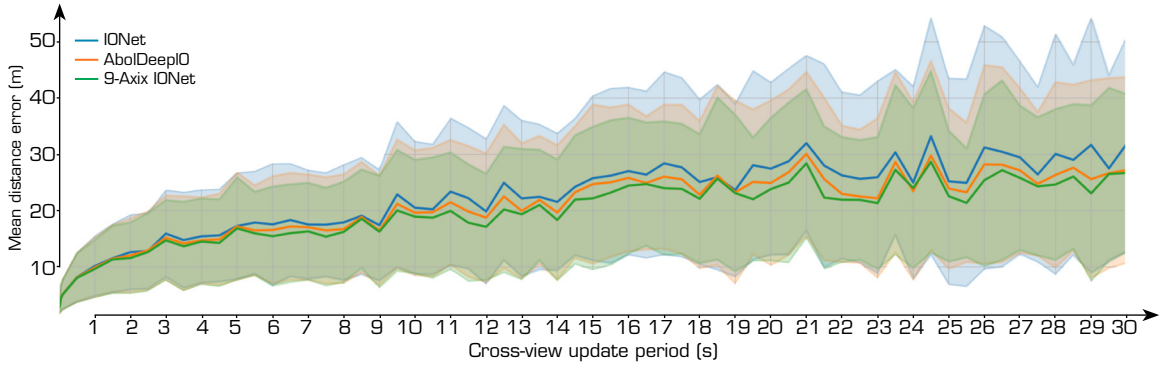
37.23 m to around 14-15 m and maximum errors from 184.16 m to approximately 59-65 m. AbolDeepIO achieved the lowest mean error in the north axis (9.35 m), while IONet Adapted reached the lowest median error (14.32 m). In the east axis, median errors decreased from 188.52 m to 19-20 m and maximum errors from 465.36 m to 62-81 m, with Nine-Axis IONet Adapted showing the lowest maximum error (62.35 m). For altitude estimation, the UAV's internal system maintained the lowest median and mean errors (0.67 m and 0.42 m), whereas the integrated methods exhibited higher variability, with Nine-Axis IONet Adapted reaching a maximum error of 56.18 m. Regarding yaw estimation, the internal system remained superior, with median and mean errors close to zero, while the integrated approaches presented median values around 0.55 rad and maximum errors up to 3.14 rad. Overall, these results confirm that all updated DIO models improved horizontal positional accuracy compared to the UAV's internal system, with AbolDeepIO providing the most balanced performance across axes.

A detailed analysis of sensor measurements or flight conditions was not conducted to investigate the anomalies during flights and their influence on the observed error amplitude. In addition, the GPS-denied periods occurred naturally, with no controlled or previously expected operation.

Cross-view contribution to navigation performance

In general, because the cross-view estimates are anchored to the UAV's true position, even when incorporating the Gaussian error model, the integrated frameworks consistently outperformed the UAV's internal navigation system. This section focuses on evaluating the specific impact of cross-view corrections on navigation accuracy. As observed, all combinations of DIO and cross-view modules led to measurable improvements in position and orientation estimation, demonstrating the effectiveness of the switching mechanism and highlighting the contribution of cross-view integration to reducing navigation errors.

The robustness of the OPsCV methodology is illustrated in Fig. 4, which shows the distance error for different cross-view geolocation update intervals, ranging from 0.01 s to 30 s. As a reference, the UAV's internal navigation system, when operating under GPS failure, exhibits a mean error of approximately 66 m, according to Table 1. As shown, the proposed frameworks maintain comparable error performance for update intervals shorter than 3 s. Within this range, mean errors increase up to approximately 15 m with a standard deviation of around 5 m. For longer intervals, errors gradually rise to approximately 25 m with a standard deviation of ± 10 m at a 15 s update interval. For even longer update periods, the growth of the mean error occurs at a slower rate.



Source: Elaborated by the authors.

Figure 4. Mean distance error of the proposed frameworks with updated DIO models, Updated IONet (blue), Updated AbolDeepIO (orange), and Updated Nine-Axis IONet (green), were highlighted, under varying cross-view update intervals, from 0.01 s to 30 s. The black line represents the average error of the UAV's internal navigation system. Shaded regions indicate the standard deviation.

Overall, the methodology using the Nine-Axis IONet Adapted DIO technique exhibits slightly lower errors than AbolDeepIO Adapted and IONet Adapted, in that order. Considering both the mean and standard deviation values, even at a 30 s update interval, all approaches still achieve errors smaller than the mean error of the UAV's internal navigation system.

The presented results demonstrate consistent and comparable performance across all updated DIO models. The simulated cross-view geolocation error, derived from a statistical performance model, indicates that the proposed framework would provide

meaningful improvements for UAV navigation even in real-world scenarios. Notably, under conservative conditions with relatively sparse position updates, the framework still outperforms the UAV's internal navigation system. It is important to highlight that the primary goal of this study is to benchmark the proposed method against the UAV's native navigation solution, which serves as a well-established baseline. Furthermore, the temporary loss of GPS did not compromise the flight, as it was safely managed by trained pilot and copilot from the Brazilian Air Force.

CONCLUSION

This study presented a robust and adaptable navigation framework for UAVs operating in GNSS-denied environments, integrating DIO techniques with a simulated cross-view geolocalization module through an ESKF, referred to as OPsCV. The system automatically transitions from GNSS-based updates to a fusion of DIO predictions and cross-view position estimates as GPS signal quality degrades. The results demonstrate that this approach maintains reliable state estimation even in the absence of satellite signals, validating its effectiveness for resilient UAV navigation in contested or constrained environments.

Three updated DIO models, Updated IONet, Updated AbolDeepIO, and Updated Nine-Axis IONet, were evaluated using real flight data collected under persistent GPS denial. These models build upon our previous work (Xavier *et al.* 2025b). The proposed system showed consistent improvements over the UAV's internal navigation solution, particularly in north-east positioning. The statistical cross-view geolocation model is also based on our prior study (Xavier *et al.* 2025a), resulting in an integrated solution for UAV navigation. The performance achieved by the proposed framework outperforms the UAV's native system across all tested scenarios, even with sparse cross-view update intervals of up to 30 s. Such large update intervals reflect conservative scenarios, likely caused by limited visual overlap, restricted communication with ground systems, or camera occlusion.

These results validate the effectiveness of integrating learned motion priors with image-based geolocation, confirming that even under degraded conditions, the framework preserves navigational performance. Overall, these findings support the viability of the proposed approach for deployment in operational scenarios where GNSS availability cannot be guaranteed, contributing to resilient and autonomous aerial navigation.

It is important to highlight that the primary goal of this study was to benchmark the proposed method against the UAV's native navigation system, which serves as a well-established baseline. Furthermore, despite GPS signal loss, flight safety was maintained throughout, as the aircraft was manually controlled by trained pilot and copilot from the Brazilian Air Force.

Looking ahead, there are several avenues for continued research with OPsCV, including:

- Investigating additional extended multi-sensor fusion techniques to further enhance navigation robustness.
- Exploring hybrid approaches that combine complementary methods for improved accuracy.
- Incorporating alternative computer vision techniques to complement the cross-view geolocalization module.
- Deploying the developed models on embedded platforms to facilitate real-time operation.
- Planning and executing fully autonomous UAV missions in GNSS-denied environments.
- Investigating novel strategies to mitigate inertial drift more effectively.
- Conducting comprehensive ablation studies on training hyperparameters to optimize model performance.

CONFLICTS OF INTEREST

Nothing to declare.

AUTHORS' CONTRIBUTION

Conceptualization: Xavier NAZ, Shiguemori EH, Maximo MROA; **Methodology:** Xavier NAZ, Shiguemori EH, Maximo MROA; **Software:** Xavier NAZ; **Validation:** Xavier NAZ, Shiguemori EH, Maximo MROA; **Formal analysis:** Xavier NAZ;

Investigation: Xavier NAZ; **Resources:** Shiguemori EH; **Data Curation:** Xavier NAZ; **Writing - Original Draft:** Xavier NAZ; **Writing - Review & Editing:** Shiguemori EH, Maximo MROA; **Visualization:** Xavier NAZ; **Supervision:** Shiguemori EH, Maximo MROA; **Project administration:** Shiguemori EH; **Final approval:** Xavier NAZ.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable.

FUNDING

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
Project No: 88887.929508/2023-00

Conselho Nacional de Desenvolvimento Científico e Tecnológico
Grants No: 316948/2023-3; 307525/2022-8

DECLARATION OF USE OF ARTIFICIAL INTELLIGENCE TOOLS

The authors declare that no artificial intelligence tools were used in the preparation, writing, data analysis, or review of this manuscript.

ACKNOWLEDGEMENTS

This research was developed within the IDeepS, which is supported by the Laboratório Nacional de Computação Científica (LNCC/MCTI, Brazil) via resources of the SDumont supercomputer.

REFERENCES

[DJI] Dà-Jiāng Innovations (2022) Matrice 600 Pro. https://www.dji.com/br/matrice600-pro

[GSI] Gabinete de Segurança Institucional (2023) Release do Exercício Geral Integrado de Resposta à Emergência e Segurança Física Nuclear em Angra dos Reis. In Portuguese. https://www.gov.br/gsi/pt-br/centrais-de-conteudo/noticias/2023-1/release-do-exercicio-geral-integrado-de-resposta-a-emergencia-e-seguranca-fisica-nuclear-em-angra-dos-reis-1

Afraimovich EL, Lesyuta OS, Ushakov II (2000) Magnetospheric disturbances, and the GPS operation. https://arxiv.org/abs/physics/0009027

Allauddin MS, Kiran GS, Raj K, Srinivas G, Mouli GUR, Prasad PV (2019) Development of a surveillance system for forest fire detection and monitoring using drones. Paper presented 2019 IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE; Yokohama, Japan. https://doi.org/10.1109/IGARSS.2019.8900436

Almeida DRA, Broadbent E, Zambrano AMA, Ferreira MP, Brancalion PHS (2021) Fusion of LiDAR and hyperspectral data from drones for ecological questions: The GatorEye Atlantic Forest restoration case study. Paper presented 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE; Brussels, Belgium. https://doi.org/10.1109/IGARSS47720.2021.9554023

Al-Sharman MK, Zweiri Y, Jaradat MA, Al-Husari R, Gan D, Seneviratne LD (2020) Deep-learning-based neural network training for state estimation enhancement: Application to attitude estimation. IEEE Trans Instrum Meas 69(1):24-34. https://doi.org/10.1109/tim.2019.2895495

Balaska V, Bampis L, Gasteratos A (2022) Self-localization based on terrestrial and satellite semantics. Eng Appl Artif Intell 111:104824. https://doi.org/10.1016/J.ENGAPPAI.2022.104824

Berntorp K, Hoang T, Di Cairano S (2019) Motion planning of autonomous road vehicles by particle filtering. IEEE Trans Intell Veh 4(2):197-210. https://doi.org/10.1109/TIV.2019.2904394

Bhowmick J, Singh A, Gupta H, Nallanthighal R (2021) A novel approach to computationally lighter GNSS-denied UAV navigation using monocular camera. Paper presented 2021 7th International Conference on Automation, Robotics and Applications (ICARA). IEEE; Prague, Czech Republic. https://doi.org/10.1109/ICARA51699.2021.9376502

Brossard M, Barrau A, Bonnabel S (2019) RINS-W: Robust inertial navigation system on wheels. Paper presented 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE; Macau, China. https://doi.org/10.1109/iros40897.2019.8968593

Brossard M, Barrau A, Bonnabel S (2020) AI-IMU dead-reckoning. IEEE Trans Intell Veh 5(4):585-595. https://doi.org/10.1109/tiv.2020.2980758

Buchanan R, Camurri M, Dellaert F, Fallon M (2021) Learning inertial odometry for dynamic legged robot state estimation. Paper presented 2021 5th Conference on Robot Learning. PMLR; London, UK. https://doi.org/10.48550/ARXIV.2111.00789

Burri M, Nikolic J, Gohl P, Schneider T, Rehder J, Omari S, Achtelik MW, Siegwart R (2016) The EuRoC micro aerial vehicle datasets. Int J Robot Res 35(10):1157-1163. https://doi.org/10.1177/0278364915620033

Cao R, Zhu J, Tu W, Li Q, Cao J, Liu B, Zhang Q, Qiu G (2018) Integrating aerial and street view images for urban land use classification. Remote Sens 10(10):1553. https://doi.org/10.3390/RS10101553

Carrillo-Arce LC, Nerurkar ED, Gordillo JL, Roumeliotis SI (2013) Decentralized multi-robot cooperative localization using covariance intersection. Paper presented 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE; Tokyo, Japan. https://doi.org/10.1109/iros.2013.6696534

Chang Y, Cheng Y, Manzoor U, Murray J (2023) A review of UAV autonomous navigation in GPS-denied environments. Robot Auton Syst 170:104533. https://doi.org/10.1016/j.robot.2023.104533

Chen C, Lu X, Markham A, Trigoni N (2018a) IONet: Learning to cure the curse of drift in inertial odometry. Proc AAAI Conf Artif Intell 32(1). https://doi.org/10.1609/aaai.v32i1.12102

Chen C, Miao Y, Lu CX, Xie L, Blunsom P, Markham A, Trigoni N (2019) MotionTransformer: Transferring neural inertial tracking between domains. Proc AAAI Conf Artif Intell 33(1):8009-8016. https://doi.org/10.1609/aaai.v33i01.33018009

Chen C, Pan X (2024) Deep learning for inertial positioning: A survey. IEEE Trans Intell Transp Syst 25(9):10506-10523. https://doi.org/10.1109/tits.2024.3381161

Chen C, Zhao P, Lu CX, Wang W, Markham A, Trigoni N (2018b) OxIOD: The dataset for deep inertial odometry. https://arxiv.org/abs/1809.07491

Chiella ACB, Teixeira BOS, Pereira GAS (2019) State estimation for aerial vehicles in forest environments. Paper presented 2019 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE; Atlanta, USA. https://doi.org/10.1109/ICUAS.2019.8797822

Chui CK, Chen G (2017) Kalman filtering. Cham: Springer. https://doi.org/10.1007/978-3-319-47612-4

Cobb S, Lawrence D, Christie J, Walter T, Chao YC, Powell D, Parkinson B (1995) Observed GPS signal continuity interruptions. Paper presented 1995 Proceedings of Ion GPS. Institute of Navigation; California, EUA.

Cohen N, Klein I (2024) Inertial navigation meets deep learning: A survey of current trends and future directions. Results Eng 24:103565. https://doi.org/10.1016/j.rineng.2024.103565

Cortes S, Solin A, Kannala J (2018) Deep learning based speed estimation for constraining strapdown inertial navigation on smartphones. Paper presented 2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE; Aalborg, Denmark. https://doi.org/10.1109/mlsp.2018.8516710

Cui Z, Zhou P, Wang X, Zhang Z, Li Y, Li H, Zhang Y (2023) A novel geo-localization method for UAV and satellite images using cross-view consistent attention. Remote Sens 15(19):4667. https://doi.org/10.3390/RS15194667

Cunha RDS (2017) A comunicação dos riscos na preparação para emergências nucleares: um estudo de caso em Angra dos Reis, Rio de Janeiro (PhD thesis). São Paulo: Universidade de São Paulo. https://doi.org/10.11606/T.85.2017.tde-06092017-085924

Dai J, Hao X, Liu S, Ren Z (2022) Research on UAV robust adaptive positioning algorithm based on IMU/GNSS/VO in complex scenes. Sensors 22(8):2832. https://doi.org/10.3390/S22082832

Dantas BNC (2023) Fusão de técnicas de inteligência artificial e visão computacional aplicada à estimação de posição de veículos aéreos com auxílio de imagens satelitais (master's thesis). São José dos Campos: Instituto Tecnológico de Aeronáutica.

Deilamsalehy H, Havens TC (2016) Sensor fused three-dimensional localization using IMU, camera and LiDAR. Paper presented 2016 IEEE SENSORS. IEEE; Orlando, USA. https://doi.org/10.1109/ICSENS.2016.7808523

Deuser F, Habel K, Oswald N (2023) Sample4Geo: Hard negative sampling for cross-view geo-localisation. Paper presented 2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE; Paris, France. https://doi.org/10.1109/ICCV51070.2023.01545

Dong X, Hu G (2016) Time-varying formation control for general linear multi-agent systems with switching directed topologies. Automatica 73:47-55. https://doi.org/10.1016/j.automatica.2016.06.024

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, *et al.* (2020) An image is worth 16×16 words: Transformers for image recognition at scale. https://doi.org/10.48550/arxiv.2010.11929

Duan R, Wang J, Jiang C, Yao H, Ren Y, Qian Y (2019) Resource allocation for multi-UAV aided IoT NOMA uplink transmission systems. IEEE Internet Things J 6(4):7025-7037. https://doi.org/10.1109/jiot.2019.2913473

Durgam A, Paheding S, Dhiman V, Devabhaktuni V (2024) Cross-view geo-localization: A survey. IEEE Access 12:192028-192050. https://doi.org/10.1109/access.2024.3507280

Elhashash M, Qin R (2022) Cross-view SLAM solver: Global pose estimation of monocular ground-level video frames for 3d reconstruction using a reference 3d model from satellite images. ISPRS J Photogramm Remote Sens 188:62-74. https://doi.org/10.1016/J.ISPRSJPRS.2022.03.018

Esfahani MA, Wang H, Wu K, Yuan S (2020a) AbolDeepIO: A novel deep inertial odometry network for autonomous vehicles. IEEE Trans Intell Transp Syst 21(5):1941-1950. https://doi.org/10.1109/TITS.2019.2909064

Esfahani MA, Wang H, Wu K, Yuan S (2020b) OriNet: Robust 3-d orientation estimation with a single particular IMU. IEEE Robot Autom Lett 5(2):399-406. https://doi.org/10.1109/LRA.2019.2959507

Farrell J, Farrell JA (2008) Aided navigation. Electronic engineering. New York: McGraw-Hill.

Fervers F, Bullinger S, Bodensteiner C, Arens M, Stiefelhagen R (2023a) C-BEV: Contrastive bird's eye view training for cross-view image retrieval and 3-dof pose estimation. https://arxiv.org/abs/2312.08060

Fervers F, Bullinger S, Bodensteiner C, Arens M, Stiefelhagen R (2023b) Uncertainty-aware vision-based metric cross-view geolocalization. Paper presented 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; Vancouver, Canada. https://doi.org/10.1109/CVPR52729.2023.02071

Fu S, Hamilton M, Brandt L, Feldman A, Zhang Z, Freeman WT (2024) FeatUp: A model-agnostic framework for features at any resolution. https://arxiv.org/abs/2403.10516

Ge F, Zhang Y, Liu Y, Wang G, Coleman S, Kerr D, Wang L (2024) Multibranch joint representation learning based on information fusion strategy for cross-view geo-localization. IEEE Trans Geosci Remote Sens 62:1-16. https://doi.org/10.1109/TGRS.2024.3378453

Giubilato R, Chiodini S, Pertile M, Debei S (2020) MiniVO: Minimalistic range enhanced monocular system for scale correct pose estimation. IEEE Sensors J 20(20):11874-11886. https://doi.org/10.1109/JSEN.2020.2978334

Goltz GAM, Shiguemori EH, De Campos Velho HF (2016) Position estimation of UAV by image processing with neural networks. Paper presented 2016 10º Congresso Brasileiro de Inteligência Computacional (CBIC). SBIC; Fortaleza, CE. https://doi.org/10.21528/cbic2011-03.6

Gong J, Zhang X, Huang Y, Ren J, Zhang Y (2021) Robust inertial motion tracking through deep sensor fusion across smart earbuds and smartphone. Proc ACM Interact Mob Wearable Ubiquitous Technol 5(2):1-26. https://doi.org/10.1145/3463517

Hegarty CJ, Chatre E (2008) Evolution of the global navigation satellite system (GNSS). Proc IEEE 96(12):1902-1917. https://doi.org/10.1109/jproc.2008.2006090

Hu S, Lee GH (2019) Image-based geo-localization using satellite imagery. Int J Comput Vis 128(5):1205-1219. https://doi.org/10.1007/S11263-019-01186-0

Huang S, Dissanayake G (2007) Convergence and consistency analysis for extended Kalman filter based SLAM. IEEE Trans Robot 23(5):1036-1049. https://doi.org/10.1109/TRO.2007.903811

Huang Y-E, Tsai S, Liu H-Y, Chiang K-W, Tsai M-L, Lee P-L, El-Sheimy N (2023) The development and validation of a tactical grade EGI system for land vehicular navigation applications. Int Arch Photogramm Remote Sens Spatial Inf Sci XLVIII-1/W2-2023:821-828. https://doi.org/10.5194/isprs-archives-xlviii-1-w2-2023-821-2023

Jin R, Zhang G, Hsu L-T, Hu Y (2024) A survey on cooperative positioning using GNSS measurements. IEEE Trans Intell Veh 9(11):7402-7420. https://doi.org/10.1109/tiv.2024.3397879

Joyce KE, Anderson K, Bartolo RE (2021) Of course we fly unmanned – we're women! Drones 5(1):21. https://doi.org/10.3390/drones5010021

Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82(1):35-45. https://doi.org/10.1115/1.3662552

Kartal S, Kaya YB, Nergiz F, Özbağ E, Yilmaz Y, Dar T (2023) Analysis of GDOP based on GEO satellite. Paper presented 2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST). IEEE; Istanbul, Turkey. https://doi.org/10.1109/rast57548.2023.10197979

Ke J, Xu T, Zeng J, Duan Z (2023) Distributed observer-based tracking of multi-agent systems with bounded input amplitudes and rates. Int J Robust Nonlinear Control 34(4):2781-2805. https://doi.org/10.1002/rnc.7108

Kim W-Y, Seo H-I, Seo D-H (2021) Nine-axis IMU-based extended inertial odometry neural network. Expert Syst Appl 178:115075. https://doi.org/10.1016/J.ESWA.2021.115075

Ko NY, Youn W, Choi IH, Song G, Kim TS (2018) Features of invariant extended Kalman filter applied to unmanned aerial vehicle navigation. Sensors 18(9):2855. https://doi.org/10.3390/s18092855

Kopf J, Cohen MF, Lischinski D, Uyttendaele M (2007) Joint bilateral upsampling. ACM Trans Graph 26(3):96. https://doi.org/10.1145/1276377.1276497

Korsah GA, Stentz A, Dias MB (2013) A comprehensive taxonomy for multi-robot task allocation. Int J Robot Res 32(12):1495-1512. https://doi.org/10.1177/0278364913496484

Li F, Zhao C, Ding G, Gong J, Liu C, Zhao F (2012) A reliable and accurate indoor localization method using phone inertial sensors. Paper presented 2012 ACM Conference on Ubiquitous Computing (Ubicomp'12). ACM; New York, USA. https://doi.org/10.1145/2370216.2370280

Li H, Deuser F, Yin W, Luo X, Walther P, Mai G, Huang W, Werner M (2025) Cross-view geolocalization and disaster mapping with street-view and VHR satellite imagery: A case study of hurricane IAN. ISPRS J Photogramm Remote Sens 220:841-854. https://doi.org/10.1016/j.isprsjprs.2025.01.003

Liu B, Teng Y, Huang Q (2017) GDOP minimum in multi-GNSS positioning. Adv Space Res 60(7):1400-1403. https://doi.org/10.1016/j.asr.2017.06.049

Liu C, Zhao J, Sun N (2022) A review of collaborative air-ground robots research. J Intell Robot Syst 106(3). https://doi.org/10.1007/s10846-022-01756-4

Liu D, Bao W, Zhu X, Fei B, Xiao Z, Men T (2022) Vision-aware air-ground cooperative target localization for UAV and UGV. Aerosp Sci Technol 124:107525. https://doi.org/10.1016/j.ast.2022.107525

Liu L, Li H (2019) Lending orientation to neural networks for cross-view geo-localization. Paper presented 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; Long Beach, USA. https://doi.org/10.1109/CVPR.2019.00577

Loianno G, Watterson M, Kumar V (2016) Visual inertial odometry for quadrotors on SE(3). Paper presented 2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE; Stockholm, Sweden. https://doi.org/10.1109/ICRA.2016.7487292

Lupton T, Sukkarieh S (2009) Efficient integration of inertial observations into visual SLAM without initialization. Paper presented 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE; Saint Louis, EUA. https://doi.org/10.1109/IROS.2009.5354267

Miller ID, Cladera F, Smith T, Taylor CJ, Kumar V (2022) Stronger together: Air-ground robotic collaboration using semantics. IEEE Robot Autom Lett 7(4):9643-9650. https://doi.org/10.1109/lra.2022.3191165

Minetto A, Gurrieri A, Dovis F (2020) A cognitive particle filter for collaborative DGNSS positioning. IEEE Access 8:194765-194779. https://doi.org/10.1109/access.2020.3033626

Mohiuddin A, Tarek T, Zweiri Y, Gan D (2020) A survey of single and multi-UAV aerial manipulation. Unmanned Syst 8(2):119-147. https://doi.org/10.1142/s2301385020500089

Nemra A, Aouf N (2009) Robust airborne 3D visual simultaneous localization and mapping with observability and consistency analysis. J Intell Robot Syst 55(4-5):345-376. https://doi.org/10.1007/s10846-008-9306-6

Or B, Klein I (2022) A hybrid model and learning-based adaptive navigation filter. IEEE Trans Instrum Meas 71:1-11. https://doi.org/10.1109/tim.2022.3197775

Parker LE (2009) Multiple Mobile Robot Teams, Path Planning and Motion Coordination. New York: Springer. https://doi.org/10.1007/978-0-387-30440-3_344

Peng K, Du J, Lu F, Sun Q, Dong Y, Zhou P, Hu M (2019) A hybrid genetic algorithm on routing and scheduling for vehicle-assisted multi-drone parcel delivery. IEEE Access 7:49191-49200. https://doi.org/10.1109/access.2019.2910134

Rahman AAA, Jaafar WSWM, Maulud KNA, Noor NM, Mohan M, Cardil A, Silva CA, Che'Ya NN, Naba NI (2019) Applications of drones in emerging economies: A case study of Malaysia. Paper presented 2019 6th International Conference on Space Science and Communication (IconSpace). IEEE; Bahru, Malaysia. https://doi.org/10.1109/ICONSPACE.2019.8905962

Ran H, Sun L, Cheng S, Ma Y, Yan S, Meng S, Shi K, Wen S (2021) A novel cooperative searching architecture for multi-unmanned aerial vehicles under restricted communication. Asian J Control 24(2):510-516. https://doi.org/10.1002/asjc.2517

Regmi K, Borji A (2019) Cross-view image synthesis using geometry-guided conditional GANs. Comput Vis Image Underst 187:102788. https://doi.org/10.1016/J.CVIU.2019.07.008

Regmi K, Shah M (2019) Bridging the domain gap for ground-to-aerial image matching. Paper presented 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE; Seoul, Korea. https://doi.org/10.1109/ICCV.2019.00056

Rigatos GG (2012) Nonlinear Kalman filters and particle filters for integrated navigation of unmanned aerial vehicles. Robot Auton Syst 60(7):978-995. https://doi.org/10.1016/j.robot.2012.03.001

Rodrigues R, Tani M (2023) SemGeo: Semantic keywords for cross-view image geo-localization. Paper presented 2023 ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE; Rhodes Island, Greece. https://doi.org/10.1109/ICASSP49357.2023.10094763

Romero H, Salazar S, Santos O, Lozano R (2013) Visual odometry for autonomous outdoor flight of a quadrotor UAV. Paper presented 2013 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE; Atlanta, USA. https://doi.org/10.1109/icuas.2013.6564748

Roos DR (2018) Aprendizado de máquina aplicado à odometria visual para estimação de posição de veículos aéreos não tripulados (master's thesis). São Paulo: Universidade Federal de São Paulo. https://doi.org/10.20950/1678-2305/bip.2024.51.e910

Savage PG (2007) Savage, volume 2 of Strapdown Analytics. Maple Plain: Strapdown Associates.

Shen C, Zhang Y, Li Z, Gao F, Shen S (2017) Collaborative air-ground target searching in complex environments. Paper presented 2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR). IEEE; Shanghai, China. https://doi.org/10.1109/ssrr.2017.8088168

Shetty A, Gao GX (2019) UAV pose estimation using cross-view geolocalization with satellite imagery. Paper presented 2019 International Conference on Robotics and Automation (ICRA). IEEE; Montreal, Canada. https://doi.org/10.1109/ICRA.2019.8794228

Shi Y, Wu F, Perincherry A, Vora A, Li H (2023) Boosting 3-dof ground-to-satellite camera localization accuracy via geometry-guided cross-view transformer. Paper presented 2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE; Paris, France. https://doi.org/10.1109/ICCV51070.2023.01967

Shi Y, Yu X, Liu L, Campbell D, Koniusz P, Li H (2022) Accurate 3-dof camera geo-localization via ground-to-satellite image matching. IEEE Trans Pattern Anal Mach Intell:1-16. https://doi.org/10.1109/TPAMI.2022.3189702

Shi Y, Yu X, Liu L, Zhang T, Li H (2020) Optimal feature transport for cross-view image geo-localization. Proc AAAI Conf Artif Intell 34(07):11990-11997. https://doi.org/10.1609/AAAI.V34I07.6875

Sottile F, Wymeersch H, Caceres MA, Spirito MA (2011) Hybrid GNSS-terrestrial cooperative positioning based on particle filter. Paper presented 2011 IEEE Global Telecommunications Conference - GLOBECOM. IEEE; Houston, USA. https://doi.org/10.1109/glocom.2011.6134002

Sun S, Melamed D, Kitani K (2021) IDOL: Inertial deep orientation-estimation and localization. Proc AAAI Conf Artif Intell 35(7):6128-6137. https://doi.org/10.1609/aaai.v35i7.16763

Tang H, Niu X, Zhang T, Li Y, Liu J (2022) OdoNet: Untethered speed aiding for vehicle navigation without hardware wheeled odometer. IEEE Sens J 22(12):12197-12208. https://doi.org/10.1109/jsen.2022.3169549

Teixeira BOS, Tôrres LAB, Aguirre LA (2010) Filtragem de Kalman com restrições para sistemas não lineares: revisão e novos resultados. Sba: Controle Automação Soc Bras Autom 21(2):127-146. https://doi.org/10.1590/S0103-17592010000200003

Toker A, Zhou Q, Maximov M, Leal-Taixe L (2021) Coming down to earth: Satellite-to-street view synthesis for geo-localization. Paper presented 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; Nashville, USA. https://doi.org/10.1109/CVPR46437.2021.00642

Torres VAMF, Jaimes BRA, Ribeiro ES, Braga MT, Shiguemori EH, Velho HFC, Torres LCB, Braga AP (2020) Combined weightless neural network FPGA architecture for deforestation surveillance and visual navigation of UAVs. Eng Appl Artif Intell 87:103227. https://doi.org/10.1016/J.ENGAPPAI.2019.08.021

Ulmer MW, Thomas BW (2018) Same-day delivery with heterogeneous fleets of drones and vehicles. Networks 72(4): 475-505. https://doi.org/10.1002/net.21855

Velusamy P, Rajendran S, Mahendran RK, Naseer S, Shafiq M, Choi J-G (2021) Unmanned aerial vehicles (UAV) in precision agriculture: Applications and challenges. Energies 15(1):217. https://doi.org/10.3390/en15010217

Wagstaff B, Peretroukhin V, Kelly J (2020) Robust data-driven zero-velocity detection for foot-mounted inertial navigation. IEEE Sens J 20(2):957-967. https://doi.org/10.1109/jsen.2019.2944412

Wang N (2021) "As it is Africa, it is ok"? Ethical considerations of development use of drones for delivery in Malawi. IEEE Trans Technol Soc 2(1):20-30. https://doi.org/10.1109/TTS.2021.3058669

Wang P, Yang Z, Chen X, Xu H (2023) A Transformer-Based Method for UAV-View Geo-Localization. Cham: Springer. https://doi.org/10.1007/978-3-031-44223-0_27

Wang Q, Luo H, Ye L, Men A, Zhao F, Huang Y, Ou C (2019) Pedestrian heading estimation based on spatial transformer networks and hierarchical LSTM. IEEE Access 7:162309-162322. https://doi.org/10.1109/access.2019.2950728

Wang T, Fan S, Liu D, Sun C (2022) Transformer-guided convolutional neural network for cross-view geolocalization. https://arxiv.org/abs/2204.09967

Wang W, Chen W, Qiu Q, Chen L, Wu B, Lin B, He X, Liu W (2024) CrossFormer++: A versatile vision transformer hinging on cross-scale attention. IEEE Trans Pattern Anal Mach Intell 46(5):3123-3136. https://doi.org/10.1109/TPAMI.2023.3341806

Wang Y, Cheng H, Meng MQ-H (2020) Pedestrian motion tracking by using inertial sensors on the smartphone. Paper presented 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE; Las Vegas, EUA. https://doi.org/10.1109/iros45743.2020.9341173

Wang Y, Kuang J, Li Y, Niu X (2022) Magnetic field-enhanced learning-based inertial odometry for indoor pedestrian. IEEE Trans Instrum Meas 71:1-13. https://doi.org/10.1109/tim.2022.3186358

Wang Y, Kuang J, Niu X, Liu J (2023) LLIO: Lightweight learned inertial odometer. IEEE Internet Things J 10(3):2508-2518. https://doi.org/10.1109/jiot.2022.3214087

Wen W, Bai X, Zhang G, Chen S, Yuan F, Hsu L-T (2020) Multi-agent collaborative GNSS/camera/INS integration aided by inter-ranging for vehicular navigation in urban areas. IEEE Access 8:124323-124338. https://doi.org/10.1109/access.2020.3006210

Williams J (2024) UAV survey mapping of illegal deforestation in Madagascar. Plants People Planet 6(6):1413-1424. https://doi.org/10.1002/ppp3.10533

Wilson D, Zhang X, Sultani W, Wshah S (2023) Image and object geo-localization. Int J Comput Vis 132(4):1350-1392. https://doi.org/10.1007/s11263-023-01942-3

Workman S, Souvenir R, Jacobs N (2015) Wide-area image geolocalization with aerial reference imagery. Paper presented 2015 IEEE International Conference on Computer Vision (ICCV). IEEE; Santiago, Chile. https://doi.org/10.1109/ICCV.2015.451

Wu S, Tang H, Jing X-Y, Qian J, Sebe N, Yan Y, Zhang Q (2022) Cross-view panorama image synthesis with progressive attention GANs. Pattern Recognit 131:108884. https://doi.org/10.1016/J.PATCOG.2022.108884

Xavier NAZ, Shiguemori EH, Maximo MROA (2025a) UAV inertial navigation using deep neural networks in simulated environments. Eng Anal Bound Elem 179:106376. https://doi.org/10.1016/j.enganabound.2025.106376

Xavier NAZ, Shiguemori EH, Maximo MROA, Shah M (2025b) A guided approach for cross-view geolocalization estimation with land cover semantic segmentation. Biomim Intell Robot 5(2):100208. https://doi.org/10.1016/j.birob.2024.100208

Xia Y, Song M, Zhang J, Hu C (2018) An autonomously navigation system for forestry quadrotor within GPS-denied below-canopy environment. Paper presented 2018 IEEE CSAA Guidance, Navigation and Control Conference (CGNCC). IEEE; Xiamen, China. https://doi.org/10.1109/GNCC42960.2018.9019136

Xia Z, Booij O, Kooij JFP (2024) Convolutional cross-view pose estimation. IEEE Trans Pattern Anal Mach Intell 46(5): 3813-3831. https://doi.org/10.1109/TPAMI.2023.3346924

Xia Z, Booij O, Manfredi M, Kooij JFP (2022) Visual cross-view metric localization with dense uncertainty estimates. Tel Aviv: Springer. European Conference on Computer Vision (ECCV); p. 90-106. https://doi.org/10.48550/ARXIV.2208.08519

Xie X, Yang T, Ning Y, Zhang F, Zhang Y (2021) A monocular visual odometry method based on virtual-real hybrid map in low-texture outdoor environment. Sensors 21(10):3394. https://doi.org/10.3390/S21103394

Xu H, Wang C, Bo Y, Jiang C, Liu Y, Yang S, Lai W (2022) An aerial and ground multi-agent cooperative location framework in GNSS-challenged environments. Remote Sens 14(19):5055. https://doi.org/10.3390/rs14195055

Xu Y, Wei Y, Wang D, Jiang K, Deng H (2023) Multi-UAV path planning in GPS and communication denial environment. Sensors 23(6):2997. https://doi.org/10.3390/S23062997

Yan H, Shan Q, Furukawa Y (2018) RIDI: Robust IMU Double Integration. Cham: Springer. https://doi.org/10.1007/978-3-030-01261-8_38

Yan Z, Jouandeau N, Cherif AA (2013) A survey and analysis of multi-robot coordination. Int J Adv Robot Syst 10(12). https://doi.org/10.5772/57313

Yang M, Sun X, Jia F, Rushworth A, Dong X, Zhang S, Fang Z, Yang G, Liu B (2022) Sensors and sensor fusion methodologies for indoor odometry: A review. Polymers 14(10):2019. https://doi.org/10.3390/polym14102019

Ye Z, Bao C, Liu X, Bao H, Cui Z, Zhang G (2022) Crossview mapping with graph-based geolocalization on city-scale street maps. Paper presented 2022 International Conference on Robotics and Automation (ICRA). IEEE; Philadelphia, USA. https://doi.org/10.1109/ICRA46639.2022.9811743

Yi S, Jin X, Wang Z, Liu Z, Zorzi M (2025) Data-driven robust UAV position estimation in GPS signal-challenged environment. https://arxiv.org/abs/2504.07842

Yu X, Liu B, Lan X, Xiao Z, Lin S, Yan B, Zhou L (2019) AZUPT: Adaptive zero velocity update based on neural networks for pedestrian tracking. Paper presented 2019 IEEE Global Communications Conference (GLOBECOM). IEEE; Waikoloa, USA. https://doi.org/10.1109/globecom38437.2019.9014070

Yuan D, Maire F, Dayoub F (2024) Cross-attention between satellite and ground views for enhanced fine-grained robot geo-localization. Paper presented 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE; Waikoloa, USA. https://doi.org/10.1109/WACV57701.2024.00128

Yue P, Xin J, Huang Y, Zhao J, Zhang C, Chen W, Shan M (2025) UAV autonomous navigation system based on air-ground collaboration in GPS-denied environments. Drones 9(6):442. https://doi.org/10.3390/drones9060442

Zhang L, Gao F, Deng F, Xi L, Chen J (2023) Distributed estimation of a layered architecture for collaborative air-ground target geolocation in outdoor environments. IEEE Trans Ind Electron 70(3):2822-2832. https://doi.org/10.1109/tie.2022.3165245

Zhang M, Zhang M, Chen Y, Li M (2021) IMU data processing for inertial aided navigation: A recurrent neural network based approach. Paper presented 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE; Xi'an, China. https://doi.org/10.1109/icra48506.2021.9561172

Zhang Q, Zhu Y (2024) Aligning geometric spatial layout in cross-view geo-localization via feature recombination. Proceedings of the AAAI Conference on Artificial Intelligence 38(7):7251-7259. https://doi.org/10.1609/AAAI.V38I7.28554

Zhang X, Li X, Sultani W, Zhou Y, Wshah S (2023) Cross-view geo-localization via learning disentangled geometric layout correspondence. Proceedings of the AAAI Conference on Artificial Intelligence 37(3):3480-3488. https://doi.org/10.1609/AAAI.V37I3.25457

Zhang Y, Shi Y, Wang S, Vora A, Perincherry A, Chen Y, Li H (2024) Increasing SLAM pose accuracy by ground-to-satellite image registration. Paper presented 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE; Yokohama, Japan. https://doi.org/10.1109/ICRA57147.2024.10611079

Zhao J, Zhai Q, Zhao P, Huang R, Cheng H (2023) Co-visual pattern-augmented generative transformer learning for automobile geo-localization. Remote Sens 15(9):2221. https://doi.org/10.3390/RS15092221

Zhou B, Gu Z, Gu F, Wu P, Yang C, Liu X, Li L, Li Y, Li Q (2022) DeepVIP: Deep learning-based vehicle indoor positioning using smartphones. IEEE Trans Veh Technol 71(12):13299-13309. https://doi.org/10.1109/tvt.2022.3199507

Zhou B, Krahenbuhl P (2022) Cross-view transformers for real-time map-view semantic segmentation. Paper presented 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; New Orleans, USA. https://doi.org/10.1109/CVPR52688.2022.01339

Zhou L, Jiang Y, Jia H, Zhang L, Xu F, Tian Y, Ma Z, Liu X, Guo S, Wu Y, *et al.* (2024) UAV vision-based crack quantification and visualization of bridges: system design and engineering application. Struct Health Monit 24(2):1083-1100. https://doi.org/10.1177/14759217241251778

Zhu S, Shah M, Chen C (2022) TransGeo: Transformer is all you need for cross-view image geo-localization. Paper presented 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; New Orleans, USA. https://doi.org/10.1109/CVPR52688.2022.00123

Zhu S, Yang T, Chen C (2021a) Revisiting street-to-aerial view image geo-localization and orientation estimation. Paper presented 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE; Waikoloa, USA. https://doi.org/10.1109/WACV48630.2021.00080

Zhu S, Yang T, Chen C (2021b) VIGOR: Cross-view image geo-localization beyond one-to-one retrieval. Paper presented 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; Nashville, USA. https://doi.org/10.1109/CVPR46437.2021.00364

Zhu Y, Sun B, Lu X, Jia S (2022) Geographic semantic network for cross-view image geo-localization. IEEE Trans Geosci Remote Sens 60:1-15. https://doi.org/10.1109/TGRS.2021.3121337

Zocca S, Minetto A, Dovis F (2021) Adaptive bayesian state estimation integrating non-stationary DGNSS inter-agent distances. Paper presented 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring). IEEE; Helsinki, Finland. https://doi.org/10.1109/vtc2021-spring51267.2021.9448952

Zou Z, Huang T, Ye L, Song K (2020) CNN based adaptive Kalman filter in high-dynamic condition for low-cost navigation system on highspeed UAV. Paper presented 2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS). IEEE; Singapore, Singapore. https://doi.org/10.1109/acirs49895.2020.9162601