

Article

UAV Sensor Data Fusion for Localization Using Adaptive Multiscale Feature Matching Mechanisms Under GPS-Deprived Environment

Yu-Shun Wang *  and Chia-Hao Chang

Avionics and Electrical Engineering Section, Aeronautical Systems Research Division, National Chung-Shan Institute of Science and Technology, Taichung 32546, Taiwan

* Correspondence: yu.shun.wang.tw@gmail.com

Abstract

The application of unmanned vehicles in civilian and military fields is increasingly widespread. Traditionally, unmanned vehicles primarily rely on Global Positioning Systems (GPSs) for positioning; however, GPS signals can be limited or completely lost in conditions such as building obstructions, indoor environments, or electronic interference. In addition, countries are actively developing GPS jamming and deception technologies for military applications, making precise positioning and navigation of unmanned vehicles in GPS-denied or constrained environments a critical issue that needs to be addressed. In this work, authors propose a method based on Visual-Inertial Odometry (VIO), integrating the extended Kalman filter (EKF), an Inertial Measurement Unit (IMU), optical flow, and feature matching to achieve drone localization in GPS-denied environments. The proposed method uses the heading angle and acceleration data obtained from the IMU as the state prediction for the EKF, and estimates relative displacement using optical flow. It further corrects the optical flow calculation errors through IMU rotation compensation, enhancing the robustness of visual odometry. Additionally, when re-selecting feature points for optical flow, it combines a KAZE feature matching technique for global position correction, reducing drift errors caused by long-duration flight. The authors also employ an adaptive noise adjustment strategy that dynamically adjusts the internal state and measurement noise matrices of the EKF based on the rate of change in heading angle and feature matching reliability, allowing the drone to maintain stable positioning in various flight conditions. According to the simulation results, the proposed method is able to effectively estimate the flight trajectory of drones without GPS. Compared to results that rely solely on optical flow or feature matching, it significantly reduces cumulative errors. This makes it suitable for urban environments, forest areas, and military applications where GPS signals are limited, providing a reliable solution for autonomous navigation and positioning of drones.

Keywords: visual-inertial odometry; EKF; optical flow method; drone image localization



Academic Editors: Bosko Rasuo and Lamei Zhang

Received: 13 September 2025

Revised: 18 November 2025

Accepted: 23 November 2025

Published: 25 November 2025

Citation: Wang, Y.-S.; Chang, C.-H. UAV Sensor Data Fusion for Localization Using Adaptive Multiscale Feature Matching Mechanisms Under GPS-Deprived Environment. *Aerospace* **2025**, *12*, 1048. <https://doi.org/10.3390/aerospace12121048>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background

Unmanned Aerial Vehicles (UAVs) have become essential platforms in various civil and military applications, including reconnaissance, logistics, environmental monitoring, and surveying. Their capability for autonomous flight and flexible deployment has made navigation and localization technologies a central topic of UAV research. Traditionally, most

UAVs rely on the Global Positioning System (GPS) for positioning and navigation. However, when operating in environments with terrain obstructions, dense urban structures, or electronic interference—such as under bridges, inside tunnels, or during GPS jamming—satellite signals may become unreliable or completely unavailable. This limitation severely restricts the performance and safety of UAV autonomous missions.

Many studies have therefore explored strategies for detecting, compensating, or mitigating GPS interference [1–3], reflecting the growing importance of improving UAV localization capabilities under GPS-denied conditions. To address these challenges, researchers have integrated other onboard sensors such as the Inertial Measurement Unit (IMU) and cameras, which provide complementary sensing modalities. IMUs can measure angular velocity and linear acceleration at high frequencies, enabling continuous state prediction, but their performance deteriorates over time due to bias drift. Vision sensors, on the other hand, can provide rich environmental information for position correction through Visual Odometry (VO) or Simultaneous Localization and Mapping (SLAM).

With the rapid development of computer vision and image-processing technologies, vision-based localization has become an effective supplement or even an alternative to GPS. Early work by Lowe introduced the classical Scale-Invariant Feature Transform (SIFT) algorithm [4], which maintains feature stability under scale, rotation, and illumination changes, making it widely used in object recognition, tracking, and image stitching. Subsequently, Bay et al. proposed Speeded-Up Robust Features (SURF) [5], and Rublee et al. introduced Oriented FAST and Rotated BRIEF (ORB) [6], both of which improved computational efficiency and feature-matching accuracy. However, these methods still face limitations when dealing with nonlinear deformations and environmental noise. To address these issues, Alcantarilla et al. proposed the KAZE algorithm in 2012 [7], which innovatively employs a nonlinear scale space to construct image pyramids, effectively preserving edges and fine details. Experimental results demonstrate that KAZE achieves higher matching stability and accuracy than traditional methods under variations in lighting, scale, and perspective. By estimating geometric transformations between consecutive frames, these feature-based approaches enable UAVs to estimate relative motion and reconstruct the surrounding three-dimensional environment.

For lightweight UAV platforms, camera configurations can generally be categorized into monocular and stereo systems. Stereo vision uses two synchronized cameras with a fixed baseline to directly compute scene depth through image disparity, achieving high spatial accuracy. However, stereo systems require precise calibration, rigid baseline maintenance, and high computational resources, which increase payload and power consumption. Monocular vision, in contrast, uses a single camera and offers the advantages of low cost, compactness, and easy integration, making it more suitable for onboard applications. Nevertheless, monocular systems suffer from scale ambiguity and higher sensitivity to lighting and texture conditions.

To better understand the development of UAV visual-inertial localization technologies, recent studies on stereo- and monocular-based systems are reviewed in the following section.

1.2. Literature Review

Many studies have applied computer vision techniques to the positioning and autonomous navigation of unmanned aerial vehicles (UAVs). In 2014, Forster et al. proposed the Semi-Direct Visual Odometry (SVO) method [8], which uses pixel intensity for computation and achieves a balance between speed and accuracy for real-time UAV navigation. Qin et al. later introduced VINS-Mono [9], a robust monocular visual-inertial estimator that integrates image and IMU data to improve motion estimation accuracy under dynamic

conditions. Weiss et al. [10] combined monocular vision with Simultaneous Localization and Mapping (SLAM) to enable UAV navigation in GPS-denied environments, while Xiao et al. [11] applied monocular feature extraction and camera pose estimation to indoor mobile robot localization.

Li and Mourikis [12] presented a landmark analysis of monocular EKF-based visual-inertial odometry, showing that filter inconsistency and unmodeled correlations can severely degrade long-term accuracy. Their work established the importance of maintaining observability and ensuring consistent state updates in EKF-based VIO systems, providing theoretical support for lightweight onboard monocular frameworks such as the one adopted in this study. OpenVINS [13] further provides a modern EKF-based visual-inertial odometry platform with on-manifold state representation, FEJ consistency maintenance, and a modular feature-tracking front-end. Its monocular configuration aligns well with lightweight VIO designs, demonstrating the effectiveness of EKF-based fusion for real-time onboard applications. Although monocular systems are computationally efficient, they inherently suffer from scale ambiguity and increased sensitivity to illumination and texture variations, which limit long-term robustness.

Beyond purely monocular pipelines, Qin et al. [14] proposed a general optimization-based multi-sensor fusion framework that integrates local VO/VIO with global measurements to suppress accumulated drift. Their findings highlight the benefit of periodic global corrections and motivate the GPS-free, feature-matching-based drift-suppression strategy adopted in this study. Stereo-based systems also offer valuable insights. ORB-SLAM2 [15] demonstrates how stereo depth can effectively resolve scale ambiguity and how loop closing can eliminate long-term drift in mapped environments. Although full SLAM with loop closure lies beyond the scope of this lightweight VIO framework, such mechanisms indicate promising directions for future extensions.

Recent stereo VIO systems such as Fu and Lu [16] show that direct sparse photometric alignment tightly fused with IMU pre-integration can achieve high accuracy on benchmarks including KITTI and EuRoC, illustrating the upper performance bound achievable with stereo sensing and computationally intensive optimization. These results provide a useful contrast to the lightweight monocular EKF formulation explored in this work. Finally, Wang et al. [17] reported that even stereo visual-inertial localization suffers substantial drift in large-scale GPS-denied environments without external references. Their use of fiducial marker corrections restored global consistency, reinforcing the broader observation that pure VIO pipelines struggle in long-duration GPS-denied scenarios and underscoring the need for lightweight drift-suppression mechanisms such as those developed in this study.

More recently, Xue et al. [18] proposed a feature-matching-based visual positioning algorithm using Accelerated-KAZE (AKAZE) descriptors to improve matching robustness in complex scenes. Similarly, Wang, Xu, Cheng, and other researchers [19–21] emphasized multi-sensor fusion strategies that integrate visual, inertial, and optical flow data to enhance localization reliability when GPS signals are unavailable. These developments highlight the growing trend toward adaptive and hybrid fusion architectures for UAV navigation. Building upon the authors' previous work [22], this study presents an enhanced Visual-Inertial Odometry (VIO) framework that integrates the extended Kalman filter (EKF), Inertial Measurement Unit (IMU), optical flow (OF), and feature matching (FM) modules to achieve accurate localization of lightweight UAVs in GPS-denied environments. A dynamic weighting mechanism is introduced to adaptively balance the contributions of visual and inertial inputs according to flight conditions, while a global feature-matching module is activated to compensate accumulated drift when optical flow tracking deteriorates.

Designed to satisfy the practical requirements of small-scale UAV platforms, the proposed system focuses on low-cost monocular sensing and ease of system inte-

gration, enabling real-time onboard implementation under constrained payload and computational resources.

2. Inertia Visual Localization Algorithm

2.1. System Architecture

The system architecture is shown in Figure 1, consisting of a visual-inertial positioning module that integrates optical flow and IMU information using an EKF, along with feature matching for correction. Based on confidence levels, the algorithm and sensor influence are dynamically adjusted to achieve autonomous drone positioning in a GPS-denied environment. Optical flow is used to estimate the drone's relative motion by calculating pixel displacement between images using the Lucas–Kanade method, which is then converted to ground displacement information. The IMU provides flight attitude (roll, pitch, and yaw) and acceleration data, allowing for rotation compensation to reduce visual drift errors. The EKF uses IMU-measured information as the state transition model to predict the drone's position and velocity changes, while integrating the displacement estimated from optical flow as observations. Noise covariance (Q , R) is dynamically adjusted based on the uncertainty of the optical flow and the reliability of the IMU to enhance robustness.

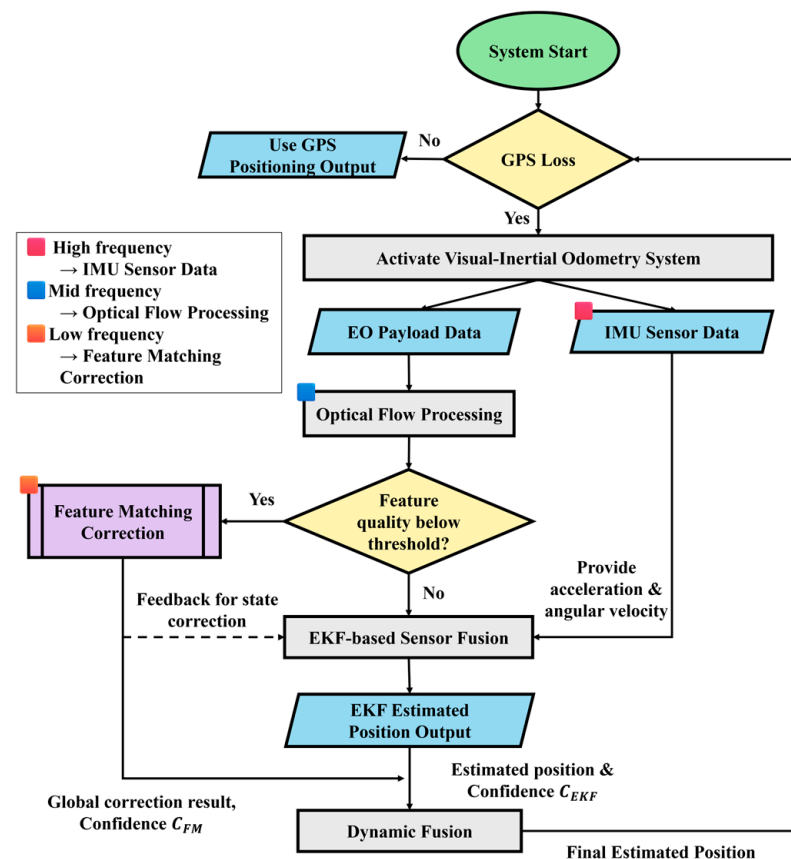


Figure 1. Flowchart of the proposed algorithm.

2.2. Proposed Algorithm

A whole algorithm procedure is given in Section 2.2. The proposed system maintains continuous position estimation through a hybrid mechanism that combines high-frequency inertial propagation, mid-frequency visual updates, and low-frequency global map correction. During each frame interval, the EKF predicts the UAV's state based on IMU acceleration and angular velocity, ensuring temporal continuity of motion estimation. The

optical flow module provides frame-to-frame relative displacement, which serves as the observation update of the EKF to correct short-term drift.

When the feature-tracking quality drops or the accumulated drift exceeds a predefined threshold, the global feature-matching module is triggered. It performs image registration between the current aerial frame and the offline map tiles to realign the VIO-estimated position to the global coordinate reference. The resulting correction is dynamically fused with the formula according to the confidence evaluation mechanism described in the Dynamic Weighted Fusion Strategy section.

This hierarchical fusion strategy enables the system to continuously maintain a consistent position estimation, the IMU ensures high-rate prediction continuity, the optical flow refines short-term accuracy, and the global map matching corrects long-term drift. Together, these modules form a closed-loop localization framework capable of stable positioning result under GPS-denied conditions.

The proposed algorithm can be divided into three main parts: optical flow method, EKF design, and feature matching correction. The content and implementation methods are as follows:

2.2.1. Lucas–Kanade Optical Flow

In this study, optical flow estimation is a crucial component of VIO for drones, primarily used to estimate the visual displacement of the camera between consecutive frames. We employed the Lucas–Kanade optical flow method to track image feature points and compensated the optical flow with IMU information. The author applied sparse optical flow to track the motion of the drone’s feature points and estimated the body movement based on their changes. The overall process is as follows:

(1) Feature Point Selection:

After receiving optical payload data, apply KAZE + Harris feature detection methods to extract robust features from different perspectives in the image $p_t = (x_t, y_t)$.

(2) Optical flow tracking:

Calculate the displacement of feature points in a new frame using the Lucas–Kanade method $p_{t+1} = (x_{t+1}, y_{t+1})$.

(3) RANSAC filtering:

Use RANSAC (Random Sample Consensus) to remove erroneous match point pairs.

(4) Displacement estimation:

Calculate the pixel displacement between consecutive frames according to Equation (1).

$$\Delta p = (\Delta x, \Delta y) = \frac{1}{N} \sum_{i=1}^N p_{t+1}^i - p_t^i \quad (1)$$

where N is the number of feature points filtered by the RANSAC algorithm.

(5) Attitude Compensation:

When estimating the visual motion of a drone, it is essential to consider the impact of attitude on optical flow measurements. This study assumes that the camera’s mounting position aligns with the drone’s z -axis rotation, so lens offset compensation is not considered. To compensate for posture changes, the optical flow displacement needs to be converted from image coordinates to the camera’s normalized coordinate system. The relationship

between image coordinates P_{img} and camera coordinates P_{cam} is expressed in Equation (2) based on the camera's intrinsic parameter matrix K .

$$P_{cam} = K^{-1}P_{img} \quad (2)$$

The camera intrinsic matrix K_{cam} is given by Equation (3), which defines the mapping between the 3D camera coordinate system and the 2D image plane. Where (c_x, c_y) represents the center coordinates of the image.

$$K_{cam} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Through the inverse intrinsic camera matrix K^{-1} , convert optical flow to motion in camera coordinates using Equation (4).

$$\Delta P_{cam} = K_{cam}^{-1} \Delta P_{img} \quad (4)$$

Between $t - 1$ and t , the drone's attitude changes to $\{\Delta\psi \ \Delta\theta \ \Delta\phi\}$, representing changes in heading, pitch, and roll angles, defining the coordinate rotation matrix R as Equation (5).

$$R = R_z(\Delta\psi)R_y(\Delta\theta)R_x(\Delta\phi) \quad (5)$$

By applying a rotation matrix, the optical flow displacement can be compensated to the state before the drone's rotation, as shown in Equation (6).

$$\Delta P'_{cam} = R^{-1} \Delta P_{cam} \quad (6)$$

Finally, the camera coordinates are converted back to image coordinates using Equations (7) and (8).

$$\Delta P'_{img} = K \Delta P'_{cam} \quad (7)$$

$$\Delta x' = \frac{\Delta P'_{img}(1)}{\Delta P'_{img}(3)}, \Delta y' = \frac{\Delta P'_{img}(2)}{\Delta P'_{img}(3)} \quad (8)$$

2.2.2. Adaptive EKF Design

After obtaining the position of the optical flow estimation, this study uses an EKF to fuse optical flow information with IMU information, transforming body coordinates to inertial coordinates through coordinate transformation and selecting a state vector (Equation (9)), where X , Y , and Z are the positions in the inertial coordinates; ψ , θ , ϕ , are the yaw, pitch, and roll angles; V_x , V_y , and V_z are the velocities in the inertial coordinates.

$$X = [X^I, Y^I, Z^I, \psi, \theta, \phi, V_x^I, V_y^I, V_z^I]^T \quad (9)$$

According to the drone kinematics model, the state change is affected by IMU measurements, and the state transition equation in discrete time can be expressed as Equation (10), where X_k is the state vector, $u_k = [a^b, \omega]$ is the control input provided by IMU (body-fixed acceleration and angular velocity), and w_k is the process noise, which follows a zero-mean Gaussian distribution.

$$X_{k+1} = f(X_k, u_k) + w_k \quad (10)$$

The specific state transition model is as follows (Equation (11)):

$$\begin{aligned}
 X_{k+1}^I &= X_k^I + V_x^I \cdot dt \\
 Y_{k+1}^I &= Y_k^I + V_y^I \cdot dt \\
 Z_{k+1}^I &= Z_k^I + V_z^I \cdot dt \\
 \psi_{k+1} &= \psi_k + \omega_\psi \cdot dt \\
 \theta_{k+1} &= \theta_k + \omega_\theta \cdot dt \\
 V_{x_{k+1}}^I &= V_{x_k}^I + A_{x_k}^I \cdot dt \\
 V_{y_{k+1}}^I &= V_{y_k}^I + A_{y_k}^I \cdot dt \\
 V_{z_{k+1}}^I &= V_{z_k}^I + A_{z_k}^I \cdot dt
 \end{aligned} \tag{11}$$

where $A^I = R_B^I \cdot A^B$, A^I is the acceleration in inertial coordinates, A^B is the acceleration in body coordinates, and R_B^I is the rotation matrix. The state transition function F is then defined as (Equation (12)).

$$F = \begin{bmatrix} I_3 & 0 & \text{diag}(dt, dt, dt) \\ 0 & I_3 & 0 \\ 0 & 0 & I_3 \end{bmatrix} \tag{12}$$

In the meantime, the error covariance matrix P is updated according to Equation (13), where Q is the system noise covariance matrix:

$$P_{k|k-1} = F P_{k-1} F^T + Q \tag{13}$$

In the update stage, the predicted state is combined with the flow data, assuming a measurement vector z (Equation (14)) comprising the horizontal displacement estimated from optic flow and the height (from a barometer):

$$z = \begin{bmatrix} X_{meas} \\ Y_{meas} \\ Z_{meas} \end{bmatrix} \tag{14}$$

The corresponding measurement matrix H is given in Equation (15).

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{15}$$

Calculate the residual using Equation (16).

$$y = z - H x_{k|k-1} \tag{16}$$

Compute the Kalman gain K by Equation (17).

$$K = P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1} \tag{17}$$

where R is the measurement noise covariance matrix; then, update the state estimate using Equation (18).

$$x_k = x_{k|k-1} + K y \tag{18}$$

Finally, update the error covariance matrix P with Equation (19).

$$P_k = (I - K H) P_{k|k-1} \tag{19}$$

The system noise covariance matrix Q mainly describes the uncertainty of the system model, while the measurement noise covariance matrix R mainly describes the uncertainty of the optical flow measurement. According to the flight situation and measurement conditions of the drone, dynamic adjustments are made. When the drone turns, the reliability of the optical flow often decreases. Therefore, this paper adjusts the position noise of Q based on the turning angular velocity ψ to trust the IMU information more when the drone rotates faster, reducing the impact of optical flow errors:

$$Q = \text{diag}(f(\psi), f(\psi), 0.1, 0.001 \cdot I_6)$$

$$f(\dot{\psi}) = \frac{0.01}{100|\dot{\psi}|} \quad (20)$$

Adjust R according to Equation (21), using $\dot{\psi}$ and the number of inlier feature points N_{inlier} as the criteria for measuring the noise covariance matrix R elements. When the drone turns (with a large turning angular velocity) or the number of matched feature points is small, the displacement calculated by optical flow may be less accurate, so it is necessary to dynamically increase the measurement noise and reduce the confidence in the optical flow results.

$$R = \begin{bmatrix} f(\dot{\psi}, N_{inlier}) & 0 \\ 0 & f(\dot{\psi}, N_{inlier}) \end{bmatrix} \quad (21)$$

$$f(\dot{\psi}, N_{inlier}) = \frac{(15|\dot{\psi}| + \frac{50}{N_{inlier}})}{r_{inlier}}$$

The visual-inertial odometry algorithm takes the first two elements of the state vector X as the position of the drone in the inertial coordinate system, measured in pixels and projected onto an offline map, and can also convert pixel coordinates to latitude and longitude based on known information.

2.2.3. Feature Matching Calibration

During the algorithm processing, when there are too few feature points in the frame in optical flow method and feature points are reselected, feature matching is performed at the same time, matching the current frame with the offline map and determining the weight of this correction based on the confidence of the match. The concept of feature matching calibration is shown in Figure 2.

Global error correction is achieved through feature matching using offline maps and aerial photos. First, obtain the satellite image of the target area (Figure 3), set as I_{map} and obtain the longitude and latitude coordinates of the four corner points (assuming a flat earth and located in the northern hemisphere), according to Equation (22).

$$\Delta Lon = \frac{Lon_R - Lon_L}{W_{map}}$$

$$\Delta Lat = \frac{Lat_T - Lat_B}{H_{map}} \quad (22)$$

The latitude and longitude change per pixel (ΔLat , ΔLon) on the map are used to achieve the conversion between latitude and longitude and pixel coordinates, where Lon_R , Lon_L , Lat_T , and Lat_B are the longitudes of the east and west sides of the offline map, and the latitudes of the north and south, W_{map} and H_{map} , are the horizontal and vertical resolutions of the map.

Divide an offline map according to Formula (23) and obtain a set of $N_{row} \times N_{col}$ images.

$$N_{row} = \left\lceil \frac{H_{map}}{S_{crop}} \right\rceil \times S_{crop}$$

$$N_{col} = \left\lceil \frac{W_{map}}{S_{crop}} \right\rceil \times S_{crop} \quad (23)$$

where $\lceil \cdot \rceil$ denotes rounding up, each cropped image is sized $S_{crop} \times S_{crop}$, and after numbering, they are stored on the onboard computer storage device (Figure 4). The corresponding image number and surrounding small map are selected and stitched together according to the estimated location in subsequent algorithms (Figure 5). The purpose of cropping is to improve the efficiency of reading offline maps.

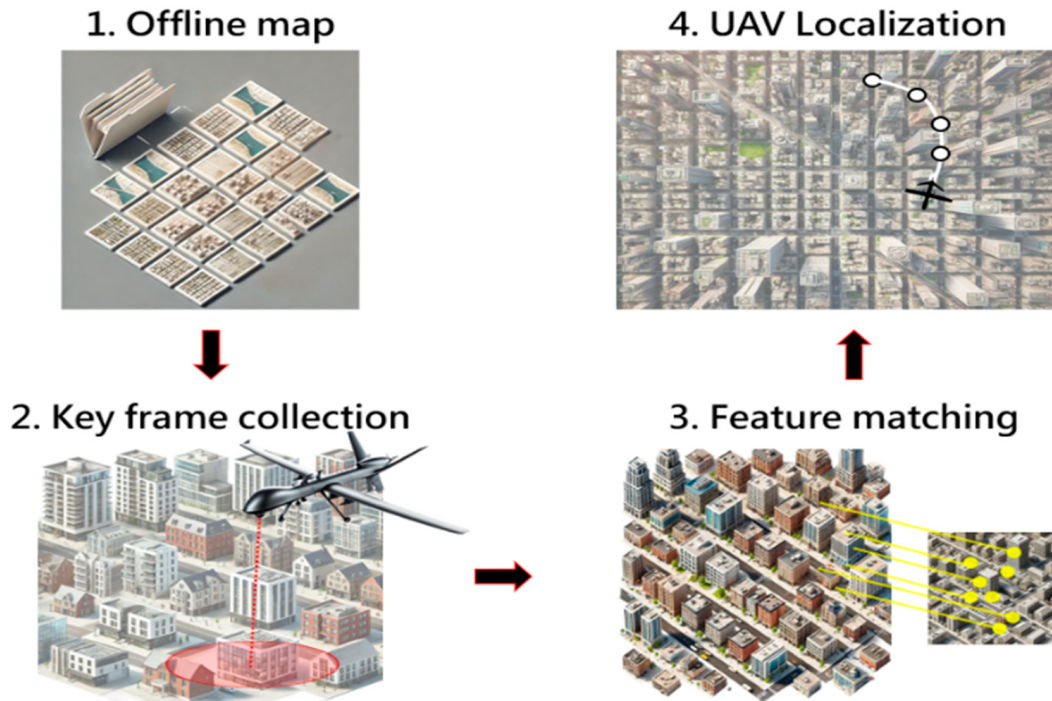


Figure 2. Feature matching calibration.



Figure 3. Offline map (4103×3894).



Figure 4. Offline map fragmentation.

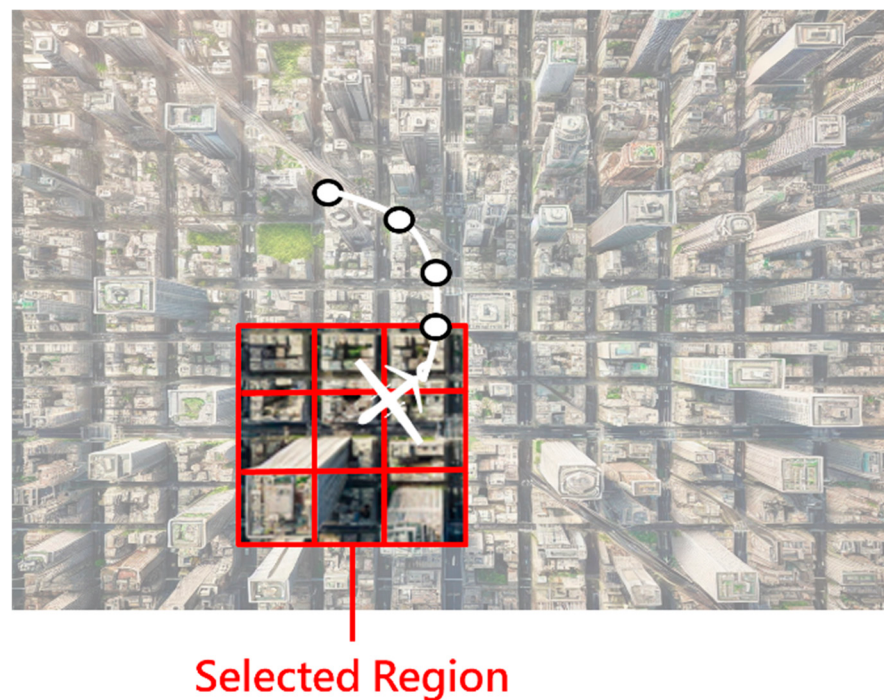


Figure 5. Offline map stitching (white dots represent previously visited waypoints).

When the visual-inertial odometer has accumulated error, perform global feature matching correction according to the following steps:

a. Feature capture

According to Section 1, the small segment optical flow method, due to the lack of feature points in the image, the current frame (Figure 6) is captured and subsequent image preprocessing is performed.



Figure 6. Current image frame (816 × 459).

b. Selecting the stitching image number

Based on the estimated location. Initially, the algorithm uses the last available GPS data as the starting point, and subsequently, it uses the calculated location from the algorithm, according to Equations (24) and (25).

$$\begin{aligned} Bx &= \left\lceil \frac{P_x}{S_{crop}} \right\rceil \\ By &= \left\lceil \frac{P_y}{S_{crop}} \right\rceil \end{aligned} \quad (24)$$

$$M_{ID} = (N_{row}/S_{crop}) \times (Bx - 1) \times By \quad (25)$$

Calculate the picture ID M_{ID} of the current drone location, and select the corresponding picture and surrounding map for stitching based on M_{ID} (Figure 7), where P_x , P_y are the current position estimated by visual-inertial odometry in pixels.

Stitched Image



Figure 7. Stitched offline map (300 × 300).

c. Image Preprocessing

After obtaining the aerial image I_s perform basic image preprocessing on the stitched offline map I_r including adjusting resolution, adjusting photo orientation based on drone heading angle, and grayscale processing. Then, calculate the normalized grayscale histograms $h_r(i)$ and $h_s(i)$ for I_r and I_s , respectively. Where $i \in \{0, 1, \dots, 255\}$ denotes the pixel intensity value. After normalizing the histograms, Equation (26) is obtained.

$$\begin{aligned}\tilde{h}_r(i) &= \frac{h_r(i)}{N_r} \\ \tilde{h}_s(i) &= \frac{h_s(i)}{N_s}\end{aligned}\quad (26)$$

where N_r and N_s are the total number of pixels in I_r and I_s , respectively. Then calculate the cumulative distribution function (CDF) according to Equation (27).

$$\begin{aligned}CDF_r(i) &= \sum_{j=0}^i \tilde{h}_r(j) \\ CDF_s(i) &= \sum_{j=0}^i \tilde{h}_s(j)\end{aligned}\quad (27)$$

For each pixel value i in the aerial image, find the smallest j such that $CDF_s(i) \leq CDF_r(j)$. This mapping function $T(i) = j$ can be implemented using a lookup table (LUT). Finally, applying the LUT to I_s (using OpenCV's `cv::LUT` function) yields the histogram-matched image I_m , effectively adjusting the gray-level distribution of the aerial image to be similar to that of the offline map, thereby improving feature matching.

d. Feature detection (KAZE algorithm) and feature matching

The KAZE algorithm is a multiscale feature detection method based on nonlinear diffusion filtering. This study uses the KAZE feature detector to detect features in I_s and I_m , retaining the 2000 strongest feature points and constructing a descriptor. Then, a brute-force matcher (BFMatcher) is used for feature matching based on L_2 distance (Figure 8).

After matching, extract the corresponding feature point coordinates from the matching results, denoted as $\{p_i\}$ (feature points of the offline map) and $\{q_i\}$ (feature points of the aerial photo). Use the RANSAC algorithm to select the model with the most inliers according to Equation (28), where H is a 3×3 affine transformation matrix, and D is a pre-set Euclidean distance tolerance. Estimate the geometric transformation between the two images using Equation (29).

$$\|q_i - (Hp_i + t)\| \leq D \quad (28)$$

$$q_i = Hp_i + t \quad (29)$$

the matching result can be calculated by RANSAC algorithm (Figure 9).

e. Image fusion

After feature matching between the offline map I_r and the aerial photo I_m , the author uses the obtained transformation matrix H to map the aerial photo to the coordinate system of the offline map, and obtains the transformed aerial photo I_{warped} . According to Formula (30), the offline map I_r and I_{warped} are combined to obtain $I_{stitched}$ (Figure 10), where α is the transparency.

$$I_{stitched} = \alpha I_r + (1 - \alpha) I_{warped} \quad (30)$$

Feature Matching Result

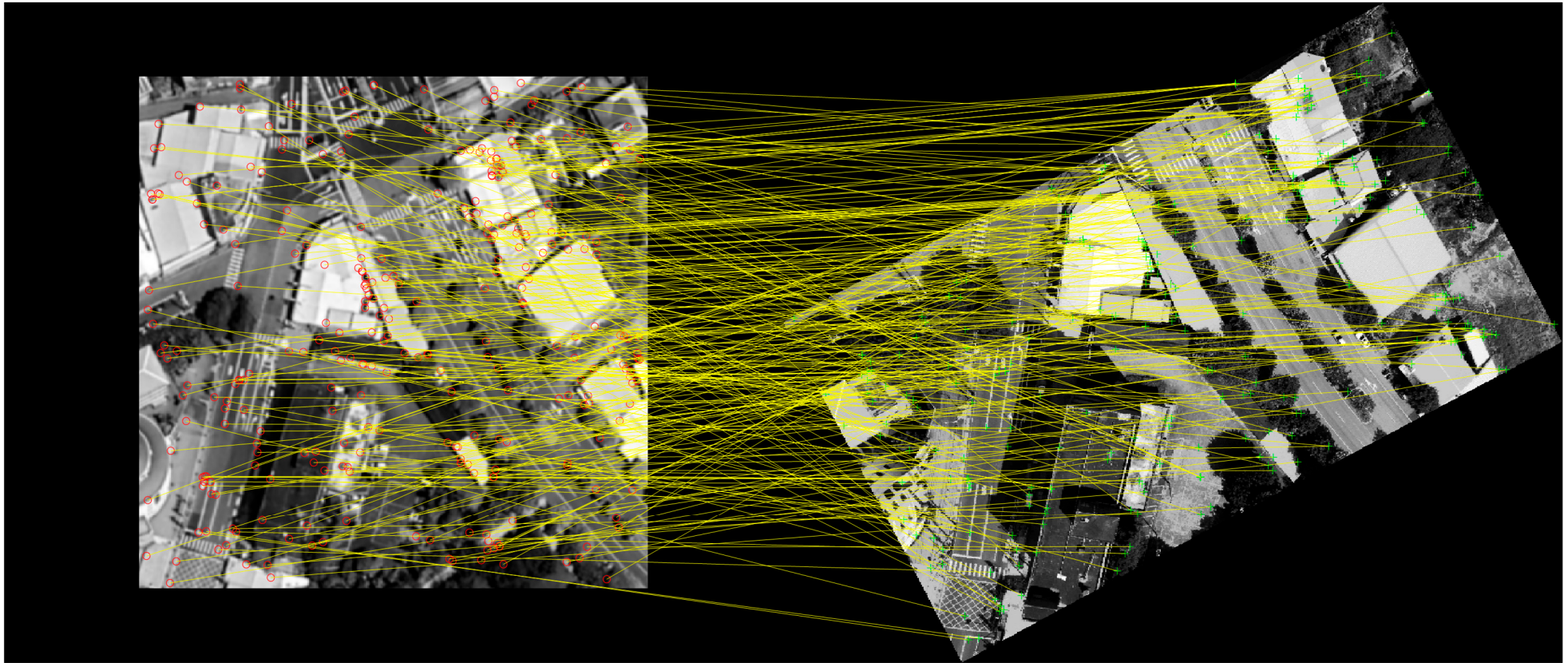


Figure 8. Feature matching results.

Feature Matching Results after Applying RANSAC

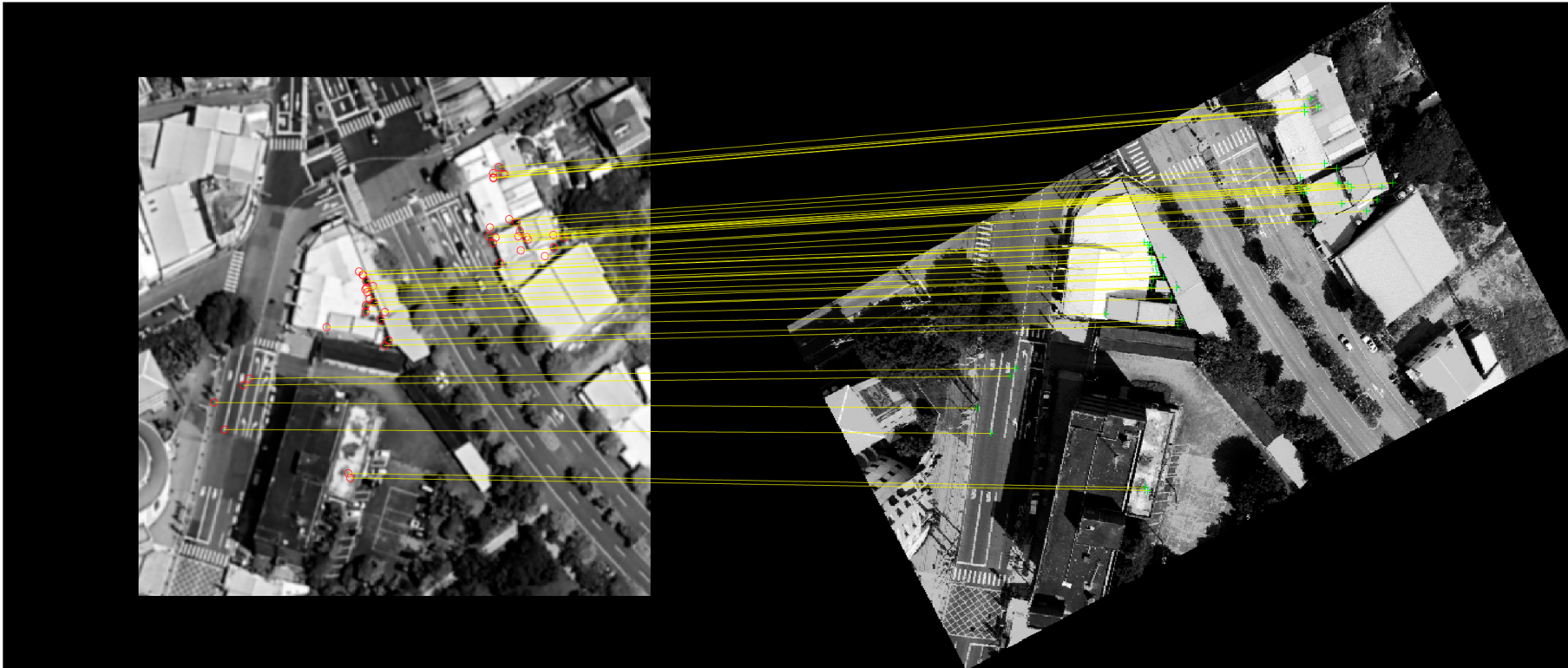


Figure 9. Feature matching using the RANSAC algorithm.

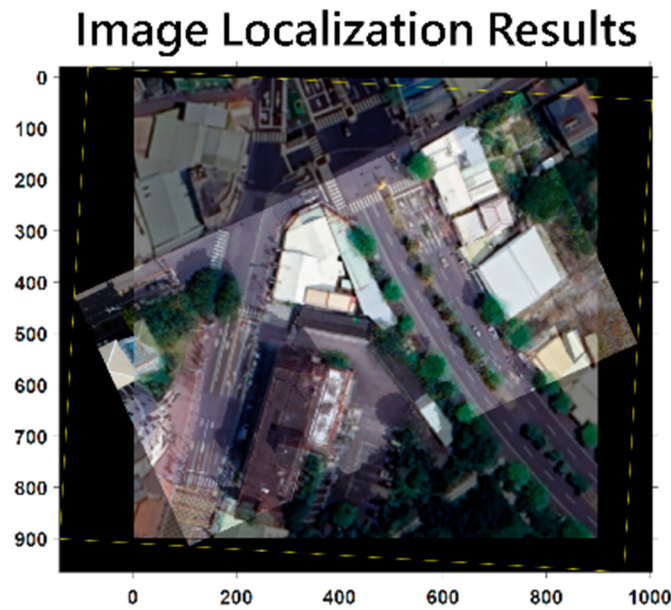


Figure 10. The result of imagine localization.

Calculate the image center (X_{center}, Y_{center}) in pixels using the coordinates of the four warped corners as per Equation (31).

$$\begin{aligned} X_{center} &= \frac{1}{4} \sum_{i=1}^4 k_x(i) \\ Y_{center} &= \frac{1}{4} \sum_{i=1}^4 k_y(i) \end{aligned} \quad (31)$$

where $(k_x(i), k_y(i)), i = 1, 2, 3, 4$ are the corner coordinates of I_{warped} in pixels.

Finally, according to Equation (32), the latitude and longitude coordinates of the tile center can be obtained.

$$\begin{aligned} Lat_{est} &= Lat_T + (k_y(1) + Y_{center}) \times \Delta Lat \\ Lon_{est} &= Lon_T + (k_x(1) + X_{center}) \times \Delta Lon \end{aligned} \quad (32)$$

Convert estimated latitude and longitude to pixel coordinates for visual-inertial odometry error correction.

2.2.4. Dynamic Weighted Fusion Strategy

In Section 2, an adaptive EKF is designed based on the angular velocity and inlier ratio. In Section III, global feature matching correction is used to determine the drone's actual position when the IMU information and optical flow accumulation position are unreliable. The author dynamically estimates the final fusion weight based on the number of matches, inlier ratio, reprojection error, and transformation matrix confidence, as shown in Equation (33), to fully utilize their respective advantages and reduce cumulative errors.

$$X_{fused} = \omega_{EKF} X_{EKF} + \omega_{FM} X_{FM} \quad (33)$$

where X_{fused} is the final calculated position. ω_{FM} and ω_{EKF} are the weights of global feature matching correction position and EKF estimated position, calculated as Equation (34). X_{EKF}, X_{FM} are the positions calculated by EKF and feature matching, respectively.

$$\begin{aligned}\omega_{FM} &= \frac{c_{FM}}{c_{EKF} + c_{FM}} \\ \omega_{EKF} &= \frac{c_{EKF}}{c_{EKF} + c_{FM}}\end{aligned}\quad (34)$$

The confidence level of feature matching correction, c_{FM} , comprehensively considers the number of matched points (N_{match}), the ratio of inner points ($\frac{N_{inlier}}{N_{match}}$), the reasonableness of the transformation matrix (T_f), and the re-projection error e_{avg} as confidence indicators.

$$s = \left(\frac{1}{1 + \left| 0.5 \left(\left\| \begin{bmatrix} T_f(1,1) \\ T_f(2,1) \end{bmatrix} + \begin{bmatrix} T_f(1,2) \\ T_f(2,2) \end{bmatrix} \right\| - 1 \right) \right| \right)} \cdot \left(\frac{1}{1 + \left(\frac{\text{atan}\left(\frac{T_f(1,1)}{T_f(2,1)}\right)}{45^\circ} \right)} \right) \quad (35)$$

$$e_{avg} = \frac{1}{N_{inlier}} \sum_{i=1}^{N_{inlier}} \|T(p_i^{map}) - (p_i^{photo})\| \quad (36)$$

Deriving the confidence level c_{FM} of global feature matching correction as in Equation (37).

$$C_{FM} = \frac{N_{inlier}}{N_{match}} \cdot \left(\frac{1}{1 + e_{avg}} \right) \cdot s \quad (37)$$

The c_{EKF} then extracts the EKF prediction uncertainty σ_{EKF}^2 from the covariance matrix $P \in \mathbb{R}^{9 \times 9}$ as in Equation (38).

$$\sigma_{EKF}^2 = P_{1,1} + P_{2,2} \quad (38)$$

The higher its value, the less reliable EKF is, so c_{EKF} is set as Equation (39).

$$C_{EKF} = \frac{1}{1 + 0.1\sigma_{EKF}} \quad (39)$$

2.2.5. Summary of the Adaptive VIO Procedure

The proposed adaptive VIO procedure integrates these components into a robust, hierarchical localization framework. The system's core is an Adaptive EKF (Section 2.2.2), which continuously predicts the state using high-frequency IMU data and corrects it using optical flow (Section 2.2.1) as the measurement update. This EKF adaptively tunes its noise matrices (Q and R) based on flight dynamics and measurement quality to optimize the fusion. To mitigate long-term drift, a feature matching calibration module (Section 2.2.3) is triggered when VIO quality degrades. This module performs global image matching using KAZE features against an offline map to compute an absolute position correction X_{FM} . Finally, a Dynamic Weighted Fusion Strategy (Section 2.2.4) intelligently arbitrates between the continuous VIO estimate (X_{EKF}) and the periodic global correction (X_{FM}). This final fusion X_{fused} (Equation (33)) is weighted by the real-time confidence of both the EKF's covariance uncertainty and the feature matching quality, ensuring that the system leverages the most reliable information source at any given time.

3. Localization Algorithm Simulation Results

This study simulates actual drone flight paths and aerial photography processes using real-world collected drone aerial images and flight log files. It verifies the actual positioning effect of the fusion algorithm using only image and IMU information.

Google Maps' local satellite images (Figure 3) are used as a reference, and the simulation area is limited to this region. The algorithm's positioning performance is validated through multiple segments of aerial images with different scenarios. The comparison ob-

jects include optical flow, standard EKF-VIO, adaptive EKF-VIO without feature matching correction, and the proposed adaptive EKF-VIO.

The simulation assumes a drone flying at a fixed altitude, with a camera mounted underneath, viewing the ground and changing with the drone's attitude. If GPS is interfered with or fails, the last available GPS location is used as the starting point, and the algorithm is initialized based on this coordinate to perform non-GPS positioning. The simulation results are compared in terms of Root Mean Square Error (RMSE), Maximum Error, and Standard Deviation (STD). The hardware specification is listed in Appendix A.

3.1. Validation Scenario 1

The first flight path is shown in Figure 11, with the drone flying at an altitude of approximately 120 m. The green line represents the actual GPS flight path extracted from the flight log file and projected onto the image coordinates as the Ground Truth. The simulation results for this flight path are shown in Table 1. The execution process of the MATLAB-based (version 1.2.2) AEKFVIO algorithm is demonstrated in a Supplementary Video available online (Video S1, available at <https://youtu.be/IQE1u4FGiBM> (accessed on 20 November 2025)).

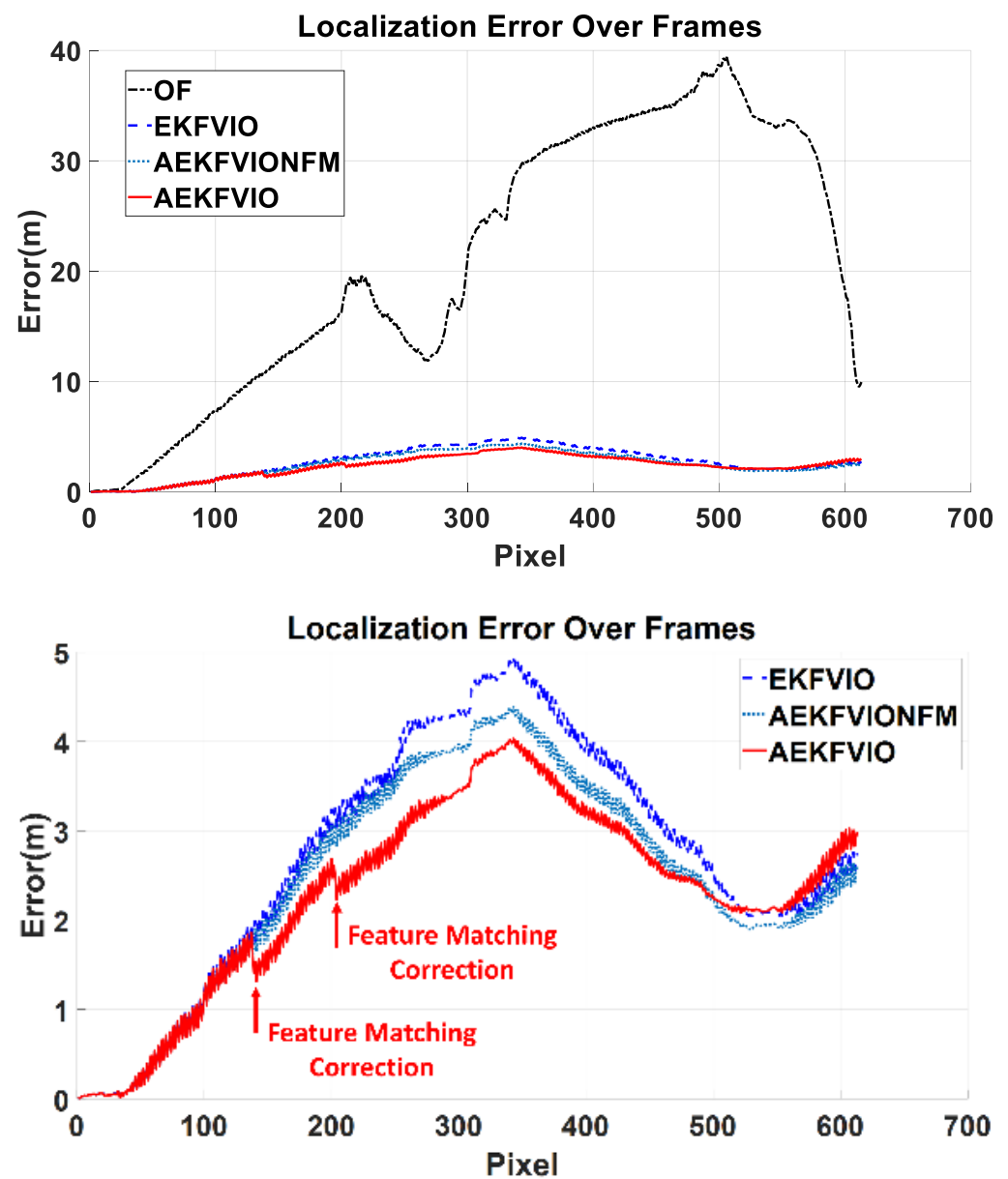


Figure 11. Localization results of each method under Validation Scenario 1.

Table 1. Simulation result comparisons under Validation Scenario 1 (unit: meter).

	RMSE	Max E	STD
OF	20.9308	39.4078	12.0733
EKFVIO	2.7105	4.9218	1.3831
AEKFVIONFM	2.4778	4.3891	1.2298
AEKFVIO	2.3042	4.0299	1.0787

The results of this flight route show that AEKF-VIO has improved in RMSE, Max Error, and STD location judgment indicators. Figure 12 shows the effect of feature matching correction in this algorithm. Compared to methods without global feature matching correction, the location error is reduced by about 7%. The jitter in the graph is due to the drift of the Ground Truth data itself.

**Figure 12.** Variation in errors regarding each method under Validation Scenario 1.

3.2. Validation Scenario 2

The second flight route is shown in Figure 13, with a drone flight height of approximately 120 m. The green line represents the actual GPS flight route extracted from the flight log file, which is converted to image coordinates through coordinate transformation. The simulation results are shown in Table 2, indicating that the AEKFVIO positioning indicators have decreased compared to other methods. The corresponding execution process is shown in Video S2 (<https://youtu.be/QPDR2GHehOU> (accessed on 20 November 2025)). According to Figure 14, the error variation over frames also shows that the algorithm with feature matching correction has a lower RMSE, decreasing by approximately 19.22% compared to the algorithm without correction.

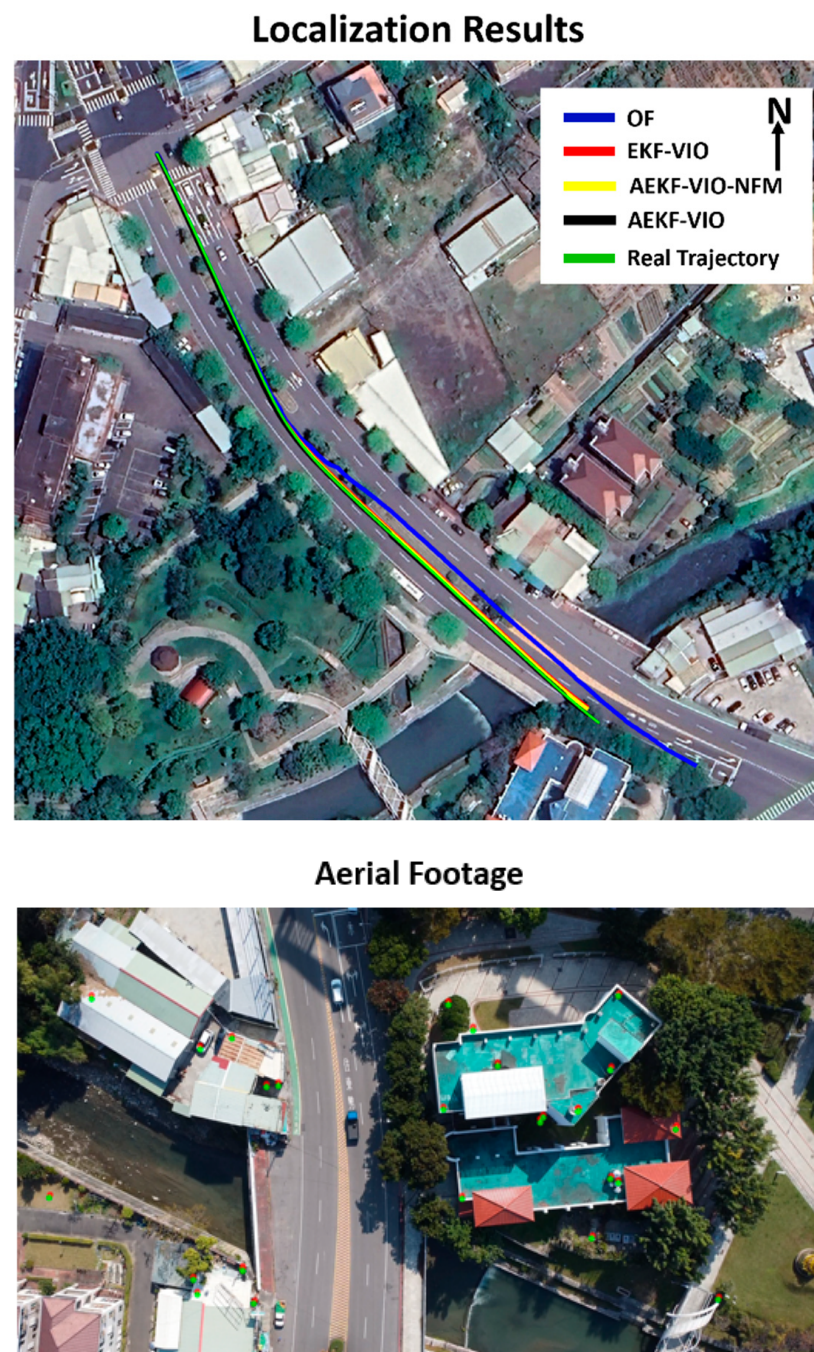
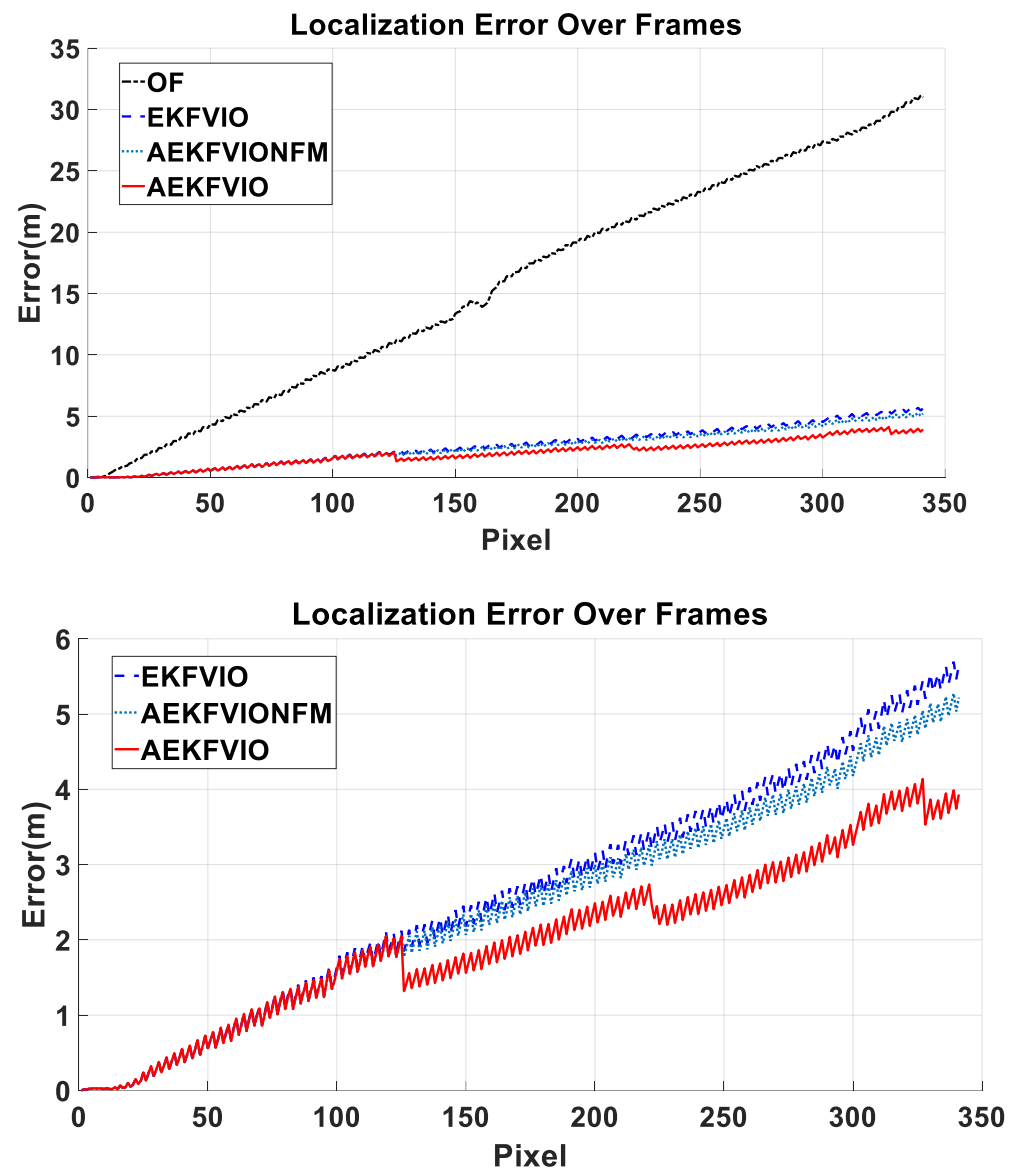


Figure 13. Localization results of each method under Validation Scenario 2.

Table 2. Simulation result comparisons under Validation Scenario 2 (unit: meter).

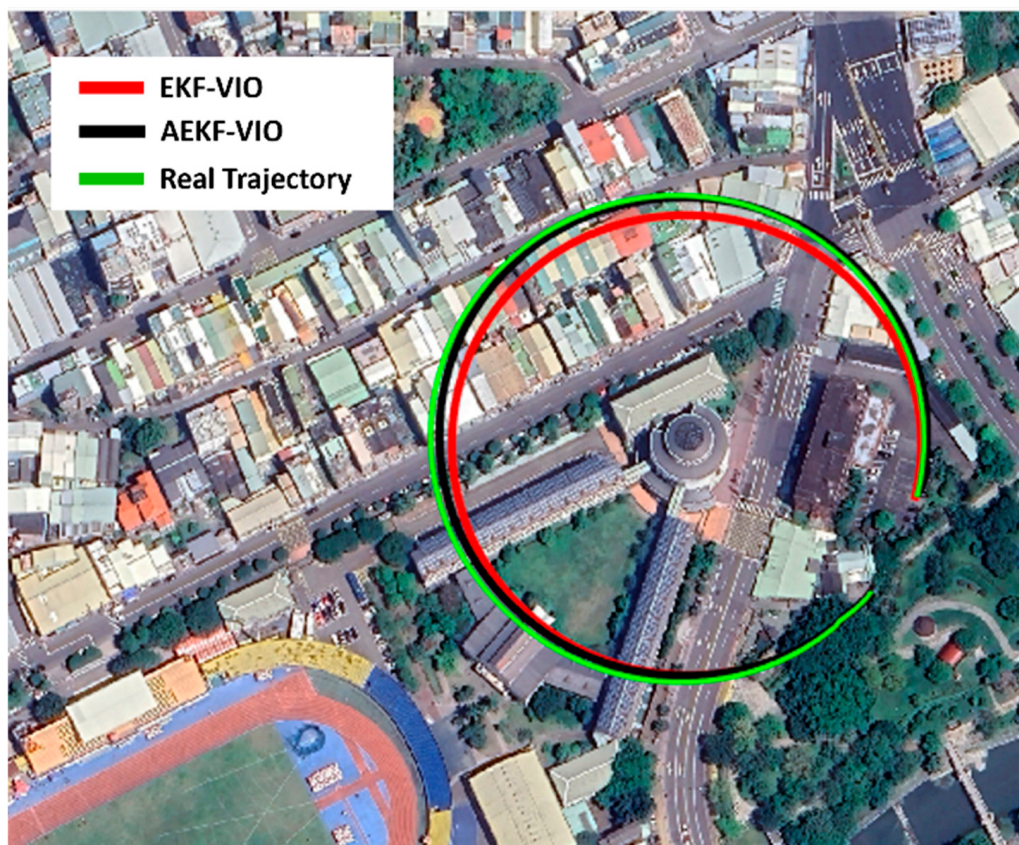
	RMSE	Max E	STD
OF	15.6346	31.1243	9.2162
EKFVIO	2.6154	5.6944	1.5932
AEKFVIONFM	2.4597	5.2740	1.4668
AEKFVIO	1.9871	4.1365	1.0935

**Figure 14.** Variation in errors regarding each method under Validation Scenario 2.

3.3. Validation Scenario 3

The flight path in this segment is a circular motion around a building, as shown in Figure 15, with a fixed pitch angle of -30 degrees. Since the feature matching correction method used in this study is based on the assumptions that the images are taken from a top-down view or with only a small angle offset, this scenario is only used to compare the differences between EKF-VIO and AEKF-VIO-NFM.

Localization Results



Aerial Footage



Figure 15. Localization results of each method under Validation Scenario 3.

As shown in Figure 16, the optical flow method performs poorly in this scenario, mainly because the flight path is rotating around a fixed target, resulting in little to no parallax change in the images, making it difficult to produce effective displacement information.

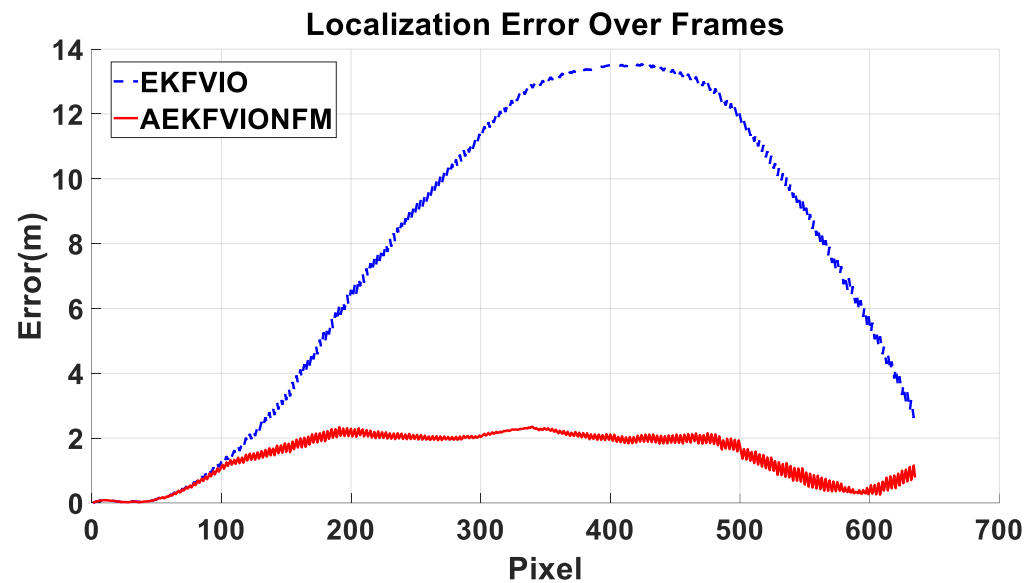


Figure 16. Variation in errors regarding each method under Validation Scenario 3.

The simulation results are shown in Table 3, which indicates that the RMSE of the EKF method with adaptive adjustment is reduced by 81%, demonstrating the importance of adaptability for precise localization. The process is also illustrated in Video S3 (<https://youtu.be/spBDzO4I3CM> (accessed on 20 November 2025)).

Table 3. Simulation result comparisons under Validation Scenario 3 (unit: meter).

	RMSE	Max E	STD
EKFVIO	7.7285	13.5460	4.7567
AEKFVIONFM	1.4541	2.3537	0.7523

4. Conclusions

This study proposes a method that combines visual-inertial odometry (VIO) with global feature matching for image-based localization of drones in GPS-denied environments. The system uses an extended Kalman filter (EKF) as its core, integrating IMU, optical flow, and KAZE feature matching algorithms. It adjusts weights dynamically to suppress cumulative errors and enhance overall robustness. The VIO part uses IMU-provided pose and acceleration information as prediction bases, while optical flow estimates relative displacement. The EKF fuses these two, adjusting state and observation noise covariance matrices based on angular velocity and feature tracking quality. To address cumulative error issues, KAZE feature matching is performed at keyframes. A confidence evaluation mechanism is designed to integrate VIO and feature matching advantages, with separate confidence indicators for EKF estimation and feature matching correction. The EKF's adaptive part reflects prediction uncertainty using covariance diagonal elements, adjusting IMU and optical flow credibility dynamically. The feature matching part considers factors like match point quantity, inlier ratio, average reprojection error, and geometric transformation scale and rotation stability, defining a comprehensive confidence function. The confidence levels determine the dynamic fusion weights for the final localization result, achieving adaptive switching and fusion of information sources. The main contribution of this article is a less complex and low-cost algorithm that can be easily applied to UAVs while maintaining a satisfactory level of accuracy.

Simulation results show that the proposed method AEKFVIO can effectively handle various flight scenarios without GPS, including straight lines, turns, and circling targets,

maintaining stable and high-precision positioning capabilities with lower root mean square error and stronger drift suppression than traditional VIO algorithms.

This system is particularly suitable for urban jungles, forests, and battlefields where GPS is blocked or spoofed, providing a practical and extensible solution for autonomous drone positioning in GPS-denied environments.

5. Future Work

This study has successfully implemented a non-GPS positioning framework that integrates visual-inertial information and feature matching, and improved positioning accuracy and system robustness through a dynamic weighting mechanism. However, the current weight and dynamic adjustment design still rely on empirical rules, relying on manually designed confidence indicators (such as re-projection error, matched point number, rotation and scale deviation, etc.), and cannot fully reflect the complex decision-making logic under different environments and scene changes.

In the future, for those situations where periodic map matching and correction cannot be performed, investigating the extent to which positioning errors accumulate over time will be a topic for future research.

Currently, the map set is updated manually in batches. In the future, functions such as integrating the latest map data sources or designing a real-time map updating mechanism can be added.

In addition, it can be combined with online learning mechanisms, and the system can adjust and reinforce the weight model in real-time according to different task requirements or environmental changes, developing towards a truly intelligent drone positioning module with decision interpretation, self-evaluation, and perception selection capabilities.

Supplementary Materials: The following supporting information can be downloaded at: <https://youtu.be/IQE1u4FGiBM>, <https://youtu.be/QPDR2GHehOU> and <https://youtu.be/spBDzO4I3CM> (accessed on 20 November 2025). Video S1. Localization Algorithm Simulation: Scenario 1. Video S2. Localization Algorithm Simulation: Scenario 2. Video S3. Localization Algorithm Simulation: Scenario 3.

Author Contributions: Conceptualization, Y.-S.W.; methodology, Y.-S.W.; software, C.-H.C.; validation, C.-H.C.; writing—original draft preparation, Y.-S.W. and C.-H.C.; writing—review and editing, C.-H.C.; All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by [National Chung-Shan Institute of Science and Technology].

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to express their sincere gratitude to Chen-Yu Yu and Pei-Yin Lin, from the Avionics and Electrical Engineering Section, Aeronautical Systems Research Division, National Chung-Shan Institute of Science and Technology, for their valuable assistance during the review process.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Hardware configuration.

Hardware Component	Specification
CPU	4th Intel Core i3-4010U 1.7 GHz Dual Core
RAM	4G DDR3 1600 MHz Memory
Storage	500 GB

References

- Ala'Darabseh, M.; Bitsikas, E.; Tedongmo, B. Detecting GPS Jamming Incidents in OpenSky Data. In Proceedings of the 7th OpenSky Workshop, Zurich, Switzerland, 21–22 November 2019; Volume 67, pp. 1–6. [\[CrossRef\]](#)
- Ni, S.; Xu, J.; Zhou, M.; Wang, Y.; Zhang, T. Detection and Elimination Method for Deception Jamming Based on an Antenna Array. *Int. J. Distrib. Sens. Netw.* **2018**, *14*, 1550147718774466. [\[CrossRef\]](#)
- Wang, H.; Chang, Q.; Xu, Y. Deception Jamming Detection Based on Beam Scanning for Satellite Navigation Systems. *IEEE Commun. Lett.* **2021**, *25*, 2703–2707. [\[CrossRef\]](#)
- Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
- Bay, H.; Tuytelaars, T.; Van Gool, L. SURF: Speeded Up Robust Features. In *Computer Vision—ECCV 2006, Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006*; Proceedings, Part I; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3951, pp. 404–417. [\[CrossRef\]](#)
- Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An Efficient Alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; IEEE: Barcelona, Spain, 2011; pp. 2564–2571. [\[CrossRef\]](#)
- Alcantarilla, P.F.; Bartoli, A.; Davison, A.J. KAZE Features. In *Computer Vision—ECCV 2012, Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*; Proceedings, Part VI; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7577, pp. 214–227. [\[CrossRef\]](#)
- Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. *IEEE Int. Conf. Robot. Autom.* **2014**, *15*, 15–22. [\[CrossRef\]](#)
- Qin, T.; Li, P.; Shen, S. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* **2018**, *34*, 1004–1020. [\[CrossRef\]](#)
- Weiss, S.; Scaramuzza, D.; Siegwart, R. Monocular vision for autonomous micro helicopters. *J. Field Robot.* **2012**, *28*, 854–874. [\[CrossRef\]](#)
- Xiao, L.; Wang, J.; Qiu, X.; Rong, Z.; Zou, X. Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment. *Robot. Auton. Syst.* **2019**, *117*, 1–16. [\[CrossRef\]](#)
- Li, M.; Mourikis, A.I. High-Precision, Consistent EKF-Based Visual-Inertial Odometry. *Int. J. Robot. Res.* **2013**, *32*, 690–711. [\[CrossRef\]](#)
- Geneva, P.; Ekenhoff, K.; Lee, W.; Yang, Y.; Huang, G. OpenVINS: A Research Platform for Visual-Inertial Estimation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 11–17. [\[CrossRef\]](#)
- Qin, T.; Cao, S.; Pan, J.; Shen, S. A General Optimization-Based Framework for Global Pose Estimation with Multiple Sensors. *arXiv* **2019**, arXiv:1901.03642. [\[CrossRef\]](#)
- Mur-Artal, R.; Tardós, J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [\[CrossRef\]](#)
- Wang, Z.; Wang, T.; Shen, S. Stereo DSO: Large-Scale Direct Sparse Visual Odometry with Stereo Cameras. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 3923–3931. [\[CrossRef\]](#)
- Wang, J.; Xu, B.; Cheng, X. UAV navigation in large-scale GPS-denied bridge environments using fiducial marker-corrected stereo visual-inertial localisation. *Autom. Constr.* **2023**, *147*, 104711. [\[CrossRef\]](#)
- Xue, B.; Yang, Z.; Liao, L.; Zhang, C.; Xu, H.; Zhang, Q. High precision visual localization method of UAV based on feature matching. *Front. Comput. Neurosci.* **2022**, *16*, 1037623. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, T.; Cai, Z.; Wang, Y. Integrated vision/inertial navigation method of UAVs in indoor environment. *J. Beijing Univ. Aeronaut. Astronaut.* **2018**, *44*, 176–186.
- Xu, G.; Zeng, J.; Liu, X. Visual odometry based on the fusion of optical flow method and feature matching. *Laser Optoelectron. Prog.* **2020**, *57*, 201501. [\[CrossRef\]](#)
- Cheng, C.; Zhang, Y.; Li, J.; Wang, Y. Monocular visual odometry based on optical flow and feature matching. In Proceedings of the 29th Chinese Control and Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 5487–5492. [\[CrossRef\]](#)
- Hsieh, C.-Y.; Wang, Y.-S. A GPS-free air vehicle on-board positioning mechanism applying feature matching technique. In Proceedings of the 2024 AIAA/IEEE 43rd Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 29 September–3 October 2024; pp. 1–7. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.