GNSS-denied geolocalization of UAVs using terrain-weighted constraint optimization

Fushan Yao [*], Chaozhen Lan, Longhao Wang, Hongfa Wan, Tian Gao, Zijun Wei

PLA Strategic Support Force Information Engineering University, Zhengzhou, 450001, China

ARTICLE INFO

ABSTRACT

Accurate geolocation using Global Navigation Satellite Systems (GNSS) is essential for safe and long-range unmanned aerial vehicles (UAVs) flights. However, GNSS systems are susceptible to blockages, jamming, and spoofing attacks. Localization using onboard cameras and satellite images provides a promising solution for UAVs operating in GNSS-denied environments. In this paper, we developed a novel UAV visual localization system for GNSS-denied situations, both day and night, that integrates image matching, visual odometry (VO), and terrain-weighted constraint optimization. First, an effective map management strategy is designed for satellite image chunking, real-time scheduling, and merging. Then, a 2D–3D geo-registration method, combining Bidirectional Homologous Points Search, is introduced to obtain accurate 3D virtual control points for UAV absolute localization. Lastly, a position estimation and optimization method, integrating the sliding window with terrain weighting constraints, is proposed to control position error accumulation and reduce position drift. Twenty experiments were conducted in typical and complex scenarios to validate our system's resilience to altitude changes, trajectory variations, and rolling terrain. Our system demonstrated drift-free and viewpoint-robust, maintaining stability even in feature-poor environments and seasonal variations. It does not require loop closure, allowing for re-localization after positioning failures. Additionally, we utilized thermal infrared images to demonstrate the system's performance in night-time conditions. With a Mean Absolute Error of less than 7 m, it can be a powerful complement to GNSS in the event of GNSS-Denied environments. All demonstration videos of our system can be found at https://github.com/YFS90/GNSS-Denied-UAV-Geolocalization.

1. Introduction

Unmanned aerial vehicles (UAVs) have become increasingly indispensable in various applications, from agriculture to military surveillance. Most UAVs depend on integrating Global Navigation Satellite Systems (GNSS) and Inertial Navigation Systems (INS) to ensure precise positioning and navigation, which is essential for executing complex tasks. Although GNSS is beneficial, it is vulnerable to challenges such as signal blockages, intentional jamming, and spoofing attacks, which can severely reduce navigation accuracy or cause mission failure. Additionally, the inertial measurement units (IMUs) within INS tend to accumulate errors over time due to sensor drift, rendering INS increasingly unreliable for prolonged navigation without correction from external references. Therefore, the ability of UAVs to determine their position in GNSS-denied situations is crucial for ensuring safe flight and mission execution.

Vision-based localization methods can be an alternative in GNSS-denied environments because cameras act as passive sensors, avoid blockage or interference, and offer low SWaP-C (Size, Weight, Power, and Cost) (Couturier and Akhloufi, 2021). Recent advancements in visual-inertial odometry (VIO) and Simultaneous Localization and Mapping (SLAM) (Luo et al., 2023) demonstrate the potential for autonomous robots to achieve accurate localization in GNSS-denied environments. However, SLAM suffers from cumulative drift, particularly over long distances, which diminishes its effectiveness in long-range UAV operations. Although techniques such as loop closure can mitigate drift, they can only partially reduce its effects and cannot eliminate positional errors entirely. Furthermore, SLAM systems are inherently limited, and without external georeferencing, they cannot generate earth-fixed coordinates, restricting their use in applications requiring absolute positional accuracy (He et al., 2023). Another approach for localization in GNSS-denied environments involves matching UAV images with satellite images (Kinnari et al., 2021, 2022) or retrieving a satellite patch corresponding to a query UAV frame (Hu et al., 2024). The UAV–satellite image matching method outputs geographical

coordinates and offers the advantage of drift-free position estimation. However, most existing work focuses on top-down view image matching or orthorectifying UAV images with an IMU (Kinnari et al., 2021; Bianchi and Barfoot, 2021; Mei et al., 2023). Due to the influence of airflow, it is challenging for the UAV to maintain a downward-facing during flight. Moreover, the positioning accuracy of this method is limited by the quality of the matching process. Factors such as image rotation, viewing angle discrepancies, and seasonal changes must be addressed. Additionally, it is challenging to perform re-localization after positioning failure in matching-based localization methods (Qiu et al., 2024). The retrieval-based method retrieves image patches similar to the query with the highest similarity from large-scale satellite images and obtains the corresponding geographical information. However, the problem is more than a one-to-all retrieval task in real-world scenarios. The performance of this approach is affected by the density and distribution of satellite images in the database, potentially resulting in significant errors (Hu et al., 2024). Moreover, the challenge of GNSS-denied positioning at night is an urgent problem that needs to be addressed and has seldom been considered in previous work.

In this paper, we investigate and develop a UAV visual localization system that integrates image matching, visual odometry, and terrain-constrained optimization for GNSS-denied environments, both day and night. The system matches UAV video frames or images with satellite images, obtains precise virtual 3D control points through feature association for rough positioning, estimates the UAV pose using visual odometry and improves localization accuracy through terrain constraint optimization. The main contributions of this paper are as follows:

(1) We have designed and implemented a novel visual localization system that utilizes geo-referenced data, including publicly accessible satellite images and elevation data, for precise 3D positioning of UAVs in GNSS-denied environments. The system operates without the need for the UAV's camera to capture downward-facing photos, nor does it require an IMU or altimeter. It only requires the initial position of the UAV to achieve precise visual positioning in high-altitude scenarios ranging from 150 to 1500 m, over distances of up to 20 km, and across coverage areas of approximately 10 km^2.

(2) We introduce a visual odometry positioning method that combines image matching and terrain-weighted constraint optimization. This method demonstrates resilience to altitude changes, trajectory shapes, and rolling terrain. It is drift-free and viewpoint-robust, maintaining stability even in feature-poor environments and seasonal variations. Additionally, it does not need loop closure, allowing for relocation after positioning failure. Furthermore, we validated the system's nighttime localization capability.

(3) Twenty sets of real-world data covering plains, hilly terrain, and urban and rural areas are employed to comprehensively validate the system's robustness and accuracy under complex scenarios.

The manuscript is organized as follows: Section 2 gives the related works. Section 3 introduces the proposed algorithm in detail. Section 4 presents and discusses the experimental results. Section 5 drew the conclusions.

2. Related work

This section reviews recent research on UAV localization in GNSS-denied environments, focusing on UAV–satellite image matching, SLAM-based methods, and retrieval methods.

2.1. UAV-satellite image matching-based visual localization

The fundamental concept of localization based on image matching is to convert the task of localization into matching UAV images with satellite maps. Handcrafted detectors, like Histogram of Oriented Gradient (HOG) (Shan et al., 2015), SIFT (Hamidi and Samadzadegan, 2015), and ORB (Chan and Yakimenko, 2022), are employed for GNSS-denied localization. However, since UAV and satellite images are captured using different sensors, they can vary in season, lighting, and viewpoints. As a result, handcrafted features are prone to failure during matching localization tasks. Liu et al. (2023) matches points of interest (POI) and store signage text (LPS) in UAV images for autonomous localization. Qiu et al. (2024), Luo et al. (2024) and Ye et al. (2024) demonstrate that deep learning-based feature extraction methods, such as Superpoint (DeTone et al., 2018), Superglue (Sarlin et al., 2020), and LightGlue (Lindenberger et al., 2023), can obtain more accurate matches from images with changes in appearance and different viewpoints. Kinnari et al. (2022) proposes a localization solution that matches UAV images to orthophotos using a trained convolutional neural network (CNN) model. Chen and Jiang (2023) applies the SE(2)-steerable network and adaptive high-quality match selection to localize UAV oblique imagery in GNSS-denied environments. Li et al. (2023a) and Xiao et al. (2024) employ multimodal image registration and a coarse-to-fine depth homography estimation method to localize UAV near-infrared and thermal images. Although learning-based matching methods mitigate viewpoint and appearance differences between images, they disregard the three-dimensional constraints between UAV images, leading to non-smooth trajectories and difficulties in relocation after matching failures.

2.2. SLAM-based UAV visual localization

Recently, numerous studies have proposed solutions to address SLAM limitations, which provide only relative positioning in the absence of geo-referenced data. Jun et al. (2022) and Kinnari et al. (2023) match UAV imagery with satellite images for UAV's absolute 2-degree-of-freedom (2-DoF) pose, then use VIO to calculate the UAV's relative 6-degree-of-freedom (6-DoF) pose, integrating both through bundle adjustment (BA) for global positioning. Hou et al. (2020) uses terrain matching with the digital elevation model (DEM) to derive 3D terrain points for the UAV's absolute positioning. He et al. (2023) combines VIO, Visual Place Recognition (VPR), and map alignment for global pose estimation. Additionally, Jurevičius et al. (2019) proposed a method for UAV positioning in GNSS-denied environments by combining particle filtering and VO. Although the SLAM method with geo-referenced data provides absolute positioning, it requires an IMU or barometer and often demands vertical camera images or low altitudes, simplifying the complexities of UAV positioning in GNSS-denied environments.

2.3. Retrieval-based UAV visual localization

Previous studies Ali-Bey et al. (2023) and Dai et al. (2023) addressed UAV positioning as scene recognition by retrieving the most similar satellite images from a database and using their locations as the UAV's query frame location. As a result, retrieval-based methods can only provide relatively coarse 2D positioning for the UAV. Wang et al. (2024), Ye et al. (2024) and Li et al. (2023b) proposed a two-stage method combining retrieval and fine-grained matching for precise UAV positioning, achieving an accuracy of 10 m. Notably, Li et al. (2023b) reported a localization error of only 2.39 m in 0.59 s in a sparsely textured village scene. FoundLoc (He et al., 2023) developed a GNSS-denied localization approach using VIO and VPR with a foundation model. Hu et al. (2024) enhanced NetVLAD for image retrieval and used Swin-Descriptors with Perspective-n-Point (PnP) for localizing 16
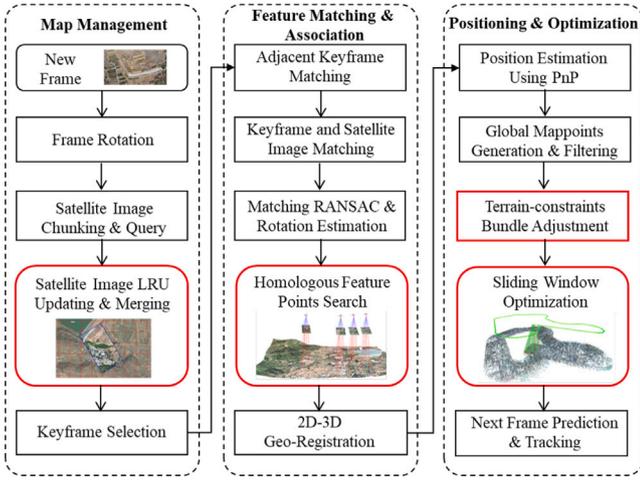
Fig. 1. The proposed system framework for the location of UAVs in GNSS-Denied environments.

Fig. 2. Satellite image chunking. In the figure, red dashed lines indicate the boundaries of the satellite image segments, the blue solid line area R represents the coordinate range of the rotated UAV image, and the yellow solid line area S denotes the corresponding satellite image range for the UAV image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

UAV datasets. However, this method does not consider geometric constraints between UAV images and involves a large search space due to one-by-one comparisons. Additionally, performance can be affected by satellite image density and distribution, leading to potential positioning errors.

3. Methodology

As depicted in Fig. 1, our UAV visual geolocation system comprises three main modules: image management module, image feature matching and association module, and pose estimation and optimization module. In the first module, we developed an efficient map management strategy for partitioning, scheduling, and merging satellite images. The second module handles image matching and feature point association, identifying virtual 3D control points via Bidirectional Homologous Points Search (BHPS) and 2D–3D geo-registration for UAV absolute positioning. The third module introduces a terrain-weighted constraint method combined with a sliding window strategy for UAV pose estimation and optimization.

3.1. Map manager and keyframe selection

For system localization, UAV images need to be matched with satellite images, which often cover large areas. Thus, chunking the satellite map is necessary to determine the corresponding range of the UAV image, enhancing matching accuracy and efficiency. As illustrated in Fig. 2, to manage the computational load of image matching, the satellite image is divided into m rows and n columns, each 128×128 pixels, for storage. Then, the four corner coordinates of the UAV image are calculated to determine the satellite image range S corresponding to the UAV image. This study assumes the UAV's approximate takeoff point is known. Based on our experience, for consumer-grade UAVs operating at an altitude of 500 m, the position error of the take-off point should be within 300 m. Satellite image blocks near this point are loaded for system initialization during the first UAV image processing. After chunking the satellite images and considering deployment on edge devices and computational efficiency, we calculate the number L of image blocks to be loaded based on the available memory and dynamically update them in memory using the least recently used (LRU) method.

$$L = \begin{cases} m \times n, & \text{if } C_{ram} \times 0.5 \geq S_{size} \\ m \times n \times k, & \text{otherwise.} \end{cases} \quad (1)$$
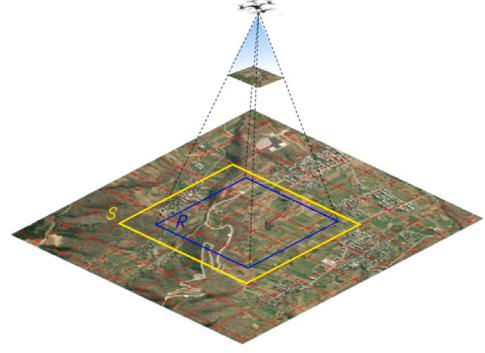
where C_{ram} represents the device's available memory, S_{size} denotes the satellite image size, and k is the empirically determined loading factor based on available memory.

$$k = \frac{C_{ram} \times 0.5}{S_{size}} \times 0.5 \quad (2)$$

This strategy manages dynamic updates of image blocks on edge devices. The satellite image blocks within range S are then merged by coordinates for image matching.

Keyframe selection affects both computational load and accuracy. Extracting keyframes at fixed intervals can lead to issues such as retaining many similar images if the UAV moves slowly or missing important images if the UAV moves quickly. We use the keyframe selection strategy from ORB-SLAM2 (Mur-Artal and Tardós, 2017). Assuming a video frame rate of f frames per second, the system assesses whether a frame should be a keyframe at intervals of $f/2$. A new keyframe is added if frame overlap is under 90%, indicating significant UAV movement, or if $f \times 2$ frames have passed since the last keyframe to preserve information and ensure localization continuity.

3.2. Feature extraction and association

3.2.1. Image matching and rotation estimation

Seasonal, lighting, and viewpoint differences between UAV and satellite images limit the robustness and reliability of traditional matching algorithms. We use Superpoint to extract deep feature points, LightGlue for matching, and Progressive Sampling Consensus (PROSAC) to filter mismatches. The sensitivity of Superpoint feature descriptors to image rotation may lead to matching failures when the rotation angle between UAV and satellite images exceeds 45°. After filtering mismatches, the system calculates the rotation angle between the UAV and satellite images using matching points. The next keyframe is then rotated according to this angle to estimate its satellite map range. Unlike (Bianchi and Barfoot, 2021), which requires a fixed heading angle, and Kinnari et al. (2021) and Mei et al. (2023), which use IMU angles, our approach pre-rotates keyframes using geometric relationships between adjacent frames. This method improves matching and provides rotation invariance, accommodating the UAV's localization across complex flight paths.

3.2.2. Bidirectional Homologous Points Search (BHPS) and 2D-3D geo-registration

In this paper, we calculate the UAV's initial pose using 3D virtual control points and optimize it to minimize localization error and pose drift, enhancing the system's re-localization capability. As shown in
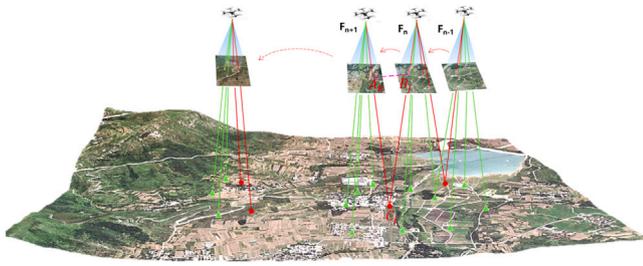
Fig. 3. Obtaining 3D virtual control points. In the figure, red dots indicate 3D virtual control points, and green triangles represent global map points from triangulation. BHPS is employed to determine the 2D control point C_i, and its elevation is interpolated on the DEM using 2D–3D Geo-Registration to acquire reliable 3D virtual control points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 3, UAV images F_n and F_{n+1} extract Superpoint feature points and use LightGlue to obtain the matching point pair set (B, A). F_{n+1} and F_n are also matched with the satellite map to get the matching point pair sets (A, C) and (B, C). If the feature point C_i extracted from the satellite map meets certain conditions, C_i is considered a virtual control point.

$$\begin{cases} B_j \langle - \rangle A_k, \left(B_j \in B, A_k \in A \right) \\ A_k \langle - \rangle C_i, \left(A_k \in A, C_i \in C \right) \\ C_i \langle - \rangle B_j, \left(C_i \in C, B_j \in B \right) \end{cases} \tag{3}$$

where $\langle - \rangle$ denotes image matching. Feature points B_j and A_k on F_n and F_{n+1} are matched, feature points C_i on the satellite image are matched with A_k and B_j on F_{n+1} and F_n, respectively. By employing bidirectional matching associations between keyframes and between keyframes and the satellite image, we validate stable and reliable 2D virtual control points C_i. In contrast, existing matching-based localization methods often rely solely on unidirectional matching between UAV and satellite images. They lack inter-frame matching for bidirectional validation, resulting in unstable points and reduced accuracy in texture-sparse areas.

Utilizing the TIFF World File (TFW) associated with GeoTIFF format satellite imagery, the geographic coordinates of the upper-left corner of the satellite image and the pixel resolutions in the X and Y directions can be obtained, allowing for the calculation of the geographic coordinates (X_i, Y_i) for any pixel C_i in the satellite image. We calculate the elevation Z_i of C_i on the DEM using 2D–3D Geo-Registration to obtain the 3D coordinates (X_i, Y_i, Z_i), which are considered as the 3D virtual control point. We determine the pixel coordinates (Dx_i, Dy_i) of C_i on the DEM and then use spatial interpolation to obtain the elevation Z_i.

$$\begin{cases} Dx_i = \dfrac{X_i - x}{rx} \\ Dy_i = \dfrac{y - Y_i}{ry} \\ Z_i = T \left(Dx_i, Dy_i \right) \end{cases} \tag{4}$$

where x and y denote the geographic coordinates of the DEM's top-left corner, and rx, and ry are the DEM resolutions. Considering the DEM resolution and terrain continuity, we use Cubic Spline Interpolation $T()$ with a 4×4 neighborhood around (Dx_i, Dy_i) to achieve accurate and smooth Z values. This process yields the 3D virtual control point set C_{3d} for UAV pose estimation and optimization. Methods (Shan et al., 2015; Hou et al., 2020) that use iterative solving for elevation determination are computationally demanding, time-consuming, and need accurate UAV pose data.

3.3. Absolute pose estimation and optimization

3.3.1. Rough estimation of image pose using PnP

After the homologous points search and 2D–3D geo-registration, a set of 3D virtual control points C_{3d} is obtained. Using the 3D coordinates of each point and the corresponding pixel coordinates, the PnP problem can be formulated to estimate the UAV's absolute position. However, in areas with sparse texture or significant changing ground features, obtaining enough 3D points for accurate positioning is challenging. Moreover, frame-by-frame PnP calculations ignore geometric constraints of co-visible points between frames, reducing accuracy and causing potential jumps in the trajectory. To address this issue, we improve UAV positioning by integrating co-visible geometric relationships between images into coarse localization. First, co-visible geometric relationships are employed to triangulate (Davison, 2003) and generate global map points. Then, global map points and 3D virtual control points are used for terrain-weighted optimization, enhancing UAV pose accuracy and smoothing the trajectory.

3.3.2. Global map points generation and filtering

After estimating the poses of images F_{n-1} and F_n with EPnP, we apply triangulation to obtain 3D global map points in the world coordinate system, shown as green triangles in Fig. 3. We remove outlier map points using the method from ORB-SLAM2. We also check whether the difference between the global map point's Z value and the DEM elevation exceeds a fixed threshold to filter out outliers.

3.3.3. Terrain-weighted constraint bundle adjustment and sliding-window optimization

In SLAM, Bundle Adjustment (BA) is commonly employed to minimize the objective function e, optimizing the UAV's pose and map points.

$$e = \min_{\xi, P} \sum_{i=1}^{m} \sum_{j=1}^{n} \left\| h \left(\xi_i, P_j \right) - p_{ij} \right\|^2 \tag{5}$$

where $h()$ denotes the reprojection function, i represents the ith image, j indicates the jth global map point on the image, ξ is the initial UAV pose estimated by PnP, P is the generated global map point, and p is the observed position of the feature point corresponding to P. However, the objective function e minimizes only the least squares difference between the projected and observed positions of the global map points, excluding external control information. This can lead to local optima, causing UAV positioning errors and pose drift.

We use DEM elevation data as prior information, project the 3D virtual control points, and compare them with their observed positions, as illustrated in Fig. 4. The figure's green triangular denotes global map points, and the red circular indicates the 3D virtual control points. During optimization, both types of points are projected and compared with their observed positions, serving as visual odometry and terrain constraints for integrated geolocation. The optimized objective function e can thus be expressed as:

$$\begin{aligned} e = \min_{\xi, P} \sum_{i=1}^{m} \sum_{j=1}^{n} \Big\{ & (1 - \beta_i) \times \rho \left(\left\| h \left(\xi_i, P_j \right) - p_{ij} \right\|^2 \right) \\ & + \alpha_j \times \beta_i \times \left\| h \left(\xi_i, C_j \right) - p_{ij} \right\|^2 \Big\} \end{aligned} \tag{6}$$

where ρ is the Huber kernel function, which reduces the influence of outlier errors. C denotes the 3D virtual control points, and the latter part of the equation represents the reprojection errors of these virtual control points, used as terrain constraints. The coefficient α is defined as:

$$\alpha_j = \begin{cases} 1, & \text{otherwise.} \\ \dfrac{1}{\left\| Z_j - z_j \right\|}, & \text{if } \left\| Z_j - z_j \right\| \geq 3 \end{cases} \tag{7}$$

where Z_j represents the elevation of the virtual control point C_j, and z_j denotes the elevation of the global map point corresponding to C_j. As the difference between Z_j and z_j increases, α decreases, reducing the weight of the terrain constraint during optimization. In such cases, optimization relies on the odometry constraint to minimize the impact
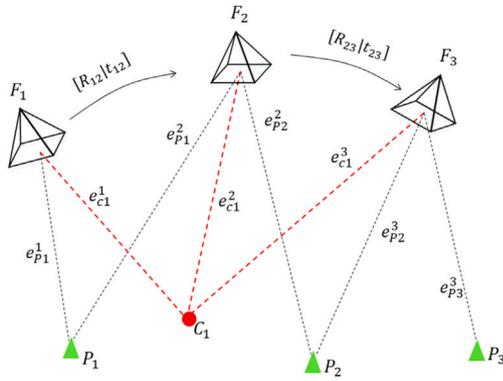
Fig. 4. The pose optimization process utilizes 3D virtual control points and global map points. In the diagram, $[R|t]$ denotes the pose of the image frame. The projection of the virtual control point C and the global map point P onto image frames F are compared with the observed positions to compute the reprojection errors e.
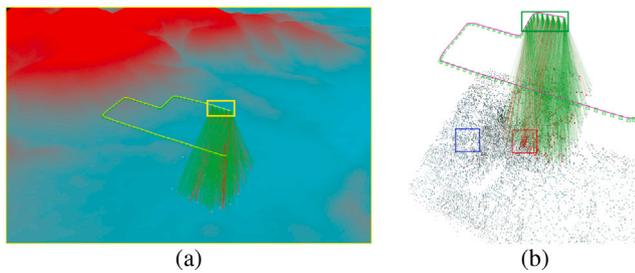

Fig. 5. The diagram illustrates the integration of sliding window and terrain constraint optimization methods. In (a), the gradient from cyan to red indicates terrain elevation variations. In (b), the colored dots inside the blue box represent global map points, the red dots inside the red box are 3D virtual control points. In the yellow box of (a), the current frame is combined with its 5 adjacent frames in a sliding window for joint pose optimization. An optimization is performed in the green box of (b) with a sliding window size of 10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of errors in DEM or map point elevation. The coefficient β mainly adjusts the weight of the terrain constraint during optimization.

$$\beta_i = \begin{cases} 0.8, & \text{otherwise.} \\ \dfrac{\text{Var}\left(Z_i\right)}{Z_{\max} - Z_{\min}}, & \text{if } \frac{\text{Var}(Z_i)}{Z_{\max}-Z_{\min}} < 0.8 \end{cases} \tag{8}$$

where $\text{Var}(Z_i)$ represents the variance of the elevation points on the DEM corresponding to the global map points in the current image frame. In flat areas with minimal terrain variation, β is small, so the optimization relies primarily on the odometry constraint. In regions with significant terrain elevation differences, β is larger, increasing the weight of the terrain constraint during optimization. By dynamically adjusting β, the relative constraints provided by odometry and the absolute constraints provided by virtual control points are weighted and integrated. This approach effectively reduces error accumulation, minimizes positioning jumps, prevents drift in the UAV's position and elevation, and ensures system re-localization after positioning failures.

Additionally, we use a sliding window mechanism to limit the optimization computation to a specific time window and improve pose estimation performance. As shown in Fig. 5, we developed a strategy integrating large and small sliding windows. For each keyframe, a sliding window covering its 5 neighboring frames is used for joint optimization, improving the pose accuracy of keyframes within the window. Additionally, for every 15 keyframes, an optimization with a sliding window size of 10 is performed to reduce cumulative positioning errors and maintain a smooth trajectory.

3.3.4. Geo-registered feature tracking and motion model

When the UAV flies over texture-rich areas, it can collect more 3D virtual control points. The system tracks these virtual control points for localization, eliminating the need for UAV–satellite image matching and enhancing efficiency. In areas with significant ground feature changes or sparse textures, matching UAV images with satellite images can be challenging and may fail to acquire enough virtual control points for PnP. In such cases, we use the velocity model from the SLAM to estimate the current frame's pose. If the poses of frames F_{n-1} and F_n are determined as Rt_{n-1} and Rt_n, respectively, the velocity v between them is computed as:

$$v = Rt_n \cdot Rt_{n-1}^{-1} \tag{9}$$

The pose Rt_{n+1} of frame F_{n+1} can be derived from the velocity model as follows:

$$Rt_{n+1} = Rt_n \cdot v \tag{10}$$

After obtaining the poses of frames F_n and F_{n+1}, we perform triangulation to obtain global map points and then apply BA optimization. This ensures localization in texture-sparse or variable environments, enhancing system stability.

4. Experiments

This section evaluates the proposed UAV visual positioning system using data from 20 real flight scenarios, conducted on a desktop with an Intel i7-10750H CPU, 32 GB RAM, Nvidia GTX 1660 Ti Mobile GPU, and Ubuntu 20.04. The system's usability and efficiency were also tested on the NVIDIA Jetson Orin NX edge device.

4.1. Introduction of experimental data

UAV Datasets. To evaluate the system's positioning accuracy and robustness under GNSS denial, we collected 20 sets of UAV image data across various scenarios, seasons, weather conditions, terrains, flight altitudes, viewpoints, flight path shapes, and thermal infrared images, as shown in Fig. 8. The test flight paths totaled 102.3 km, with altitudes ranging from 200 to 1167 m and a terrain elevation difference of up to 412 m. The dataset summary is shown in Table 2. During flights, drones recorded image positions using GPS as the ground truth for comparison. The DJI M300, DJI phantom4 RTK and DJI Mavic3E drones exhibit a horizontal and vertical positioning accuracy of ± 0.1 m in RTK mode. In contrast, the DJI Phantom3 drone has a GNSS horizontal positioning accuracy of ± 1.5 m and a vertical accuracy of ± 0.5 m.

Satellite Image. In GNSS-denied situations, the closer the acquisition time and resolution of the satellite images are to those of the UAV images, the fewer changes in observed features, resulting in better matching performance with LightGlue and higher accuracy and robustness in UAV positioning. As depicted in Fig. 6, we selected Google and ESRI satellite images from different years and seasons compared to the UAV images to evaluate the system's performance. The acquisition date of Fig. 6(a) is June 2022, image (b) was acquired in winter 2023, image (c) in October 2022, image (d) in summer 2023, and the acquisition dates for images (e–h) are July 2022, June 2023, June 2019, and June 2014, respectively. For image resolution, we chose satellite images at levels 16 to 18 to handle altitude changes and terrain variations.

Terrain Data. Fig. 7 shows the 12.5 m resolution DEM from the Advanced Land Observing Satellite (ALOS) in the experimental region. The DEM accurately depicts terrain undulations in hilly and mountainous areas (Fig. 7(b) and (f)). However, it fails to capture urban building forms due to resolution limits (Fig. 7(c)–(e)), leading to substantial errors in Z-values of virtual control points from urban areas and complicating GNSS-denied positioning.

TerrainFusion Dataset. The TerrainFusion dataset (Wang et al., 2019) includes several scenarios such as cities, plains, mountains, and deserts. A summary of the dataset is presented in Table 1. This dataset
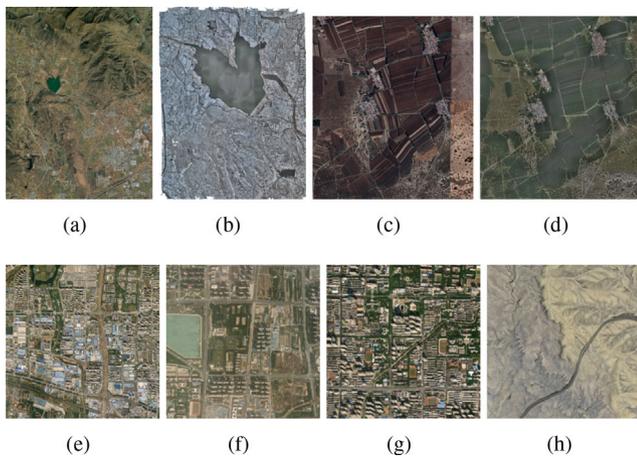
Fig. 6. Reference images for positioning tests. (a) and (b) are for sets a-c; (c) and (d) are for sets d-i and r-t; (e), (f), and (g) are for sets m-n, o, and p; and (h) is for set q. Images (b) and (d) are UAV orthophotos used in sets c and g.

Table 1
Characteristics of the TerrainFusion datasets.

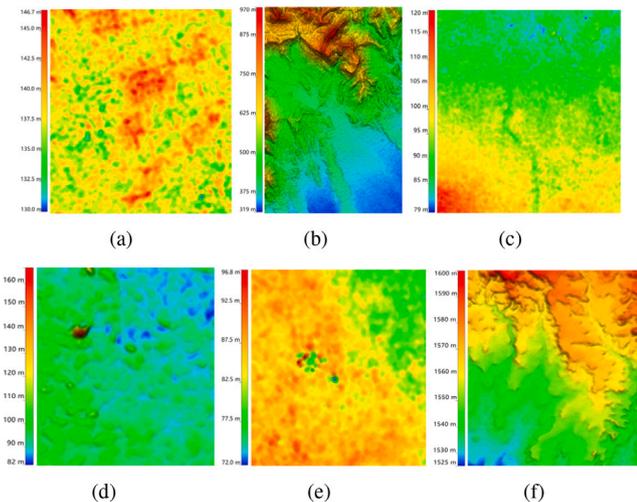| Set	Traj. (km)	Alt. (m)	Speed (m/s)	Area
olathe	3.7	150	10.1	1.59 km^2
mound60 m	0.25	50	13.4	1530 m^2
mountainlong	9.25	127	11.2	1.03 km^2
fengniao	0.59	78	15.7	3789.2 m^2
village	8.32	196	17.4	0.932 km^2
factory	4.43	126	16.3	0.432 km^2

Fig. 7. DEM data for the experiments are as follows: (a) corresponds to satellite images in Fig. 6(a)–(b); (b) corresponds to Fig. 6(c)–(d); and (c)–(f) corresponds to Fig. 6(e)–(h).

features typical UAV operational scenarios with significant variations in flight speed, flight path length, and operational area. The UAV flight paths are relatively regular, with images captured by the UAV camera oriented vertically downward. We utilize the UAV's GPS-recorded trajectory as ground truth for experimental comparisons. Since the flight altitude of the mound60 m data is too low to meet our method's testing requirements, we exclude it from the experiments.

4.2. Metric analysis

We compared each positioning method to the ground truth from the drone GNSS, reporting the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) in X, Y, and Z coordinates.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(PC_i - PT_i\right)^2}{n}} \tag{11}$$

$$MAE = \frac{\sum_{i=1}^{n} |PC_i - PT_i|}{n} \tag{12}$$

where PC denotes the calculated UAV coordinates and PT represents the GNSS ground truth. MAE reflects the positioning error magnitude, with a smaller MAE indicating higher accuracy. RMSE indicates significant anomalies in positioning, with a smaller RMSE suggesting more consistent accuracy and a smoother trajectory.

4.3. Localization performance

Fig. 10 shows the positioning results for the 20 test data from Fig. 8, demonstrating that our system effectively achieves GNSS-denied positioning in various complex scenarios. We will now analyze the system's performance.

System Performance during diverse Trajectory shapes and Varying Flight Altitude. As illustrated in Fig. 10, with effective keyframe extraction and rotation invariance strategies, our system has achieved 3D UAVs positioning under complex flight paths. Altitude changes during flight can cause significant discrepancies in image resolution and geographic coverage, complicating UAV positioning in GNSS-denied environments. Unlike the fixed flight heights in He et al. (2023) and Hou et al. (2020), the UAV in sets k and l experienced up to 360 m altitude changes, while in set p, it flew between 281 m and 500 m. Our system employs flexible satellite image scheduling and aggregation strategies, along with image pre-rotation and homologous points search methods, to improve image matching accuracy and ensure precise UAV positioning under complex flight paths and altitude variations.

System Performance under Image Obliquity and Rotation. As illustrated in Fig. 9(a), when capturing images vertically, the center point L1 of the UAV image can be considered the UAV's position after matching with the satellite image. When the UAV's camera is obliqued during capture, a deviation occurs between the image center point L2 and the actual UAV position, increasing with a larger oblique angle. Additionally, tall buildings in obliqued images obstruct surrounding features, posing further challenges for UAV positioning. Therefore, Kinnari et al. (2021), Bianchi and Barfoot (2021) and Goforth and Lucey (2019) require the UAV's camera to capture images directly downward or use angle information from the IMU to perform orthorectification, minimizing the negative effects of image oblique and rotation.

Image rotation affects overlaps and presents significant challenges for image matching. To mitigate the negative impacts of image oblique and rotation on UAV positioning, we apply a pre-rotation strategy to the image and search for precise correspondences through BHPS and use 2D–3D Geo-Registration to obtain high-quality 3D virtual control points for PnP solving to determine the UAV's pose, rather than simply using the matched image center point as the UAV's position. This approach ensures the system has rotation-invariant and viewpoint-robust capabilities during positioning. In set k, where the UAV with a 20° oblique camera performs near 360° rotations, our system still achieves stable positioning.

System Performance during High Altitude, Long-distance Flights, and Rolling Terrain. We evaluated the system's positioning performance in high-altitude and long-distance scenarios using five datasets from sets h to l. In set k, the UAV flew 9.3 km at a minimum altitude of 722 m, with terrain elevation changes exceeding 400 m. The total flight length of the datasets exceeded 100 km. To address these challenges, We use high-quality 3D virtual control points, invariant to both viewpoint and rotation, for initial pose estimation, followed by optimization of the UAV's pose with global map points and virtual control points. Additionally, dynamic updating of satellite image blocks and sliding window optimization maintain efficiency over long distances and large

Table 2
Characteristics of the UAV datasets (ED: Elevation Difference).

| Set	Location	Scene	Traj. (m)	Camera	ED (m)	Alt. (m)	Yaw (°)	Pitch (°)	Flight date
a	Zhaoyuan	Village	2333	DJI H20t	9.5	200	−179.9~179.8	−90.2	Jul. 2022
b	Zhaoyuan	Village	8483	DJI H20t	13.1	389~500	−179.9~179.8	−122~−77.8	Sept. 2023
c	Zhaoyuan	Village	3404	DJI H20t	10.8	399~498	−179.9~179.8	−119~−59	Sept. 2023
d	Dengfeng	Village	3233	DJI P4R	43.8	500	−1.5~0.8	−89.8	Sept. 2020
e	Dengfeng	Forest	2279	DJI P4R	70.2	500~580	−179.9~179.9	−89.7	Jun. 2021
f	Dengfeng	Hilly	3288	DJI P4R	45.8	500	−179.9~179.9	−89.9	Mar. 2022
g	Dengfeng	Hilly	3343	DJI P4R	33.7	500	−179.9~179.9	−89.9	Mar. 2022
h	Dengfeng	Towns	10 068	DJI H20t	88.9	500	−179.9~179.9	−89.7	Mar. 2023
i	Dengfeng	Mountain	7219	DJI H20t	201.5	500	−179.9~179.9	−89.8	Mar. 2023
j	Dengfeng	Hilly	10 529	DJI H20t	80.8	381~500	−179.8~178.7	−98.2~−65	Jun. 2023
k	Dengfeng	Mountain	9353	DJI H20t	412.6	722~1045	−179.8~178.7	−97.3~−70.9	Jan. 2024
l	Dengfeng	Mountain	5125	DJI H20t	195.3	799~1167	−179.8~178.7	−91.2~−66.8	Jan. 2024
m	Zhengzhou	City	4282	DJI H20t	12.1	500	−5~173.3	−89.9	May. 2023
n	Zhengzhou	City	5396	DJI H20t	12.1	342~500	−179.9~179.8	−120~−80	May. 2023
o	Zhengzhou	City	4855	DJI H20t	13.2	500	−179.9~179.8	−99.3~−76	Dec. 2023
p	Zhengzhou	City	5496	Mavic 3E	12.9	219~500	−179.9~180	−90~−64.7	Feb. 2024
q	Olathe	Desert	3702	Phantom3	45.8	150	−179.9~180	−89.8	May. 2016
r	Dengfeng	Hilly	4464	H20t Infrared	79.2	500	−30.4~31	−90.4	Jan. 2021
s	Dengfeng	Hilly	2487	H20t Infrared	50.6	500	6.9	−90.4	Jan. 2021
t	Dengfeng	Villages	2962	H20t Infrared	85.3	500	−179.8~179.8	−89.6	Jan. 2021
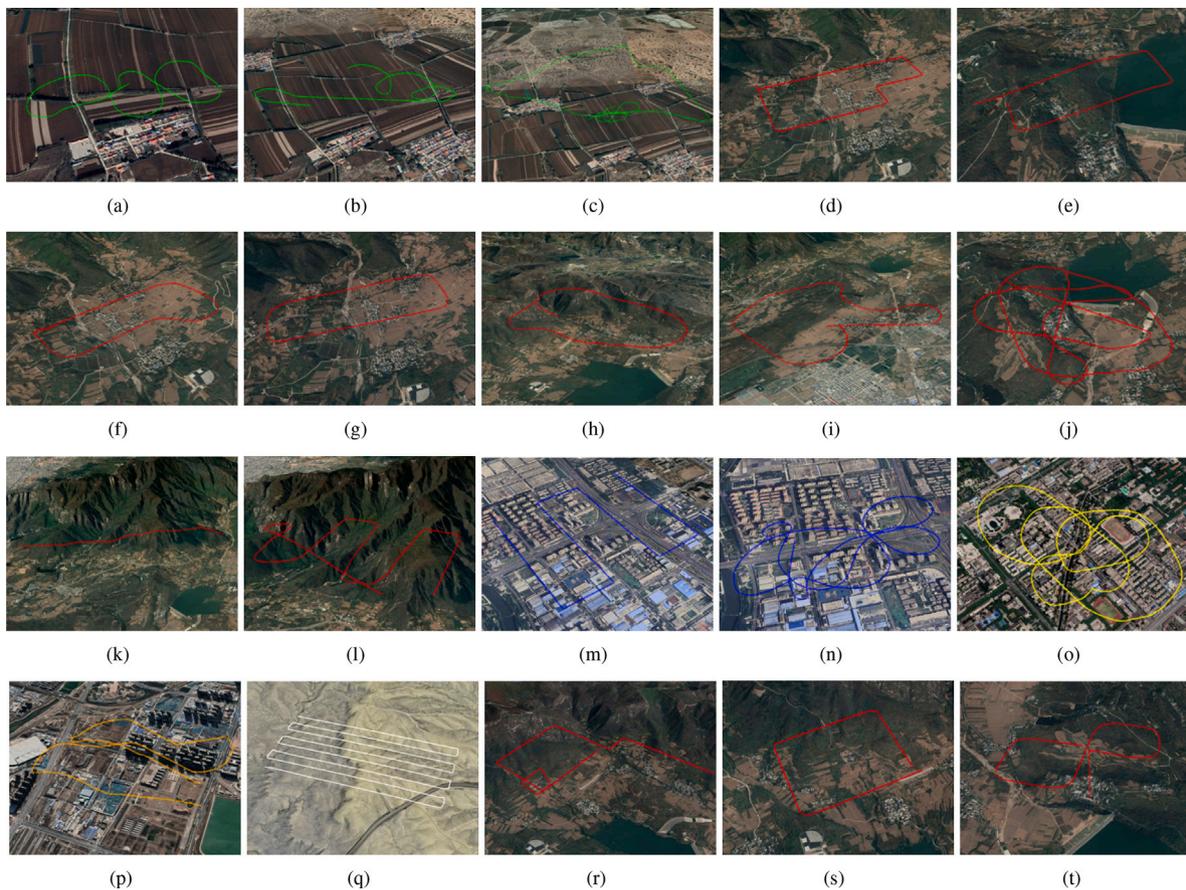
Fig. 8. The 20 sets of UAV data used in the experiments. (a)–(c) flat rural areas, (d)–(l) hilly and mountainous regions, (m)–(p) flat urban areas, (q) desert, and (r)–(t) hilly regions. The colored lines in the figures represent the UAVs' actual flight trajectories.

areas. In He et al. (2023), Kinnari et al. (2021) and Kinnari et al. (2022), UAV heights are under 100 m, while in Choi and Myung (2020), Bianchi and Barfoot (2021), and Patel et al. (2020), flight paths do not exceed 2 km, often in flat urban areas.

System Performance in Regions with Sparse Texture. In sparse texture regions (Fig. 11), finding reliable 3D virtual control points for UAV pose determination is challenging. In these cases, we adapt SLAM positioning methods by using co-visibility geometric relationships between frames. However, without 3D virtual control points for terrain constraints, the

UAV pose estimated with the velocity model may drift. As shown in Fig. 12, LightGlue struggles with image matching in sparse texture regions, leading to increasing pose errors. When the UAV moves to more textured areas, our system uses terrain-constraint optimization to correct the pose and relocalize, preventing cumulative error propagation seen in SLAM. Additionally, the system uses 3D virtual control points to constrain and optimize positioning, calculating each frame's absolute pose and scale in the world coordinate system. Thus, it does not rely on loop closures for drift correction to reduce error accumulation like
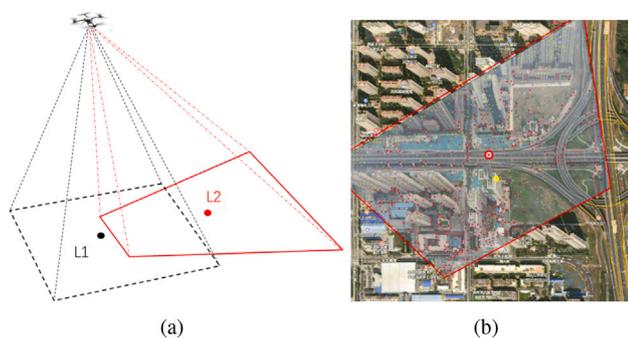
Fig. 9. The impact of oblique and rotated UAV images on positioning. (b) shows the result of registering and overlaying the oblique UAV image with the satellite image. The red dots indicate the image center points, corresponding to point L2 in (a), and the yellow dots represent the centers of all extracted Superpoint features. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

changes, leading to improved image matching and more accurate 3D virtual control points for UAV precise positioning in GNSS-denied environments. Due to satellite revisit cycles and other constraints, obtaining satellite images close to the UAV image dates may not be feasible. As shown in Fig. 13, there are noticeable seasonal changes between the UAV and satellite images in (a)–(c). The variations in vegetation and water bodies in (d)–(f) reflect significant differences in the acquisition times of the UAV images and satellite maps. In (g)–(i), the weather and lighting conditions of the UAV images and satellite maps are distinctly different. To address this, we pre-rotate the UAV images and use a velocity model to estimate the UAV's pose based on the co-visual geometric relationships between image frames. Despite potential errors with the velocity model, the system quickly corrects and repositions once image matching improves. This allows effective UAV visual positioning in GNSS-denied environments, even in sparse texture areas or under seasonal variations. In Kinnari et al. (2022), a CNN network was trained to handle appearance and viewpoint differences, achieving effective UAV image matching across seasons. However, this method did not account for other factors, resulting in positioning errors of about 30 m and limiting accuracy to 2D positioning.

System Performance During Night-time. Visible light cameras struggle to capture sufficient light for night-time imaging, making night-time localization challenging. As shown in Fig. 14, we collected three sets of thermal infrared UAV images — sets r, s, and t — to evaluate the system's performance in night-time localization. Due to significant imaging differences, LightGlue fails to match visible light satellite maps with thermal infrared images, so we use LoFTR (Sun et al., 2021) for

SLAM. In set k, the system achieves an average positioning accuracy of 4.455 m, even in straight-line flight. Our optical camera-based system ensures accurate localization unaffected by GNSS signals and is highly resistant to interference.

System Performance under Seasonal, Lighting, and Weather Variations. The closer the acquisition time and weather conditions of the UAV images and satellite maps, the smaller the ground feature and lighting
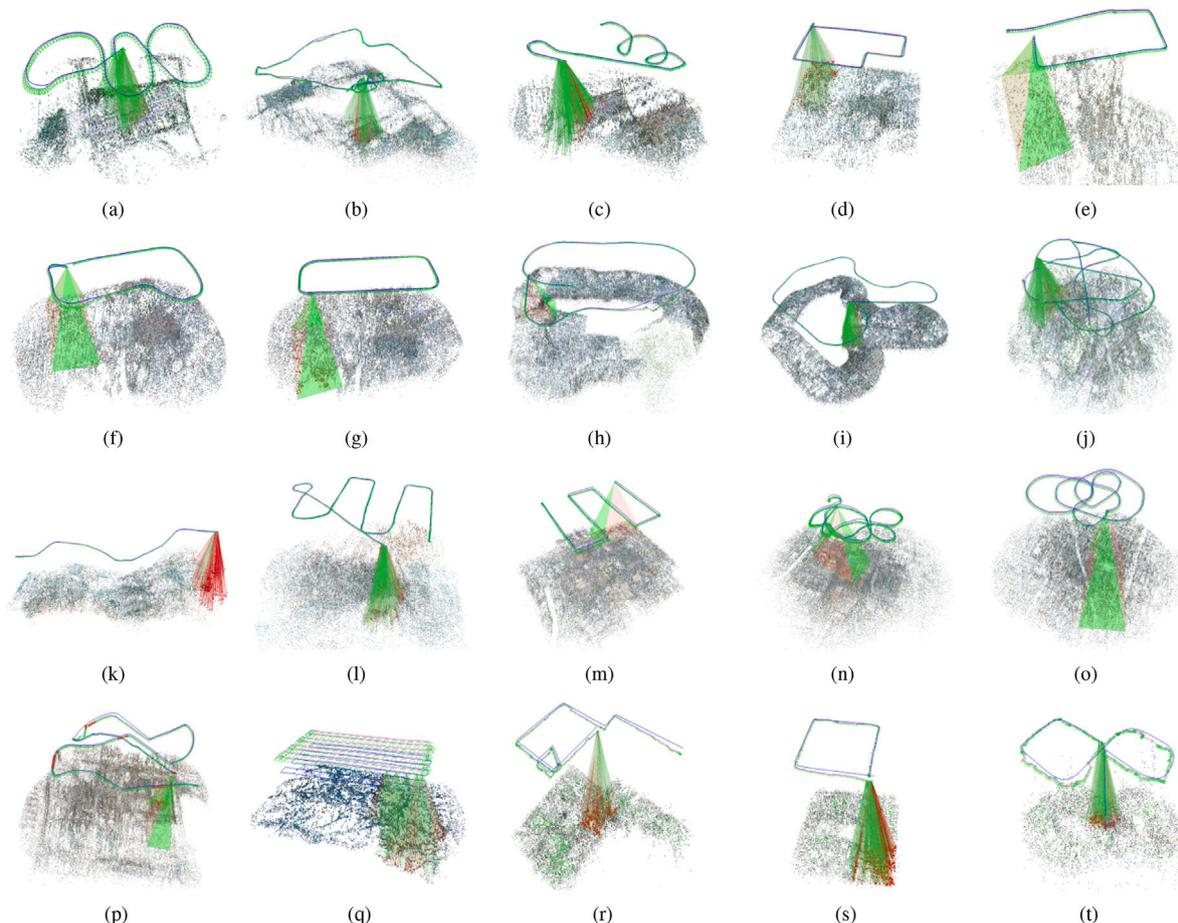

Fig. 10. Positioning results for the 20 data sets. The red curve indicates the UAV trajectory computed by the system, the blue curve shows the GNSS-recorded ground truth trajectory, and the colored points represent the global map points generated by the system. In (a)–(d), green lines link the UAV trajectory to the map points, showing the connections between image frames and their corresponding map points. In (e)–(h), pyramids between the trajectory lines and map points depict image geographic coverage. In some areas, the small error between the system's positioning and the ground truth causes the red and blue trajectories to nearly overlap. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
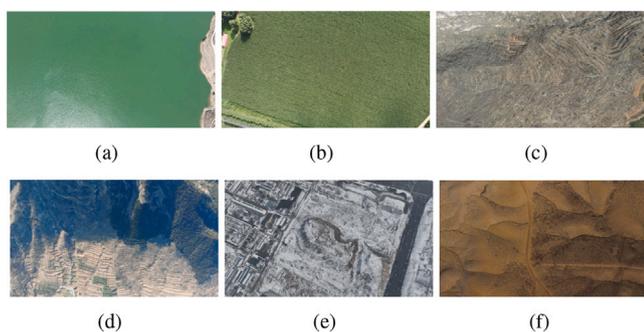
Fig. 11. UAV images of sparse texture regions. (a) shows the water area in set i. (b) depict cornfields from sets a. (c) shows the bare mountainous area in set i. (d) illustrates shadows caused by mountain obstructions in set l. (e) shows the snow-covered area in set p. (f) depicts the desert area in set q.
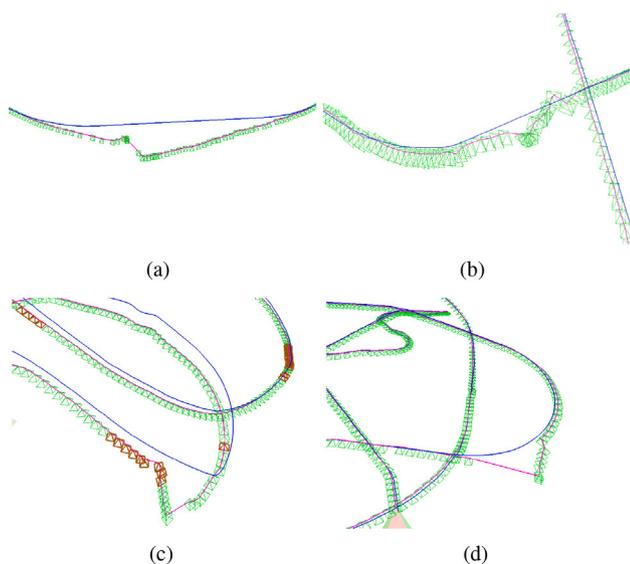

Fig. 12. System's re-localization capability. The blue line represents the GNSS-recorded UAV trajectory, while the red line represents the system-calculated trajectory. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 13. Overlay images show the result of UAV images matched with satellite maps under various scenarios, seasons, weather conditions, and lighting. UAV images are framed in red, with red dots marking geometric centers and yellow dots representing feature point centers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

end-to-end feature extraction and matching. Since end-to-end matching does not support bidirectional feature point association, we use nearest neighbor search to find corresponding points. However, the thermal infrared camera's small field of view and low resolution affect feature point accuracy, leading to positioning errors of about 20 m, which is worse than visible light images. Additionally, thermal infrared cameras provide higher contrast and more precise details at night, resulting in better night-time localization accuracy compared to daytime.

System Performance on TerrainFusion Dataset. As illustrated in Fig. 15, our system accomplished the positioning challenges in these complex scenarios, with results summarized in Table 3. While the TerrainFusion Dataset contains relatively straightforward flight paths, the sparse scene textures and low UAV flight altitudes pose challenges for our UAV positioning method. According to the experimental results, our method achieves the highest positioning accuracy in the y and z directions, while the matching method performs best in the x direction. This is due to notable deviations in keyframe positioning results in the mountainlong and village datasets, leading to larger positioning errors in the x direction. However, our system benefits from terrain constraint optimization, enabling it to correct UAV pose and re-localize even with significant deviations in key frame positioning, unlike SLAM, where positioning errors tend to accumulate and propagate. Additionally, our
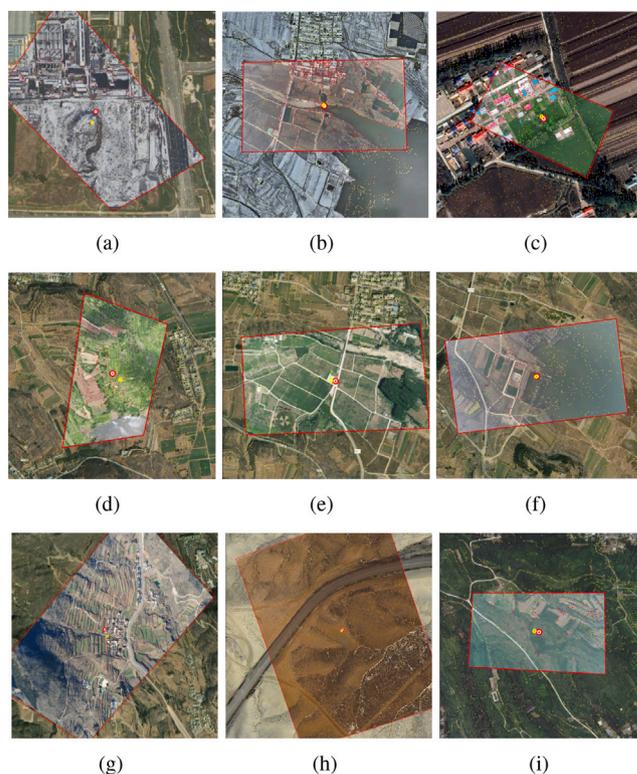
system does not require the UAV camera to be oriented vertically downward, as is necessary for matching-based methods. Moreover, the system can provide 3D positioning information for the UAV in GNSS-denied environments.

4.4. Localization precision comparison

Ablation experiment. In the ablation experiment, we disabled the terrain-weighted constraints optimization module to assess its effect on positioning accuracy, meaning that we did not use the terrain elevation information from the DEM or 3D virtual control points for UAV optimization localization. As shown in Fig. 16 and Table 4, removing the module allowed for absolute positioning but introduced significant drift and scaling errors. Without this module, the system's positioning, similar to SLAM, relied solely on inter-frame relationships and lacked terrain constraints. This absence hinders effective UAV pose correction through virtual control points, causing errors to accumulate over time, leading to drift, scaling errors, and inability to relocalize. In contrast, our method's positioning closely matches the ground truth, highlighting the precision of our system and demonstrating the module's effectiveness in controlling error accumulation and reducing drift in GNSS-denied conditions.

To assess the impact of DEM resolution on UAV localization accuracy, we used the 12.5 m ALOS DEM and the 30 m Copernicus DEM for test data Set d. With other experimental parameters constant, the MAE values for UAV localization with the Copernicus DEM were 3.882 m, 14.978 m, and 24.414 m, while those with the ALOS DEM were 3.785 m, 15.047 m, and 4.676 m. These results indicate that the higher-resolution ALOS DEM enhances UAV localization accuracy under GNSS-denied conditions.

Table 3
MAE and RMSE results for localization of the TerrainFusion datasets.

Set	Ours			Ablation study			EPnP			Matching	
	x (m)	y (m)	z (m)	x (m)	y (m)	z (m)	x (m)	y (m)	z (m)	x (m)	y (m)
MAE	22.207	**9.829**	**8.883**	36.058	22.628	44.253	28.207	20.483	12.225	**11.3**	11.370
RMSE	31.389	**21.413**	**14.356**	44.365	26.012	52.907	41.573	35.520	20.012	**23.756**	28.597

Table 4
MAE and RMSE results for localization of 20 UAV datasets.

Set	Ours			Ablation study			EPnP			Matching	
	x (m)	y (m)	z (m)	x (m)	y (m)	z (m)	x (m)	y (m)	z (m)	x (m)	y (m)
MAE	**6.701**	**8.178**	**6.966**	32.241	33.866	29.813	25.402	37.423	13.998	33.207	35.395
RMSE	**9.849**	**12.229**	**10.124**	37.960	38.122	36.845	51.326	68.789	27.943	73.684	69.062

Table 5
Summary of other method positioning results.

| Author & Year	Method	Experiments	Known initial	Accuracy	Times
Chiu et al. (2014)	Image matching + SLAM	Flight length: 38.9/26.5 km; Sensors: IMU and camera; Map: 3D map	Yes	9.35 m	1 Hz
Shetty and Gao (2019)	Deep learning + VO	UAV simulation imagery; Map: 2D Google map	Yes	36 m	0.82 s
Patel et al. (2020)	Image matching + VO	Flight length: 1.13 km; Flight height: 36–42 m; Sensor: camera with a gimbal; Map: Google Earth map	Yes	1.39/0.86/0.38 m	1 Hz
Kinnari et al. (2021)	Image matching	Flight length: 6.8/4.0/6.3 km; Flight height: 92 m; Sensor: IMU; Map: 2D map	Yes	<20 m	–
He et al. (2023)	Image Retrieval + SLAM	Flight length: 0.88/1.0 km; Flight height: 50 m; Sensors: IMU and nadir-facing camera; Map: Google map	No	19.38 m	1.93 Hz
Ye et al. (2024)	Image Retrieval + image matching	Flight height: 0.15/4 km; Map: BJ-2 satellite images	No	0.1/10 m	0.42 s
Chen and Jiang (2023)	Image matching + PnP	Flight length: 0.85/1.76 km; Flight height: 169/325 m; Map: 2D map; coverage area <3 km^2	No	5.54/5.2 m	–
Ours	Image matching + SLAM	Flight length: 102.3 km; Flight height: 200~1167 m; Map: 2D map + DEM	Yes	6.7/8.1/6.9 m	0.74/0.6 s

(a) (b)

(c) (d)
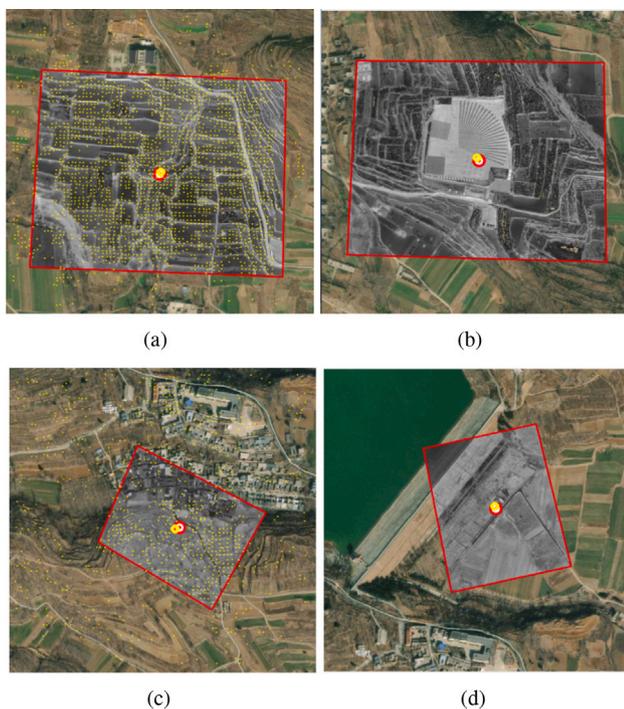
Fig. 14. Overlay of thermal infrared UAV images with satellite maps. The red box highlights the thermal infrared UAV images. In (a) and (c), the small yellow dots indicate feature points extracted by LoFTR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
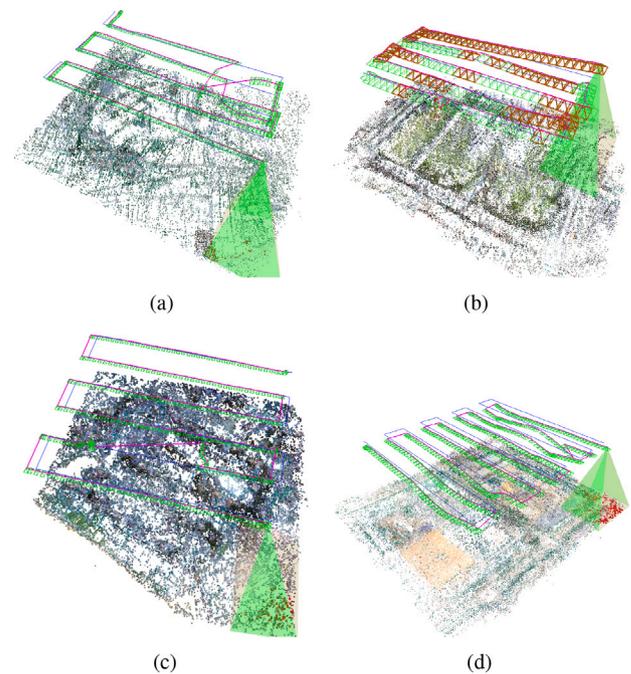

(a) (b)

(c) (d)

Fig. 15. Positioning results for the TerrainFusion dataset. (a) to (d) represent the positioning results for the mountainlong, fengniao, village, and factory datasets, respectively.
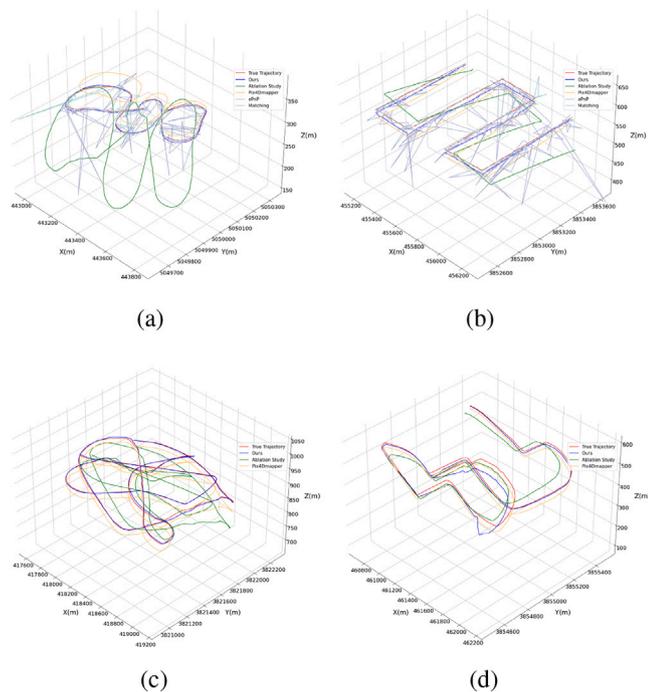
Fig. 16. Comparison of positioning trajectories. Lines of different colors represent the trajectories of various methods. Since the matching positioning method provides only X and Y coordinates, GNSS-recorded Z values are used for the Z-axis in the matching method. For clarity, (c) and (d) display only the trajectories of our method, the ablation experiment method and Pix4Dmapper.

Comparison of Our Method with Match and PnP Localization. Table 4 presents MAE and RMSE statistics for the 20 datasets. Our method significantly outperforms both EPnP and matching positioning methods, which are comparable to GNSS positioning. Furthermore, the proposed method exhibits the smallest average RMSE, indicating fewer anomalies, stable accuracy, and smoother trajectories. The matching positioning method assumes flat terrain and requires vertical UAV images, making accuracy sensitive to terrain variations and camera angle, and it does not provide altitude information. In set k, with mountainous data, the MAE for matching positioning is 117.443 m, with a high RMSE. Sparse textures make obtaining sufficient virtual control points for PnP-based positioning difficult. As shown in Fig. 16(a) and (b), positioning results exhibit significant errors and trajectory jumps. Moreover, both matching and PnP methods neglect geometric relationships between adjacent images, making relocalization impossible after a positioning failure.

Comparison of Our Method with Pix4Dmapper. We import the keyframe and their corresponding GPS coordinates extracted during our system's operation as POS data into the Pix4Dmapper software, which then uses SfM (Structure from Motion) optimization to calculate the pose of each keyframe. As illustrated in Fig. 16, the UAV trajectory computed by Pix4Dmapper closely aligns with the GPS trajectory in shape, exhibiting a smooth trajectory that highlights the high precision of SfM-based software. However, due to the absence of Ground Control Points for Bundle Adjustment, the UAV trajectory calculated by Pix4Dmapper exhibits systematic drift. Although such software achieves high computational accuracy, it requires significant processing resources and is suitable only for offline data processing, lacking real-time capabilities. Therefore, it is unsuitable for real-time UAV positioning in GNSS-denied environments. While our system's accuracy and trajectory smoothness may not match that of Pix4Dmapper, it is better suited for UAV positioning in dynamic scenarios, fulfilling real-time positioning needs in GNSS-denied conditions.

Comparison of Our Method with Other Research. Due to the lack of publicly available code and standard datasets, we summarized and compared several recent research methods, as shown in Table 5. Patel et al. (2020) and Chiu et al. (2014), like ours, require an initial position; but our system achieves 3D positioning without barometric altimeters or IMUs. Kinnari et al. (2021) requires a downward camera and relies solely on image matching, which cannot ensure trajectory continuity. He et al. (2023) uses a retrieval approach where each frame is positioned independently, providing good accuracy but requiring the UAV to maintain a fixed altitude and heading. Ye et al. (2024) excels in texture-rich environments. Shetty and Gao (2019) uses deep networks for cross-view positioning but only achieves 36 m accuracy. Chen and Jiang (2023) handles highly oblique images but was tested only in urban settings and short flights, not covering complex scenarios such as high altitude, long distances, and terrain variations. Our system uses virtual control points for pose estimation and optimization, preventing error accumulation and avoiding loop closure correction, with relocalization capability after failure. In GNSS-denied conditions, our system, which requires no additional equipment, achieves MAE values of 6.701 m, 8.178 m, and 6.966 m for visual-only positioning, with values for sets c, h, and g under 5 m. While methods Patel et al. (2020) and Chen and Jiang (2023) also achieve under 5 m accuracy, our approach handles various complex conditions, including diverse flight paths, altitudes, image tilts, terrain, and seasonal changes. Additionally, existing methods rarely address night-time UAV positioning under GNSS denial. In contrast, our system demonstrates the feasibility and effectiveness of using thermal infrared imagery for night-time positioning, which opens the possibility of achieving all-day, all-weather navigation under GNSS denial.

4.5. Computational efficiency

Effective strategies for satellite image chunking, dynamic scheduling, geo-registered feature tracking, and sliding window optimization result in average keyframe positioning times of 0.74 s per frame for visible light images and 0.609 s per frame for thermal infrared images in our system. Other methods also achieve positioning times of around 1 s, which is sufficient for UAV positioning in GNSS-denied scenarios. We deployed and validated the system's accuracy and effectiveness on the NVIDIA Jetson Orin NX, the average keyframe positioning time is 3.25 s.

5. Conclusions

This article presents a high-precision, robust UAV visual localization system for GNSS-denied scenarios day and night, integrating image matching, visual odometry, and terrain-weighted constraint optimization. The system demonstrated excellent adaptability and positioning performance across 20 UAV test datasets. Compared to existing methods, our system achieves pure visual 3D absolute positioning under GNSS denial without relying on altimeters or IMUs. It does not require a top-down camera and remains robust against UAV rotation, altitude variations, trajectory shapes, rolling terrain, feature-poor environments, and seasonal changes. Additionally, the system avoids the need for loop correction, prevents error accumulation and pose drift, and enables re-localization after positioning failure. Our system also addresses night-time denial positioning challenges. The average keyframe positioning time is under 1 s, with mean MAE values for XYZ positioning across 20 UAV datasets being 6.701 m, 8.178 m, and 6.966 m. This performance achieves GNSS-comparable accuracy and can be a robust supplement to GNSS in denial environments. In the future, We will explore multi-sensor fusion localization techniques to enhance the system's robustness and improve positioning efficiency.

CRediT authorship contribution statement

Fushan Yao: Writing – original draft, Resources, Methodology, Formal analysis, Conceptualization. **Chaozhen Lan:** Writing – review & editing, Supervision, Conceptualization. **Longhao Wang:** Writing – review & editing, Software, Formal analysis. **Hongfa Wan:** Writing – review & editing, Software, Formal analysis. **Tian Gao:** Writing – review & editing, Software, Formal analysis. **Zijun Wei:** Writing – review & editing, Software, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Ali-Bey, A., Chaib-Draa, B., Giguere, P., 2023. Mixvpr: Feature mixing for visual place recognition. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2998–3007.

Bianchi, M., Barfoot, T.D., 2021. UAV localization using autoencoded satellite images. IEEE Robot. Autom. Lett. 6 (2), 1761–1768.

Chan, Y.C., Yakimenko, O., 2022. Assessment of image-matching technique for tactical aerial platforms to navigate in the GPS-degraded environments. In: 2022 17th International Conference on Control, Automation, Robotics and Vision. ICARCV, IEEE, pp. 692–700.

Chen, Y., Jiang, J., 2023. An oblique-robust absolute visual localization method for GPS-denied UAV with satellite imagery. IEEE Trans. Geosci. Remote Sens..

Chiu, H.-P., Das, A., Miller, P., Samarasekera, S., Kumar, R., 2014. Precise vision-aided aerial navigation. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 688–695.

Choi, J., Myung, H., 2020. BRM localization: UAV localization in GNSS-denied environments based on matching of numerical map and UAV images. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 4537–4544.

Couturier, A., Akhloufi, M.A., 2021. A review on absolute visual localization for UAV. Robot. Auton. Syst. 135, 103666.

Dai, M., Zheng, E., Feng, Z., Qi, L., Zhuang, J., Yang, W., 2023. Vision-based UAV self-positioning in low-altitude urban environments. IEEE Trans. Image Process..

Davison, 2003. Real-time simultaneous localisation and mapping with a single camera. In: Proceedings Ninth IEEE International Conference on Computer Vision. IEEE, pp. 1403–1410.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 224–236.

Goforth, H., Lucey, S., 2019. GPS-denied UAV localization using pre-existing satellite imagery. In: 2019 International Conference on Robotics and Automation. ICRA, IEEE, pp. 2974–2980.

Hamidi, M., Samadzadegan, F., 2015. Precise 3D geo-location of UAV images using geo-referenced data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 40, 269–275.

He, Y., Cisneros, I., Keetha, N., Patrikar, J., Ye, Z., Higgins, I., Hu, Y., Kapoor, P., Scherer, S., 2023. Foundloc: Vision-based onboard aerial localization in the wild. arXiv preprint arXiv:2310.16299.

Hou, H., Xu, Q., Lan, C., Lu, W., Zhang, Y., Cui, Z., Qin, J., 2020. UAV pose estimation in GNSS-denied environment assisted by satellite imagery deep learning features. IEEE Access 9, 6358–6367.

Hu, B., Chen, L., Chen, R., Bu, S., Han, P., Li, H., 2024. CurriculumLoc: Enhancing cross-domain geolocalization through multi-sage refinement. IEEE Trans. Geosci. Remote Sens..

Jun, M., Lilian, Z., Xiaofeng, H., Hao, Q., Xiaoping, H., 2022. A 2d georeferenced map aided visual-inertial system for precise uav localization. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 4455–4462.

Jurevičius, R., Marcinkevičius, V., Šeibokas, J., 2019. Robust GNSS-denied localization for UAV using particle filter and visual odometry. Mach. Vis. Appl. 30 (7), 1181–1190.

Kinnari, J., Renzulli, R., Verdoja, F., Kyrki, V., 2023. LSVL: Large-scale season-invariant visual localization for UAVs. Robot. Auton. Syst. 168, 104497.

Kinnari, J., Verdoja, F., Kyrk, V., 2022. Season-invariant GNSS-denied visual localization for UAVs. IEEE Robot. Autom. Lett. 7 (4), 10232–10239.

Kinnari, J., Verdoja, F., Kyrki, V., 2021. GNSS-denied geolocalization of UAVs by visual matching of onboard camera images with orthophotos. In: 2021 20th International Conference on Advanced Robotics. ICAR, IEEE, pp. 555–562.

Li, H., Liu, Z., Lyu, Y., Wu, F., 2023a. Multimodal image registration for GPS-denied UAV navigation based on disentangled representations. In: 2023 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 1228–1234.

Li, H., Wang, J., Wei, Z., Xu, W., 2023b. Jointly optimized global-local visual localization of UAVs. arXiv preprint arXiv:2310.08082.

Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. Lightglue: Local feature matching at light speed. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 17627–17638.

Liu, Y., Bai, J., Wang, G., Wu, X., Sun, F., Guo, Z., Geng, H., 2023. Uav localization in low-altitude gnss-denied environments based on poi and store signage text matching in uav images. Drones 7 (7), 451.

Luo, H., Li, G., Zou, D., Li, K., Li, X., Yang, Z., 2023. UAV navigation with monocular visual inertial odometry under GNSS-denied environment. IEEE Trans. Geosci. Remote Sens..

Luo, X., Wan, X., Gao, Y., Tian, Y., Zhang, W., Shu, L., 2024. JointLoc: A real-time visual localization framework for planetary UAVs based on joint relative and absolute pose estimation. arXiv preprint arXiv:2405.07429.

Mei, C., Fan, Z., Zhu, Q., Yang, P., Hou, Z., Jin, H., 2023. A Novel scene matching navigation system for UAVs based on vision/inertial fusion. IEEE Sens. J. 23 (6), 6192–6203.

Mur-Artal, R., Tardós, J.D., 2017. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. IEEE Trans. Robot. 33 (5), 1255–1262.

Patel, B., Barfoot, T.D., Schoellig, A.P., 2020. Visual localization with google earth images for robust global pose estimation of uavs. In: 2020 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 6491–6497.

Qiu, X., Yang, D., Liao, S., Wang, S., Li, Y., 2024. Image moment extraction based aerial photo selection for UAV high-precision geolocation without GPS. Measurement 226, 114141.

Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4938–4947.

Shan, M., Wang, F., Lin, F., Gao, Z., Tang, Y.Z., Chen, B.M., 2015. Google map aided visual navigation for UAVs in GPS-denied environment. In: 2015 IEEE International Conference on Robotics and Biomimetics. ROBIO, IEEE, pp. 114–119.

Shetty, A., Gao, G.X., 2019. Uav pose estimation using cross-view geolocalization with satellite imagery. In: 2019 International Conference on Robotics and Automation. ICRA, IEEE, pp. 1827–1833.

Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X., 2021. LoFTR: Detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8922–8931.

Wang, Z., Shi, D., Qiu, C., Jin, S., Li, T., Shi, Y., Liu, Z., Qiao, Z., 2024. Sequence matching for image-based UAV-to-satellite geolocalization. IEEE Trans. Geosci. Remote Sens..

Wang, W., Zhao, Y., Han, P., Zhao, P., Bu, S., 2019. TerrainFusion: Real-time digital surface model reconstruction based on monocular SLAM. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 7895–7902.

Xiao, J., Zhang, N., Tortei, D., Loianno, G., 2024. STHN: Deep homography estimation for UAV thermal geo-localization with satellite imagery. arXiv preprint arXiv:2405.20470.

Ye, Q., Luo, J., Lin, Y., 2024. A coarse-to-fine visual geo-localization method for GNSS-denied UAV with oblique-view imagery. ISPRS J. Photogramm. Remote Sens. 212, 306–322.