12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

UAV POSITION ESTIMATION THROUGH IMAGE MATCHING BETWEEN DRONE IMAGERY AND MAPS

RADE PAVLOVIĆ
Military Technical Institute, Ministry of Defence, Ratka Resanovića 1, Belgrade, rade.pavlovic@mod.gov.rs

SARA PAVLOVIĆ
Union University – School of Computing, 6/VI Kneza Mihaila, Belgrade, sarapavlovic1977@gmail.com

Abstract: *Accurate position estimation is a fundamental requirement for the autonomous navigation of unmanned aerial vehicles (UAVs). Most modern drones rely on Global Navigation Satellite Systems (GNSS), such as GPS, to determine their geographical position. However, satellite navigation can become unreliable in environments with signal obstruction, interference, or poor satellite visibility. An alternative approach is visual localization, which estimates the position of a drone by comparing images captured by its onboard camera with reference maps or georeferenced satellite imagery. This paper presents the principles and methods of UAV position estimation through image matching between drone imagery and maps. The main stages of the process include image acquisition, preprocessing, feature extraction, image matching, geometric transformation, and geographical coordinate estimation. Particular attention is given to traditional computer vision techniques, such as Scale-Invariant Feature Transform (SIFT), Oriented FAST and Rotated BRIEF (ORB), and feature matching using geometric models. The paper also discusses deep learning-based approaches, image registration, and the challenges associated with differences in viewing angle, image resolution, illumination, and environmental conditions. Finally, evaluation metrics and possible applications are examined. Visual localization can provide an additional source of positioning information and may complement conventional navigation systems. Its effectiveness, however, depends on the quality of reference maps, the reliability of image correspondences, and the accuracy of the geometric transformation.*

Keywords: *unmanned aerial vehicle, visual localization, image matching, satellite imagery, geographical positioning, computer vision, image registration, GNSS-denied navigation.*

1. INTRODUCTION

Unmanned aerial vehicles have become increasingly important in modern engineering, scientific research, and commercial applications. They are used for aerial photography, agricultural monitoring, infrastructure inspection, environmental observation, surveying, and search-and-rescue operations. Many of these applications require accurate knowledge of the UAV's geographical position.

The Global Navigation Satellite System is the most common source of positioning information for modern drones. By receiving signals from navigation satellites, a UAV can estimate its latitude, longitude, and altitude. This information is used by the flight controller to maintain a planned trajectory and execute autonomous missions.

Despite its widespread use, satellite-based positioning has several limitations. Buildings, mountains, and other obstacles can obstruct satellite signals. Signal reflections may introduce positioning errors, while interference can reduce the availability or reliability of navigation information. These limitations have encouraged research into alternative localization techniques, including inertial navigation, visual odometry, and image-based geolocation [1].

One of approaches is to determine the UAV's position by comparing an image captured by its onboard camera with a reference map. The reference image may be obtained from satellite imagery, aerial photographs, or a previously constructed georeferenced map. If a sufficiently accurate correspondence is established between the two images, the geographical location of the camera can be estimated.

This process is commonly referred to as visual localization or visual geolocalization. In aerial applications, it often involves matching images captured from a relatively low altitude with satellite images acquired from a much higher viewpoint.

The main objective of this paper is to explain the theoretical foundations and practical implementation of image-based UAV position estimation. The paper describes the image-matching process, presents commonly used computer vision algorithms, discusses geographical coordinate estimation, and examines the main factors influencing localization accuracy.

2. THEORETICAL FOUNDATIONS OF VISUAL LOCALIZATION

Visual localization is the process of determining the position of a camera or mobile platform by analyzing visual information. In the case of a UAV, the system

Corresponding author: R, Pavlović;
E-mail: rade.pavlovic@mod.gov.rs
DOI: 10.5937/OTEH26076

attempts to identify the geographical location corresponding to the area visible in the camera image [2].

The general principle is based on comparing two sources of visual information (figure 1):

- query image captured by the UAV's onboard camera.
- reference image with known geographical coordinates.

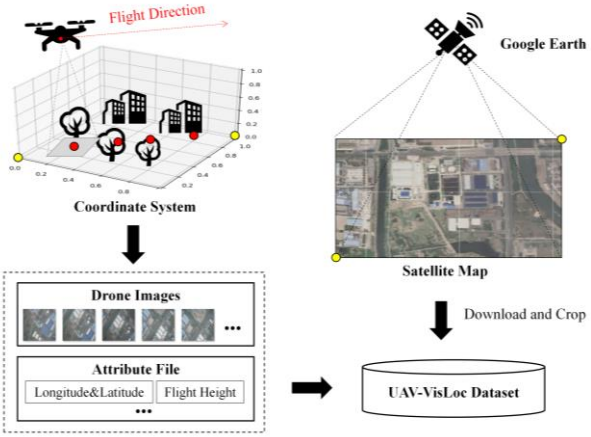


Figure 1. The dataset collecting process [3].

The query image contains visual features of the terrain, buildings, roads, fields, rivers, and other objects. The reference map contains corresponding geographical structures. By identifying common features, the system can estimate where the UAV is located.

The reference map must be georeferenced, meaning that its image coordinates are associated with real-world geographical coordinates. These coordinates may be expressed as latitude and longitude or as projected coordinates, such as easting and northing.

The localization process can be divided into two principal tasks. The first is identifying the correct geographical area represented in the UAV image. The second is estimating the precise alignment between the UAV image and the reference map [4].

The first task is often described as image retrieval or place recognition, while the second involves image matching and geometric registration. Combining these stages creates a coarse-to-fine localization pipeline.

2.1. Coordinate Systems

Accurate position estimation requires an understanding of the coordinate systems involved. A digital image uses a two-dimensional pixel coordinate system. A pixel is typically represented by coordinates ((u,v)), where (u) represents the horizontal position and (v) represents the vertical position in the image.

A map, in contrast, represents locations in a geographical coordinate system. In the WGS 84 system, a location is commonly expressed using latitude and longitude. Projected coordinate systems represent locations using linear coordinates, such as metres [5].

The purpose of image registration is to establish a mathematical relationship between these coordinate systems. For example, if a pixel in the UAV image can be matched to a known point in a georeferenced map, its corresponding geographical location can be determined. Multiple reliable correspondences can be used to estimate the transformation between the images.

However, the camera's position is not always identical to the geographical coordinates of the matched ground point. The camera is located above the Earth's surface, and its viewing direction affects the relationship between image coordinates and ground coordinates. Camera altitude, orientation, and terrain elevation must therefore be considered when estimating the UAV's actual position.

2.2. Absolute and Relative Localization

Two important localization concepts are absolute localization and relative localization. Absolute localization estimates the UAV's position within a global or geographically referenced coordinate system. Matching a camera image to a georeferenced satellite map is an example of absolute visual localization [6].

Relative localization estimates the movement of the UAV with respect to a previous position or observation. Visual odometry, for example, estimates motion by analyzing changes between consecutive camera frames.

Relative localization can provide useful motion information, but errors may accumulate over time. Absolute localization can correct this drift by connecting the current observation to a reference map.

A system that combines both approaches can use visual odometry to estimate movement between frames and map matching to periodically correct its geographical position.

3. IMAGE MATCHING ALGORITHMS

Feature-based image matching is one of the traditional approaches to visual localization. Instead of comparing every pixel directly, the algorithm identifies distinctive points in each image and establishes correspondences between them. A feature point may represent a corner, intersection, or other locally distinctive image structure. Each point is associated with a descriptor that describes its surrounding visual pattern.

The matching algorithm compares descriptors from the UAV image with descriptors from the reference map (figure 2). Similar descriptors indicate potential correspondences. Feature-based approaches are particularly useful when the images differ in scale, rotation, or moderate changes in illumination.

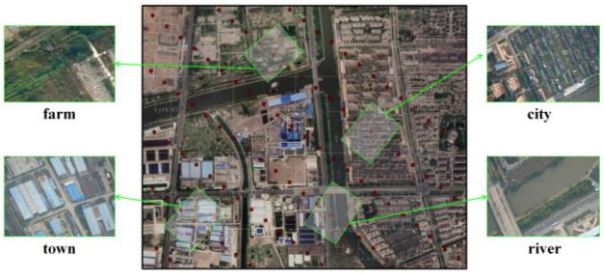


Figure 2. Drone images and satellite map from dataset [3].

Several algorithms have been developed for this purpose, including SIFT (Scale-Invariant Feature Transform), SURF (Speeded-Up Robust Features), and ORB (Oriented FAST and Rotated BRIEF). Their performance depends on image characteristics, computational resources, and the degree of difference between the two viewpoints.

3.1. Scale-Invariant Feature Transform (SIFT)

The Scale-Invariant Feature Transform is a widely used algorithm for detecting and describing local image features [7]. SIFT identifies distinctive keypoints at different image scales (figure 3). It then constructs descriptors based on the local gradient orientations surrounding each keypoint.

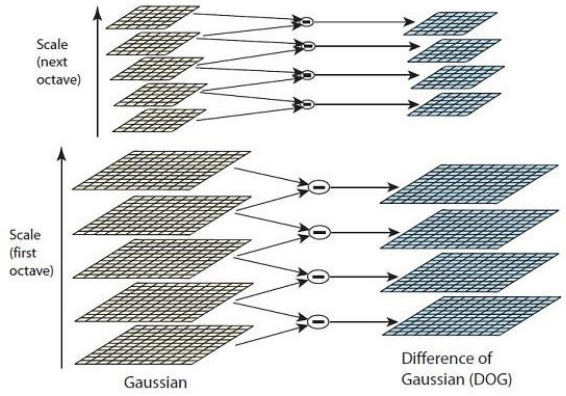


Figure 3. Diagram of SIFT process [8].

One of the principal advantages of SIFT is its relative robustness to changes in image scale and rotation. This makes it useful when a UAV image and a reference map represent the same geographical area at different resolutions. For example, a road intersection may appear relatively large in a low-altitude drone image and much smaller in a satellite image. SIFT can identify corresponding local structures even when their apparent sizes differ.

However, SIFT can require considerable computational resources when processing high-resolution images or large collections of map tiles. It may also produce unreliable matches in areas with repetitive patterns, such as similar rooftops or agricultural fields.

3.2. ORB Feature Detection

ORB combines the FAST keypoint detector with the BRIEF descriptor and an orientation estimation procedure [9]. Compared with traditional floating-point descriptors, ORB uses binary descriptors that can be matched efficiently using Hamming distance. This makes ORB attractive for real-time applications and embedded computing platforms. It can process images relatively quickly and is available in widely used computer vision libraries.

However, ORB may be less robust than some alternatives when there are substantial changes in scale, viewpoint, or image appearance (figure 4). Its effectiveness therefore depends on the particular environment and camera configuration.

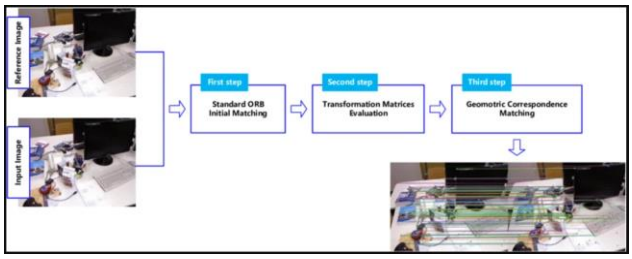


Figure 4. The framework of Oriented FAST and Rotated BRIEF [10].

For a UAV operating on a limited computational budget, ORB can provide a useful starting point. Its performance should nevertheless be evaluated against the expected range of flight altitudes and viewing angles.

3.3. Deep learning Matching

Deep learning provides an alternative approach by learning image representations directly from large datasets. Instead of relying exclusively on manually designed descriptors, a neural network can learn features that are more robust to changes between the drone image and the reference map [11]. A deep learning matcher can therefore be designed to extract high-level semantic and geometric information from both images and estimate their similarity or spatial correspondence (figure 5).

The general architecture of a deep learning-based image-to-map matching system consists of three main stages: feature extraction, feature matching, and geographical localization [12]. The first stage generates compact feature representations from the drone image and the map. The second stage establishes correspondences between the extracted representations. Finally, the matched information is transformed into an estimated geographical position.

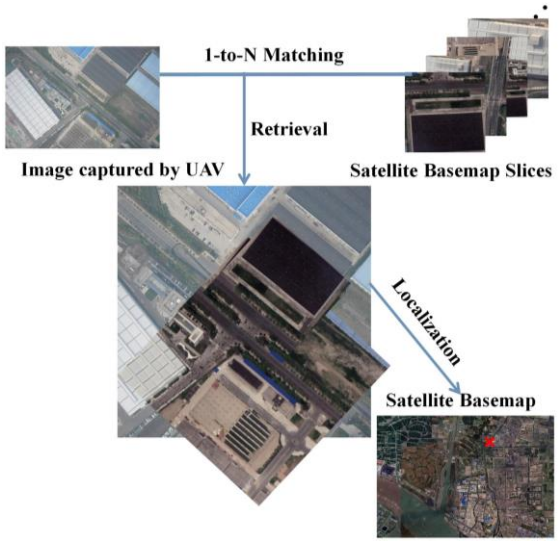


Figure 5. General architecture of a deep learning-based image-to-map matching system.

A major advantage of deep learning-based matching is its ability to learn robust representations directly from data. The learned features can be more tolerant to variations in illumination, viewpoint, scale, and scene appearance than traditional hand-crafted descriptors.

Another advantage is the possibility of combining global and local information. Global descriptors allow efficient candidate retrieval, while local matching provides accurate spatial correspondence.

However, several limitations remain. Training a reliable model requires a sufficiently large and representative dataset. Performance can decrease when the deployment environment differs significantly from the training data. Large changes in terrain, construction, vegetation, weather, or seasonal appearance may reduce the visual similarity between the drone image and the reference map.

Computational requirements are another important consideration. Deep neural networks can require significant GPU resources during both training and inference. For real-time UAV applications, model compression, quantization, efficient architectures, and hardware acceleration may therefore be necessary.

4. EXPERIMENTAL RESULTS AND POSITIONING ERROR ANALYSIS

To evaluate the effectiveness of image-based UAV localization, experiment is considered in which drone images are matched against a georeferenced satellite map. The objective is to estimate the geographical position of the UAV and measure the difference between the estimated coordinates and the known reference position. In this study dataset containing 200 UAV images is assumed. The images represent different camera orientations, flight altitudes, and environmental conditions. Each image is associated with a reference position obtained from an independent positioning system.

Three image-matching approaches are considered: SIFT, ORB, and deep learning-based feature matcher. All three methods are evaluated using the same image set and reference map. The experiment assumes that the reference map is accurately georeferenced and that the ground-truth positions are expressed in a suitable projected coordinate system, with distances measured in meters. The results of deferent measure for image-matching algorithms are shown in Table 1.

According results the deep learning-based matcher produces valid position estimates for 183 out of 200 images, corresponding to a localization success rate of 91.5%. The SIFT method produces 169 valid estimates, while ORB produces 155. Also, the mean positioning errors are 5.2 m for the deep learning-based approach, 7.5 m for SIFT, and 12.1 m for ORB. The corresponding RMSE values are 8.4 m, 11.5 m, and 19.4 m. In general, the deep learning-based matcher algorithm shows better results for all parameters**.**

Table 1. Results of comparison of image-matching algorithms.

Evaluation metric	**SIFT**	**ORB**	**Deep learning matcher**
Total test images	200	200	200
Valid position estimates	169	155	183
Localization success rate	84.5%	77.5%	91.5%
Mean error (m)	7.5	12.1	5.2
Median error (m)	4.7	7.7	3.2
RMSE (m)	11.5	19.4	8.4
90 percent error (m)	15.6	26.2	10.8

CONCLUSION

Determining a UAV's position by comparing camera images with maps is an important approach to visual localization. The method uses visual features shared by the captured image and a georeferenced reference map to establish the geographical location of the observed area.

The complete process includes image acquisition, preprocessing, feature extraction, descriptor matching, geometric verification, image registration, and coordinate transformation. Algorithms such as SIFT and ORB provide established methods for identifying and matching image features, while deep learning-based techniques offer additional possibilities for handling substantial differences in appearance and viewpoint.

The conversion from image coordinates to geographical coordinates requires an accurate reference map and a suitable geometric model. The estimated ground location must also be distinguished from the UAV's actual position, which depends on camera orientation, altitude, and terrain geometry.

Deep learning-based image-to-map matching provides a powerful framework for visual localization of drones. The proposed approach combines learned feature extraction, global image retrieval, local feature matching, attention-based correspondence estimation, and geometric verification.

The main advantage of this architecture is that it transforms the problem from direct pixel-level comparison into comparison of learned visual representations. This makes the system more robust to changes in viewpoint, scale, illumination, and image appearance. A hierarchical matching architecture is particularly suitable for large-scale applications. Global descriptors can first identify a small number of candidate

map regions, after which a local deep learning matcher performs detailed correspondence estimation. Finally, geometric verification can remove inconsistent matches and provide a reliable estimate of the drone's geographical position.

The resulting system can serve as the basis for GNSS-independent UAV localization and can be further extended by incorporating temporal information from consecutive drone frames, inertial measurements, digital elevation models, and multi-modal map information.

Although visual localization offers an alternative source of positioning information, its accuracy is affected by image quality, map resolution, environmental changes, and the reliability of feature correspondences. A robust implementation should therefore include geometric verification, performance evaluation, and methods for identifying uncertain or failed matches.

Further development of deep learning methods, multi-sensor fusion, and efficient image registration techniques may improve the reliability of visual positioning in complex environments.

Acknowledgement

This work was supported by the Ministry of Science, Technological Development and Innovation, Republic of Serbia (Contract No.451-03 34/2026-03/200325)

References

[1] COUTURIER, A., AKHLOUFI, M. A.: A review on absolute visual localization for UAV. *Robotics and Autonomous Systems, 135*, 103666 (2021). https://doi.org/10.1016/j.robot.2020.103666

[2] FISCHLER, M. A., BOLLES, R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395,(1981). https://doi.org/10.1145/358669.358692

[3] https://github.com/IntelliSensing/UAV-VisLoc

[4] HU, S., CHANG, X.: Multi-view drone-based geo-localization via style and spatial alignment. arXiv, (2020).https://arxiv.org/abs/2006.13681

[5] MOSKALENKO, I., KORNILOVA, A., FERRER, G.: Visual place recognition for aerial imagery: A survey. *Robotics and Autonomous Systems, 183*, 104837, (2025). https://doi.org/10.1016/j.robot.2024.104837

[6] *BAYKAL, F. GEDIK, M.İ. REYES-ALDASORO, C.C. KARABAĞ, C.: Image Matching for UAV Geolocation: Classical and Deep Learning Approaches*. *Journal of Imaging, 11*(11), 409, (2025). https://doi.org/10.3390/jimaging11110409

[7] LOWE, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*, 91–110, (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94

[8] https://docs.opencv.org/4.13.0/da/df5/tutorial_py_sift_intro.html

[9] RUBLEE, E., RABAUD, V., KONOLIGE, K., BRADSKI,G.: ORB: An efficient alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2564–2571, (2011). https://doi.org/10.1109/ICCV.2011.6126544

[10] https://www.researchgate.net/figure/The-framework-of-Geometric-Correspondence-Oriented-FAST-and-Rotated-BRIEF-GC-ORB_fig1_337026817

[11] FU, Y., LUO, X., QIN, X., WAN, H., CUI, J., HUANG, Z.: Deep Learning-Based Feature Matching Algorithm for Multi-Beam and Side-Scan Images. *Remote Sensing*. 2025; 17(4):675. https://doi.org/10.3390/rs17040675

[12] ZHANG, S., LI, Z., ZHANG, K., LU, Y., DENG, Y., TANG, L., MA, J.: Deep learning reforms image matching: A survey and outlook. *arXiv preprint arXiv:2506.04619*, (2025).