



Article

Vision-Based Moving-Target Geolocation Using Dual Unmanned Aerial Vehicles

Tingwei Pan ^{1,†} , Jianjun Gui ^{2,†} , Hongbin Dong ^{1,*}, Baosong Deng ² and Bingxu Zhao ¹ ¹ Department of Computer Science and Technology, Harbin Engineering University, Harbin 150009, China² Defense Innovation Institute, Chinese Academy of Military Science, Beijing 100071, China

* Correspondence: donghongbin@hrbeu.edu.cn

† These authors contributed equally to this work.

Abstract: This paper develops a framework for geolocating ground-based moving targets with images taken from dual unmanned aerial vehicles (UAVs). Unlike the usual moving-target geolocation methods that rely heavily on accurate navigation state sensors or assumptions of the known target's altitude, the proposed framework does not have the same limitations and performs geolocation of moving targets utilizing dual UAVs equipped with the low-quality navigation state sensors. Considering the Gaussian measurement errors and yaw-angle measurement bias provided by low-quality sensors, we first propose an epipolar constraint-based corresponding-point-matching method, which enables the historical measurement data to be used to estimate the current position of the moving target; after that, we propose a target altitude estimation method based on multiview geometry, which utilizes multiple images, including historical images, to estimate the altitude of the moving target; finally, considering the negative influence of yaw-angle measurement bias on the processes of target altitude estimation and parameter regression, we take advantage of multiple iterations among the two processes to accurately estimate the moving target's two-dimensional position and the yaw-angle measurement biases of two UAVs. The effectiveness and practicability of the framework proposed in this paper are proved by simulation experiments and actual flight experiments.

Keywords: altitude estimation; computer vision; corresponding point matching; moving target geolocation; unmanned aerial vehicle (UAV)



Citation: Pan, T.; Gui, J.; Dong, H.; Deng, B.; Zhao, B. Vision-Based Moving-Target Geolocation Using Dual Unmanned Aerial Vehicles. *Remote Sens.* **2023**, *15*, 389. <https://doi.org/10.3390/rs15020389>

Academic Editor: Andrzej Stateczny

Received: 1 November 2022

Revised: 4 January 2023

Accepted: 5 January 2023

Published: 8 January 2023



Copyright: © 2023 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

In recent years, with the advantages of small size, simple operation, and low price [1–3], unmanned aerial vehicles (UAVs) are increasingly being deployed for a wide variety of missions, including surveillance and reconnaissance. UAV-based computer vision capabilities, such as target detection and tracking, play particularly important roles in the above missions. However, it is not sufficient to simply detect and track a moving target; we often need to know the three-dimensional position of the moving target, which is obtained by the process of target geolocation [4–6]. In this paper, target geolocation refers to the process of using UAVs to obtain target information and estimate the position of the target in the world coordinate system. In this process, GPS (global positioning system) and AHRS (attitude and heading reference system) are used to acquire the UAV's navigation state, such as position and attitude, for geolocation in real time; the UAV's navigation state and the target's image are acquired simultaneously. There have been many studies on ground target geolocation technology [7–9]. Fabian et al. [10] analyzed the factors that affect the accuracy of moving-target geolocation. Generally speaking, a precise geolocation result depends on accurate navigation state measurement of UAVs and target's altitude. Therefore, it is difficult to obtain the accurate geolocation result of the non-cooperative target utilizing the onboard low-quality navigation state sensors. For this reason, this paper aims to achieve accurate geolocation by using only vision to mitigate the effects of Gaussian

measurement errors and the yaw-angle measurement bias, and to accurately estimate the position of the grounded moving target, which is a challenge.

In recent years, scholars have conducted a considerable amount of research on moving-target geolocation using UAV systems. In general, moving-target geolocation methods can be divided into two categories: the methods based on one-shot learning [11–14] and the methods based on multiview geometry [15–17]. As the most commonly used method, the one-shot method only obtains one image of the moving target but require the altitude of the moving target. There are many ways to get the altitude information of the moving target. Some researchers utilized a DEM (digital elevation model) to obtain the altitude of the moving target. Qiao et al. [18] proposed a vision-based geolocation frame could estimate the three-dimensional position of a moving target in real time with a gimbal-stabilized camera, but an accurate digital elevation model is required to obtain the altitude of a moving target on the ground. Alternatively, some researchers obtained the target's altitude through estimating the distance between the UAV and the moving target based on the size of the target in the image. Zhang et al. [19] proposed a relative distance estimation method based on the prior information of the target that required the size of the target to be known and then used the geometric relationship to calculate the distance between the UAV and the target. Zhu et al. [20] showed that their learning-based method could estimate the distance to a specific target. Nevertheless, their experimental results showed that the distance estimation methods based on prior information work well only if the target is close and are not suitable for accurately geolocating grounded targets with UAVs. In practical applications, the distance from the UAV to the target is usually determined by a laser rangefinder [11], which has the highest accuracy among the above-mentioned methods. However, the use of laser-based methods for the continuous geolocation of moving targets will greatly reduce the flight time of UAVs due to the weight of the laser rangefinder. It is worth noting that the above-mentioned methods all assume that the UAV's navigation state is accurate; that is, they do not consider the negative influence of the measurement errors of the UAV's navigation state on the ground moving-target geolocation.

To avoid the limits of the above-mentioned methods, some researchers utilized multiple UAVs to simultaneously acquire multiple images of a moving target and estimated its position by multiview geometry. Wang et al. [21] utilized multiple UAVs to geolocate a moving target and presented nonlinear filter based on solving the Fokker–Planck equation to address the issue of the time delay during data transmission. Xu et al. [22] proposed a method to adaptively adjust the weights according to the positions of the UAVs, which can improve the accuracy of the results of the weighted least-squares. However, these methods did not solve the problems well. On the one hand, mitigating the negative influence of Gaussian measurement errors on geolocation requires multiple images, and each UAV can only obtain one image at a time, which means that a large number of UAVs are required for accurate moving-target geolocation. On the other hand, although the above-mentioned methods using multiple measurements can mitigate the influence of Gaussian measurement errors provided by the navigation state sensors, they still have not solved the problem of yaw-angle measurement bias.

The yaw-angle bias provided by the AHRS may vary every time when the AHRS is initialized, and the assumption of unknown and constant bias is justified in a short time window [23]. Related to this, Zhang et al. [24] proposed a vision-based target geolocation method, which focuses on eliminating the effects of yaw-angle measurement bias provided by low-quality sensors on target geolocation. Further, Zhang et al. [25] focused on improving the estimation accuracy of yaw-angle bias by utilizing a particle swarm optimization algorithm to derive an expression of the optimal trajectory and perform optimal trajectory planning and then to enhance the geolocation performance. However, their methods can only be applied to the stationary-target geolocation scenarios.

In the scenario assumed in this paper, two UAVs follow a single moving target on the ground and estimate the position of the moving target in real time. Except for the images of the moving target, all the information of the target is unknown. Both UAVs

are equipped with monocular cameras to obtain target images. In addition, UAVs are equipped with GPS and AHRS to obtain their own positional and attitude information. In the navigation state measurement of UAVs, there are Gaussian measurement errors for position and attitude information, and due to different measurement principles, there are also measurement biases for yaw angle. The purpose of this paper is to mitigate the negative impact of UAV navigation state measurement errors on moving-target geolocation accuracy in the above scenario. In order to deal with this issue, we develop a ground-based moving-target geolocation framework utilizing dual UAVs, and the constructed system is shown in Figure 1. The proposed framework is free from the above-mentioned constraint of assuming the ground target is stationary, and it geolocates the moving target using sequential images taken from dual UAVs and navigation states provided by low-quality sensors.

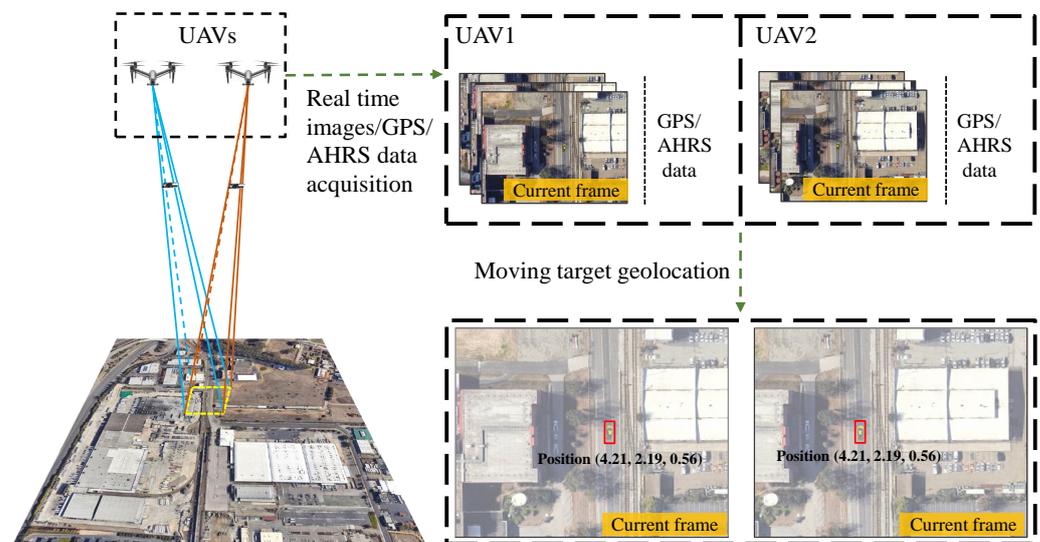


Figure 1. Demonstration that the proposed framework realizes online ground moving-target geolocation using dual UAVs.

The main contributions of this paper are as follows.

- We propose a ground moving-target geolocation framework that utilizes only the sequence of images obtained by two UAVs with low-quality sensors. The framework contains the processes of corresponding point matching, target altitude estimation, and parameter regression, which are used to mitigate the negative influences of Gaussian measurement errors and yaw-angle biases.
- In order to obtain more available images using only two UAVs, we propose a corresponding-point-matching method based on epipolar constraint. The method enables the historical images to be used to estimate the current position of the moving target. In addition, in order to filter out the wrong corresponding points, we propose an outer point filtering method based on the consistency principle.
- The effectiveness of the proposed framework was verified via the experiments in simulated and real environments.

2. Methods

The proposed ground moving-target geolocation framework based on dual UAVs is illustrated in Figure 2. The framework utilizes a multiprocessing architecture to process the four tasks of data acquisition, target detection, corresponding point matching, and moving-target geolocation in parallel. Moreover, in the moving-target geolocation, the two subtasks of target altitude estimation and parameter regression are processed iteratively.

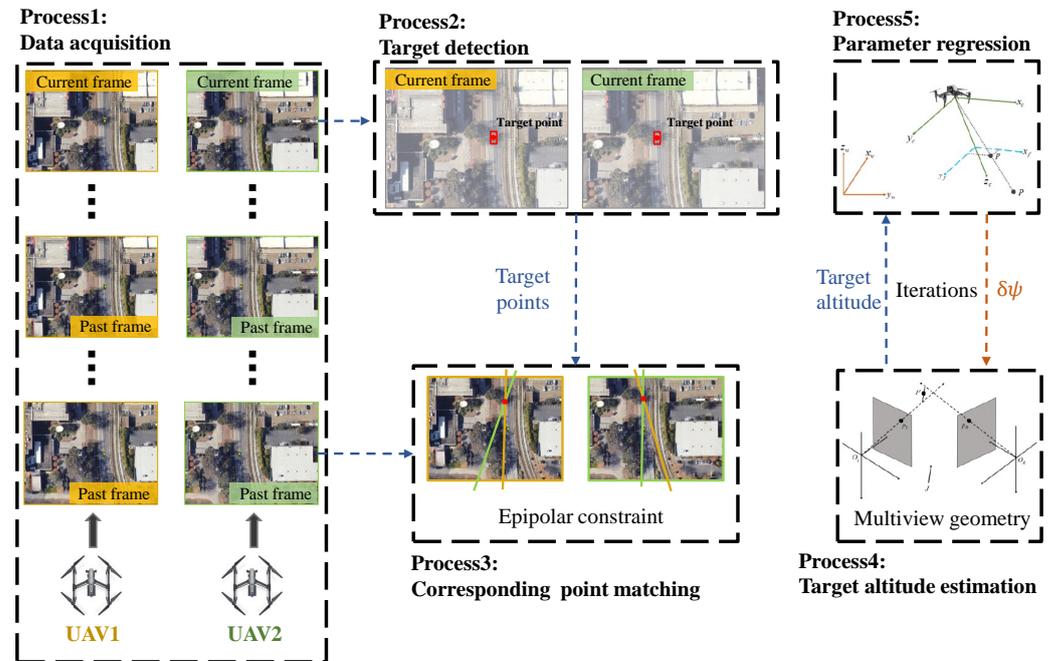


Figure 2. Overview of the proposed moving-target geolocation framework using dual UAVs.

The data acquisition process uses two UAVs equipped with a camera and low-quality navigation sensors to track the moving target and obtain the images of the target and the navigation states of two UAVs simultaneously; then, the images and navigation states are transmitted to the ground station in real time. The target detection process obtains the latest image from each UAV and achieves accurate detection of the moving target with the image pair. The corresponding point matching is to provide sufficient measurement data for the subsequent moving-target geolocation; it matches the corresponding points in the past images according to the target points provided by the target detection process, and then the past frames' images can be utilized to estimate the current position of the moving target. The process of target altitude estimation aims to estimate the target's altitude utilizing the current images and the past images. After that, the target's altitude is used by the parameter regression process to estimate the two-dimensional position of the target and the yaw-angle measurement biases of two UAVs. Considering the yaw-angle measurement biases, there are several iterations between the target altitude estimation and the parameter regression until the estimated yaw-angle biases are less than the given threshold or the number of iterations reaches the given maximum.

2.1. Parameter Regression

The parameter regression is used to estimate the two-dimensional position of target and the yaw-angle measurement biases of two UAVs. The target geolocation model based on UAV can be visualized in Figure 3.

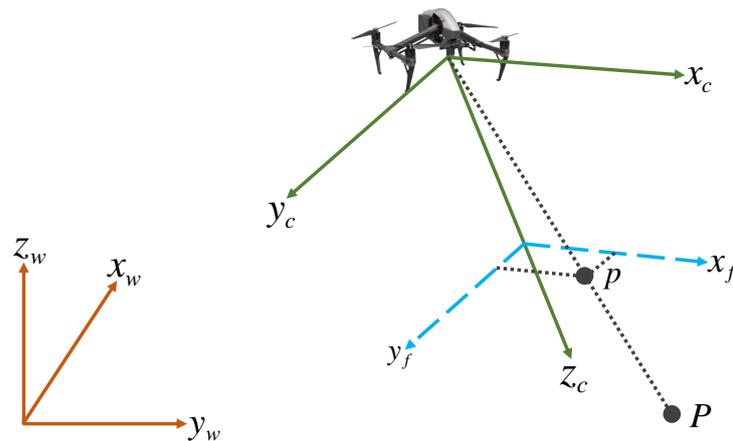


Figure 3. Target geolocation model.

According to [24], it can be expressed as

$$\begin{bmatrix} x_P \\ y_P \end{bmatrix} = \begin{bmatrix} x_v \\ y_v \end{bmatrix} + \frac{(z_P - z_v)}{(0, 0, 1)C_w^c} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} C_w^c \begin{bmatrix} x_p \\ y_p \\ f \end{bmatrix} \quad (1)$$

where $(x_P, y_P, z_P)^T$ and $(x_v, y_v, z_v)^T$ represent the coordinates of the current target’s position P and the UAV’s position in the world coordinate system w , respectively. The coordinates of projection point p of current target’s position P are represented as (x_p, y_p) in the image plane; and f is the focal length of the camera. In a simulation experiment, the precise focal lengths can be obtained directly. In a real experiment, we used the kit provided by ROS [26] to calibrate the cameras to obtain the focal lengths. The UAV’s navigation state $(x_v, y_v, z_v, \psi, \theta, \phi)$ was measured by the onboard GPS and AHRS. The rotation matrix C_w^c represents the transformation from the camera coordinate system c to the world coordinate system w . The relative altitude h between the current target’s position and the UAV is defined as

$$h = z_v - z_P \quad (2)$$

In general, the pitch angle θ and the roll angle ϕ of a UAV can be measured accurately, but the yaw-angle measurement will have large error due to the use of magnetometer. Hence, to solve the above-mentioned issue, we estimate the yaw-angle measurement biases of two UAVs and target’s position by parameter regression. The estimation process is illustrated as follows.

The goal of parameter regression is to estimate the parameter $\xi = [\xi_1, \xi_2]^T$, where

$$\xi_1 := [x_P, y_P]^T \quad (3)$$

is the two-dimensional coordinate of the moving target and

$$\xi_2 := [\delta\psi_L, \delta\psi_R]^T \quad (4)$$

are the yaw-angle measurement biases of the left UAV and right UAV, respectively. The measured navigation state of two UAVs is $\tau = [\tau_1, \tau_2, \tau_3]^T$, where

$$\tau_1 := [x_{v_L}, y_{v_L}, z_{v_L}, x_{p_L}, y_{p_L}, \theta_L, \phi_L]^T \quad (5)$$

is the navigation state of the left UAV and

$$\tau_2 := [x_{v_R}, y_{v_R}, z_{v_R}, x_{p_R}, y_{p_R}, \theta_R, \phi_R]^T \quad (6)$$

is the navigation state of the right UAV. The measured navigation state $\tau_3 := [\psi_L, \psi_R]^T$ includes the yaw angles of the left UAV and the right UAV, respectively. According to (1), the general form of the measurement equation is as follows:

$$\xi_1 = F(\tau_1, \tau_2, \tau_3) \tag{7}$$

and the actual measurements are

$$\gamma_1 = \tau_1 + w_1, w_1 \sim N(0, Q_1) \tag{8}$$

$$\gamma_2 = \tau_2 + w_2, w_2 \sim N(0, Q_2) \tag{9}$$

$$\gamma_3 = \tau_3 + \xi_2 + w_3, w_3 \sim N(0, Q_3) \tag{10}$$

where the Gaussian measurement noise covariances Q_1 and Q_2 are both 7×7 symmetric positive-definite matrices, and Q_3 is a 2×2 symmetric positive-definite matrix. Combine (8), (9) and (10) into the measurement equation as follows:

$$\xi_1 = F(\gamma_1 - w_1, \gamma_2 - w_2, \gamma_3 - (\xi_2 + w_3)) \tag{11}$$

Then, the linearized measurement equation is obtained by using Taylor’s theorem:

$$F(\gamma_1, \gamma_2, \gamma_3) \approx \xi_1 + \frac{\partial F}{\partial \tau_1} \Big|_{\gamma_1, \gamma_2, \gamma_3} \cdot w_1 + \frac{\partial F}{\partial \tau_2} \Big|_{\gamma_1, \gamma_2, \gamma_3} \cdot w_2 + \frac{\partial F}{\partial \tau_3} \Big|_{\gamma_1, \gamma_2, \gamma_3} \cdot w_3 + \frac{\partial F}{\partial \tau_3} \Big|_{\gamma_1, \gamma_2, \gamma_3} \cdot \xi_2 \tag{12}$$

Suppose that $2N$ bearing measurements of the moving target are taken by the left UAV and the right UAV at the discrete time $t = 1, \dots, N$; that is, $(\gamma_{11}, \gamma_{21}, \gamma_{31}), \dots, (\gamma_{1N}, \gamma_{2N}, \gamma_{3N})$. For $2N$ bearing measurements, the linear measurement equations are as follows:

$$\begin{pmatrix} F(\gamma_{11}, \gamma_{21}, \gamma_{31}) \\ \vdots \\ F(\gamma_{1N}, \gamma_{2N}, \gamma_{3N}) \end{pmatrix} = \begin{bmatrix} I_2 & \frac{\partial F}{\partial \tau_3} \Big|_{\gamma_{11}, \gamma_{21}, \gamma_{31}} \\ \vdots & \vdots \\ I_2 & \frac{\partial F}{\partial \tau_3} \Big|_{\gamma_{1N}, \gamma_{2N}, \gamma_{3N}} \end{bmatrix} \xi + W \tag{13}$$

where

$$W \sim N(0, Q) \tag{14}$$

is the geolocation error caused by the Gaussian measurement error of the UAV’s navigation state, and its covariance matrix is as follows:

$$Q = \text{diag} \left(\left\{ \left(\frac{\partial F}{\partial \tau_1} \Big|_{\gamma_{1k}, \gamma_{2k}, \gamma_{3k}} \right) Q_1 \left(\frac{\partial F}{\partial \tau_1} \Big|_{\gamma_{1k}, \gamma_{2k}, \gamma_{3k}} \right)^T + \left(\frac{\partial F}{\partial \tau_2} \Big|_{\gamma_{1k}, \gamma_{2k}, \gamma_{3k}} \right) Q_2 \left(\frac{\partial F}{\partial \tau_2} \Big|_{\gamma_{1k}, \gamma_{2k}, \gamma_{3k}} \right)^T + \left(\frac{\partial F}{\partial \tau_3} \Big|_{\gamma_{1k}, \gamma_{2k}, \gamma_{3k}} \right) Q_3 \left(\frac{\partial F}{\partial \tau_3} \Big|_{\gamma_{1k}, \gamma_{2k}, \gamma_{3k}} \right)^T \right\}_{k=1}^N \right) \tag{15}$$

From (13), the two-dimensional position of moving target and the yaw-angle biases of two UAVs can be solved effectively with the weighted least-square method.

2.2. Estimation of Target Altitude

It can be seen in (1) that the target’s altitude z_P is required by the process of parameter regression. In this section, the target’s altitude z_P is obtained by utilizing the multiview geometry.

The left UAV and the right UAV continuously acquire the navigation state and the target images when tracking the moving target. We select the current image and $N - 1$ past images from the sequence of images taken by each UAV as the measurement data to estimate the current position of the moving target, and the two adjacent images in these N images must meet the baseline constraint—that is, the distance between the positions of UAV when taking these two images must be greater than or equal to the given threshold. Any image $g_{L_i}(i = 1, \dots, N)$ taken by the left UAV and any image $g_{R_j}(j = 1, \dots, N)$ taken by the right UAV can be form an image pair m_{ij} , and a total of N^2 image pairs can be obtained. For each image pair m_{ij} , its multiview geometry model is shown in Figure 4. The points p_L and p_R are the corresponding points of the target’s position P projected in the left image, g_{L_i} , and the right image, g_{R_j} , respectively. The corresponding points in the current images are provided by target detection process, but the corresponding points in the past images are provided by the process of corresponding point matching described in Section 2.3. O_L and O_R are the origins of the left and right camera coordinate systems, respectively. According to the principle of multiview geometry, if there is no measurement error, the lines $O_L p_L$ and $O_R p_R$ will intersect at point P . Nevertheless, due to the Gaussian measurement error and yaw-angle biases provided by the low-quality sensors, the lines will not intersect at the point P . Therefore, we approximate the intersection point P to the nearest point P' from the two lines.

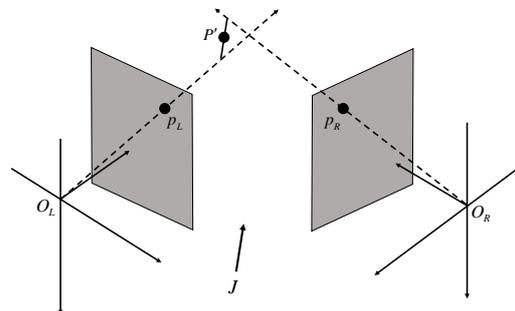


Figure 4. Target altitude estimation model.

Let $a p_L (a \in R)$, $T_{RL} + b R_{RL}^T p_R (b \in R)$ be ray $O_L p_L$ and ray $O_R p_R$ in the left coordinate system, respectively. We denote J as an orthogonal vector with the rays $O_L p_L$ and $O_R p_R$. The problem is now simplified to searching the midpoint P' of the line segment parallel to J that joins $O_L p_L$ and $O_R p_R$.

The endpoints of the line segment ($a_0 p_L$ and $T_{RL} + b_0 R_{RL}^T p_R$) are calculated to solve the following equation:

$$\begin{bmatrix} p_L & -R_{RL}^T p_R & p_L \times R_{RL}^T p_R \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = T_{RL} \tag{16}$$

for a_0 , b_0 , and c_0 . We let h_{ij} be the relative altitude between the target and the UAV when taking the image g_{L_i} , and h_{ij} is estimated utilized image g_{L_i} and g_{R_j} . Similarly, h_{ji} is the relative altitude between the target and the UAV when taking the image g_{R_j} , and h_{ji} is estimated utilizing image g_{R_j} and g_{L_i} . Thus, the relative altitude h_{ij} and h_{ji} can be obtained from the following equations:

$$h_{ij} = \frac{1}{2} [0 \ 0 \ 1] R_L (a_0 p_L + b_0 R_{RL} p_R + T_{RL}) \tag{17}$$

$$h_{ji} = \frac{1}{2} [0 \quad 0 \quad 1] R_R (b_0 p_R + a_0 R_{LR} p_L + T_{LR}) \tag{18}$$

where $p_L := [x_{p_L}, y_{p_L}, f_L]^T$ and $p_R := [x_{p_R}, y_{p_R}, f_R]^T$ are the coordinates of the corresponding points in the left and right coordinate systems, respectively. The matrices R_{RL} and T_{RL} are the rotation matrix and displacement matrix from the right coordinate system to the left coordinate system, respectively. Similarly, the matrices R_{LR} and T_{LR} are the rotation matrix and displacement matrix from the left coordinate system to the right coordinate system respectively. That is,

$$R_L := C_c^n(\psi_i + \delta\psi_{L_n}, \theta_i, \phi_i), R_R := C_c^n(\psi_j + \delta\psi_{R_n}, \theta_j, \phi_j) \tag{19}$$

where $\delta\psi_{L_n}$ and $\delta\psi_{R_n}$ is the estimated value given by parameter regression in the n_{th} iteration and the initial values $\delta\psi_{L_0} = \delta\psi_{R_0} = 0$. There are N image pairs containing image g_{L_i} or image g_{R_j} , so we can get N estimates of relative altitude for each image g_{L_i} or image g_{R_j} . Thus, we can get the average altitude \bar{h}_i and \bar{h}_j for image g_{L_i} and g_{R_j} by the following equations, respectively.

$$\bar{h}_i = \frac{\sum_{j=1}^N h_{ij}}{N}, \bar{h}_j = \frac{\sum_{i=1}^N h_{ji}}{N} \tag{20}$$

After that, we estimate the target’s altitude according to the indirect adjustment theory. After the estimation of relative altitude, each image $g_k (k = 1, \dots, 2N)$ has an estimated relative altitude \bar{h}_k . There is estimation error for relative altitude \bar{h}_k and measurement error for UAV’s altitude z_{v_k} . According to (2), we have

$$z_{v_k} + e_{v_k} = z_P + \bar{h}_k + e_{h_k} \tag{21}$$

where e_{v_k} is the measurement error of z_{v_k} and e_{h_k} is the estimation error of \bar{h}_k . For $2N$ images, we use E_k to represent $e_{v_k} - e_{h_k}$ and have

$$\begin{pmatrix} E_1 \\ \vdots \\ E_{2N} \end{pmatrix} = J_{2N,1} z_P - \begin{bmatrix} z_{v_1} - \bar{h}_1 \\ \vdots \\ z_{v_{2N}} - \bar{h}_{2N} \end{bmatrix} \tag{22}$$

where $J_{2N,1}$ represents the $2N \times 1$ matrix of ones. According to the least-squares theory, z_P can be obtained as follows:

$$z_P = (J_{2N,1}^T J_{2N,1})^{-1} J_{2N,1}^T \begin{bmatrix} z_{v_1} - \bar{h}_1 \\ \vdots \\ z_{v_{2N}} - \bar{h}_{2N} \end{bmatrix} \tag{23}$$

Finally, the estimated altitude of the moving target is used to estimate the yaw-angle biases and the two-dimensional positions described in Section 2.1.

2.3. Corresponding Point Matching

In this paper, the corresponding points refer to the projection points of the target’s current position P on different images. The target points provided by target detection in the current images g_{L_1} and g_{R_1} captured by the left and right UAVs, respectively, are corresponding points of the moving target’s current position, but the target points provided by target detection in the past images cannot be regarded as corresponding points because the target’s position has changed. Therefore, it is necessary to match the corresponding points before using the past images to estimate the current position of the moving target.

In this section, we first propose a corresponding-point-matching method for the current images and the past images based on an epipolar constraint. As mentioned above,

2N images are required for each estimation of target position, of which N images are from the left camera and the other N images are from the right camera. Among the 2N images, there are two current images, denoted as g_{L1} and g_{R1} , and the remaining $2N - 2$ images are the past images. For each matching of the corresponding point, the current images g_{L1} and g_{R1} and any past image $g_n (n = 1, \dots, 2N - 2)$ are used to build a corresponding-point-matching model, as shown in Figure 5. The points $p_L := [x_{p_L}, y_{p_L}]^T$ and $p_R := [x_{p_R}, y_{p_R}]^T$ are the corresponding points provided by target detection in the current images g_{L1} and g_{R1} captured by the left UAV and right UAV. The point $p_n := [x_p, y_p]^T$ is the corresponding point provided by the corresponding point matching in the past image g_n . $O_L, O_R,$ and O_n are the optical centers of cameras. The points $e_1, e_L,$ and e_2, e_R are the intersection points of the lines $O_L O_n$ and $O_R O_n$ with the image plane, respectively.

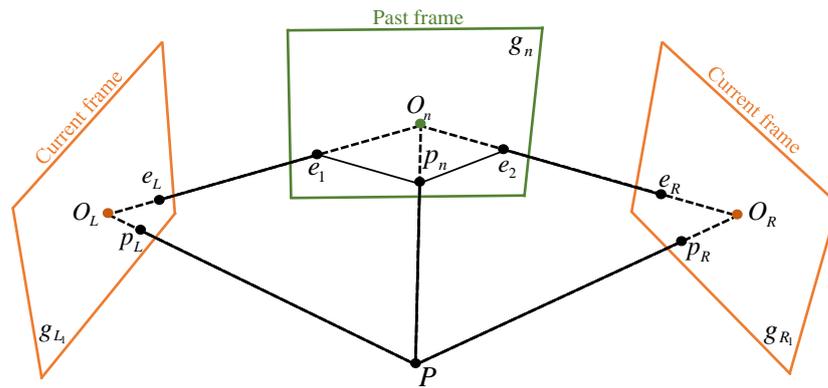


Figure 5. Corresponding-point-matching model.

The coordinate of target point p_L in the camera coordinate system O_n is represented as

$$p'_L = R_{Ln} \left(K_L^{-1} [x_{p_L}, y_{p_L}, 1]^T \right) + T_{Ln} \tag{24}$$

where the matrix K_L is the intrinsic matrix of the left camera. The matrices R_{Ln} and T_{Ln} are the rotation matrix and translation matrix from coordinate system O_L to coordinate system O_n , respectively. The coordinate of target point p_R in the camera coordinate system O_n is represented as

$$p'_R = R_{Rn} \left(K_R^{-1} [x_{p_R}, y_{p_R}, 1]^T \right) + T_{Rn} \tag{25}$$

where the matrix K_R is the intrinsic matrix of the right camera. The intrinsic matrix K reflects the attributes of the camera, and its description is as follows:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{26}$$

where f_x represents the ratio of the physical focal length to the physical size of the pixel in the x direction, and f_y represents the ratio of the physical focal length to the physical size of the pixel in the y direction. For the cameras used in simulation experiments and real experiments, there is $f_x = f_y = f$. The coordinate of the principal point of photograph is denoted as (c_x, c_y) . The intrinsic matrix K is obtained by camera calibration, and the intrinsic matrices of the left camera and the right camera are denoted as K_L and K_R , respectively. The matrices R_{Rn} and T_{Rn} are the rotation matrix and translation matrix from coordinate system O_R to coordinate system O_n , respectively. In the corresponding-point-matching model, the three-dimensional point P may be at any position on the line $O_L P$ because its depth information is unknown. However, no matter where the three-

dimensional point P is on the line $O_L P$, its projection point p_n on image g_n must be on the line $e_1 p_n$ according to the imaging principle. The line $e_1 p_n$ can be described as

$$[T_{Ln} \times p'_L] K_n^{-1} [x_p, y_p, 1]^T = 0 \tag{27}$$

where $K_n^{-1} [x_p, y_p, 1]^T$ is the coordinate of point p_n in the camera coordinate system O_n . Similarly, the line $e_2 p_n$ can also be determined on the image g_n according to the target point p_R , which can be described as

$$[T_{Rn} \times p'_R] K_n^{-1} [x_p, y_p, 1]^T = 0 \tag{28}$$

Therefore, the coordinates of corresponding point p_n on the image g_n can be solved by combining (27) and (28). In other words, the corresponding point we are looking for is the intersection of line $e_1 p_n$ and line $e_2 p_n$. In the above process, the rotation and translation matrices between cameras are necessary. For this, the corresponding point matching subtask first extracts ORB features from the current images, which is an effective and fast-computing feature point. Then, the subtask matches the feature points extracted from these two current images with the feature points extracted from the past image. After that, the eight-point algorithm with random sample consensus (RANSAC) is used to improve the accuracy of feature point matching and estimate the fundamental matrices F_{Ln} and F_{Rn} . The rotation and translation matrices between cameras can be obtained by decomposing the fundamental matrices F_{Ln} and F_{Rn} .

However, the error matching of the feature points extracted from the current images and the past image will reduce the accuracy of the fundamental matrix estimation and then reduce the accuracy of the corresponding point matching. In this regard, we propose a method based on the consistency principle to filter the wrong corresponding points. The subtask of corresponding point matching can obtain $2N - 2$ corresponding points because there are $2N - 2$ past images selected to be used to estimate the current position of the moving target. Then, the two current images and each of the $2N - 2$ past image form multiple groups of observation data—that is, $(g_{L1}, g_{R1}, g_1), \dots, (g_{L1}, g_{R1}, g_{2N-2})$. Each group is sent to the iterations of target altitude estimation and parameter regression. After that, we can get $2N - 2$ estimates of yaw-angle biases—that is, $\zeta_{2_1}, \dots, \zeta_{2_{2N-2}}$. For each estimate $\zeta_{2_n} (n = 1, \dots, 2N - 2)$, we calculate its consistency factor C_n according to the following equation:

$$C_n = \frac{\sum_{m=1; m \neq n}^{2N-2} \mathbb{I}(e_{(n,m)})}{2N - 2} \tag{29}$$

$\mathbb{I}(\cdot)$ is an indicator function defined by

$$\mathbb{I}(e_{(n,m)}) = \begin{cases} 1, & \text{if } e_{(n,m)} < TH \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

where e_{nm} is defined by

$$e_{nm} = abs(\delta\psi_L^n - \delta\psi_L^m) + abs(\delta\psi_R^n - \delta\psi_R^m) \tag{31}$$

After that, we have $C_{max} = MAX(C_1, \dots, C_{2N-2})$ and its corresponding estimates $\zeta_{2_n} = [\delta\psi_L^n, \delta\psi_R^n]^T$. The set S_p of corresponding points to be filtered is obtained; that is,

$$S_p = \{p_m | e_{nm} \geq TH\} \tag{32}$$

The corresponding points p_m in set S_p are considered as incorrect corresponding points, and their corresponding image g_m will no longer be used to estimate the position for the moving target. It should be noted that if $N = 2$ and $e_{nm} \geq TH$, the two past images

are eliminated and the geolocation is terminated because there is not enough measurement data.

In this section, a corresponding-point-matching method based on epipolar constraint is proposed. Considering the existence of a wrong corresponding point, we also propose a method based on the consistency principle to eliminate the wrong corresponding points. After corresponding point matching, multiple past images can be used to estimate the current position of the moving target.

3. Results

In this section, we evaluate the performance of the proposed geolocation framework for a moving target. In order to conduct the simulation experiments, we built a moving-target geolocation platform based on ROS and GAZEBO [26] to evaluate the robustness of the proposed framework by various simulations. In addition, we used two quad-rotor UAVs to further verify the effectiveness of the framework in a real experiment.

3.1. Simulation Experiment

For simulation experiments, a moving target detection and tracking platform in an urban environment was built. In the platform, a vehicle runs in a city environment, as shown in Figure 6, and two UAVs with vertically downward cameras detect the target utilizing Yolov5 [27], as shown in Figure 7. The flight altitude of the two UAVs is 100 m, and the baseline between the two UAVs is 70 m. In the process of target sensing, the UAVs first use the target detection algorithm to find the target, and then use the proposed method to continuously geolocate the moving target. We implemented the proposed framework using Python and ran the framework on a machine with Intel(R) an i7-10700 @2.90 GHz CPU and NVIDIA RTX 2070 Super GPU.



Figure 6. Urban environment and moving path of the target in the GAZEBO simulation.

Before geolocating the moving target using multiple images, including the past images, the corresponding points in the past images were matched using the corresponding-point-matching method described in Section 2.3. The accuracy of the corresponding-point-matching method proposed in this paper depends heavily on the estimation of the fundamental matrix F , and wrong corresponding point matching will cause an error in target geolocation. However, the estimation of the fundamental matrix is occasionally inaccurate due to the matching error of ORB features between the current image and the past image. Therefore, the method of filtering the wrong corresponding points is proposed in Section 2.3.

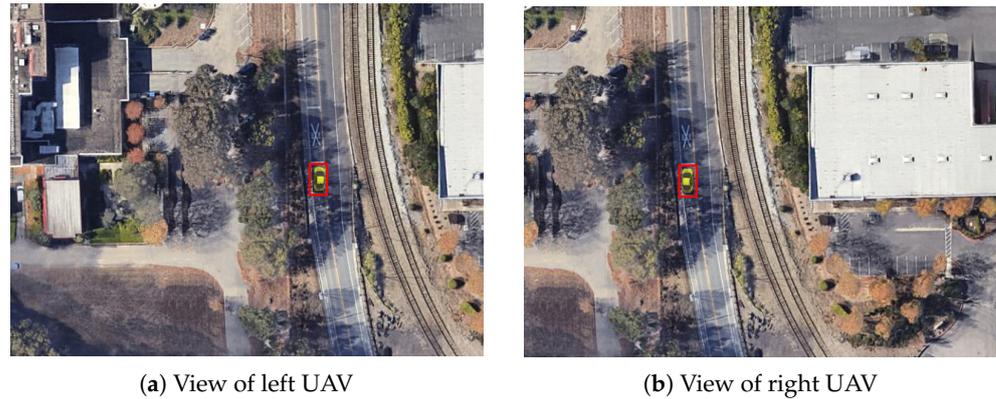


Figure 7. Two UAVs track and detect the moving target.

In order to evaluate the performance of the method for corresponding point matching and the effectiveness of the method for filtering the wrong corresponding points, we obtained the experimental results shown in Figure 8. In this experiment, we did not consider the influence of UAV navigation state measurement error on the geolocation results, but only observed the changes caused by the corresponding points. It is worth noting that eight images were used for each geolocation in this experiment. In Figure 8, the black line is the ground-truth path of the moving target, and the yellow line is the path estimated by the proposed moving-target geolocation framework without filtering the wrong corresponding points. It can be seen that the corresponding point matching introduces errors to the moving-target geolocation, and the geolocation error is large when the car is at the corner of an image, which is due to the small overlap between the current image and the past images, so there are not enough ORB features to estimate an accurate fundamental matrix. However, the geolocation error was greatly reduced after using the proposed method of filtering the wrong matching points, as shown by the red line.

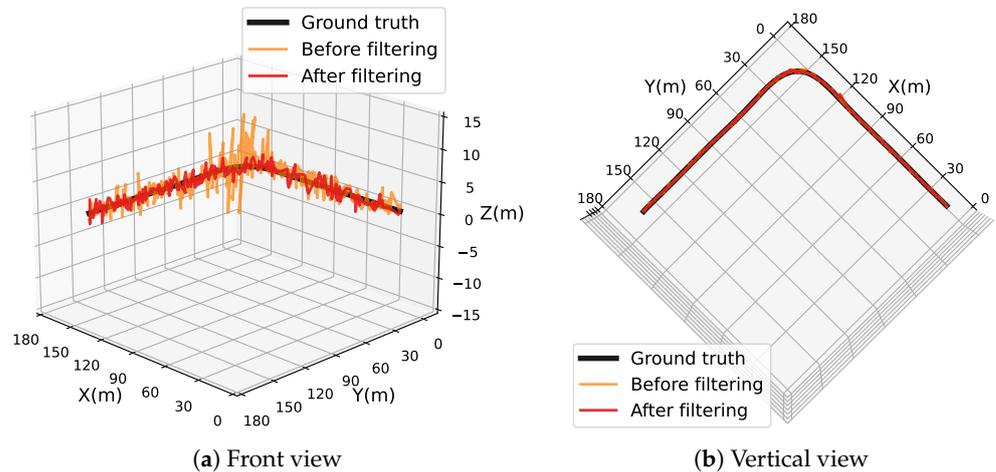


Figure 8. Paths of geolocation in the experiment of corresponding point matching.

The statistical characteristics of continuous geolocation results are shown in Table 1, in which we count the mean absolute errors (MAEs), the standard deviations (STDs), and the maximum errors (MAX) of geolocation results in each direction, X, Y, and Z. The mean absolute error (MAE) is calculated as follows:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (33)$$

where n represents the number of geolocations, and y_i and \hat{y}_i represent the estimated value and the true value, respectively. It can be seen in Table 1 that the proposed method can effectively filter out the wrong corresponding points and reduce the geolocation errors.

Table 1. Statistical results of geolocation in the experiment of corresponding point matching.

	Before Filtering			After Filtering		
	x	y	z	x	y	z
MAE (m)	0.283	0.274	1.216	0.151	0.173	0.720
STD (m)	0.373	0.408	1.561	0.125	0.164	0.547
MAX (m)	4.135	6.335	8.574	0.692	1.135	2.464

After matching corresponding points, the past images and the current images are used to estimate the target's altitude z_p . We obtained the statistical characteristics of altitude estimation in the continuous geolocation process, as shown in Table 2. We utilized 4 or 8 images to obtain the altitude estimation results, and compared the mean absolute errors, the standard deviations, and the maximum errors of altitude estimation errors when using different numbers of images. In addition, we set $\sigma_\psi = \sigma_\theta = \sigma_\phi = 2^\circ$ and $\sigma_{x_v} = \sigma_{y_v} = \sigma_{z_v} = 3m$. The yaw-angle biases $\delta\psi_L, \delta\psi_R$ of the left and right UAVs were set as follows:

$$\delta\psi_L = 0.2t + \delta\psi_0; \delta\psi_R = 0.2t - \delta\psi_0 \quad (34)$$

where $t(s)$ represents the flight time of UAVs and $\delta\psi_0$ was set to 20 in this experiment. It can be seen from the results shown in Table 2 that the more images used for target altitude estimation, the more accurate the altitude estimation will be. However, using more images means more computing costs, and the average processing speed of moving-target geolocation was 22 FPS when utilizing eight images and the above-mentioned machine in the realistic experiment.

Table 2. Statistical results of altitude estimation of the target.

Number of Images	MAE (m)	STD (m)	MAX (m)
4	3.372	2.769	9.812
8	2.448	1.931	7.819

Similarly to the experiments of target altitude estimation, we evaluated the performance of yaw-angle bias estimation. In this experiment, the true values of yaw-angle biases need to be known in order to evaluate our method, but it is difficult to measure the yaw-angle biases of a sensor in actual flight. Therefore, we set the same experimental conditions as the altitude estimation experiment to evaluate the performance of yaw-angle biases' estimation in the simulation environment. As shown in Table 3, the same conclusion can be obtained—that is, the proposed method can effectively estimate the yaw-angle biases of UAVs, and the more images used, the more reliable the estimation results will be.

Table 3. Statistical results of yaw-angle bias estimation.

Number of Images		MAE (deg)	STD (deg)	MAX (deg)
4	$\delta\psi_L$	7.642	4.270	13.943
	$\delta\psi_R$	6.633	3.477	13.125
8	$\delta\psi_L$	3.051	2.781	7.099
	$\delta\psi_R$	2.550	2.156	6.957

In the proposed framework, the altitude of a moving target and the yaw-angle biases of two UAVs are estimated iteratively until the preset conditions are reached. We preset two conditions for stopping the iteration. One is that the estimated yaw-angle biases of the current iteration are less than 0.1, and the other is that the maximum number of iterations is reached, and we set the maximum number of iterations to 20 in this study. Table 4 shows the average absolute errors of the estimated values in the iteration process of the proposed framework utilizing eight images under the above-mentioned experimental settings. It can be seen that when the number of iterations is four, the estimated yaw-angle biases and target altitude reach stable values. After that, the results of subsequent iterations fluctuate around the stable values. This is because in the iteration process, there is not only the influence of yaw-angle biases, but also the influence of Gaussian measurement errors. This makes it impossible for the whole system to reach a minimum yaw-angle biases through multiple iterations.

Table 4. Iterative process of parameter estimation.

Number of Iterations k	1	2	3	4	8	12	16	20
$\delta\psi_{Lk}$ (deg)	16.372	4.269	2.812	2.753	2.841	2.912	2.736	3.051
$\delta\psi_{Rk}$ (deg)	14.448	4.931	2.469	2.817	2.526	2.728	2.837	2.550
z_P (m)	10.448	5.931	3.819	2.769	2.815	2.635	2.864	2.448

The main factors that affect the geolocation accuracy are the measurement errors of the UAV's navigation state. Therefore, we evaluated the performance of the proposed framework under different measurement errors of the UAV's attitude angle and position. As the comparison algorithm, the one-shot method assumes that the distance from UAV to the moving target is known (e.g., provided by laser rangefinder) and utilizes only one image to estimate the three-dimensional coordinates of the moving target. This approach assumes that the UAV navigation state is provided by an accurate AHRS, so it does not consider the influence of the UAV navigation state measurement errors on the geolocation result. We first assume that only the attitude angle has Gaussian measurement error in the UAV's navigation state, and then compares the performance of the proposed framework and the one-shot method. Specifically, we set $\sigma_\psi = 1^\circ, \dots, 5^\circ$, respectively, and $\sigma_\psi = \sigma_\theta = \sigma_\phi$. In the comparisons, we compared the three-dimensional geolocation errors and the distance error, as shown in Figure 9. It should be noted that the distance error is the distance between the estimated three-dimensional position and the true three-dimensional position.

The proposed framework utilizes 4 or 8 images to obtain the geolocation results for each geolocation. It can be seen that with an increase in attitude-angle measurement errors, the geolocation errors obtained by the two methods increase. In terms of the x coordinate and y coordinate, the geolocation errors of our method are always smaller than those of the one-shot method. This is because the proposed framework can expand the measurement data by corresponding point matching, and more measurement data can mitigate the negative influence of Gaussian measurement errors. In the same way, for the proposed framework, the geolocation results obtained by using eight images are more stable than those obtained by using four images. However, in terms of z coordinate, the geolocation error of our method is larger than that of the one-shot method. There are two reasons for this phenomenon. On the one hand, the target's altitude z_P is calculated according to the distance between the UAV and the moving target, and the true distance is assumed to be known for the one-shot method. On the other hand, the altitude estimation method described in Section 2.2 regards the point closest to the two lines of sight as the target point. The geolocation error caused by attitude measurement errors is more concentrated on the target altitude because the horizontal distance between the UAV and the target is much smaller than the vertical distance. Finally, the distance error of our method with eight

images is smaller than that of the one-shot method, because the distance error contains the three-dimensional errors.

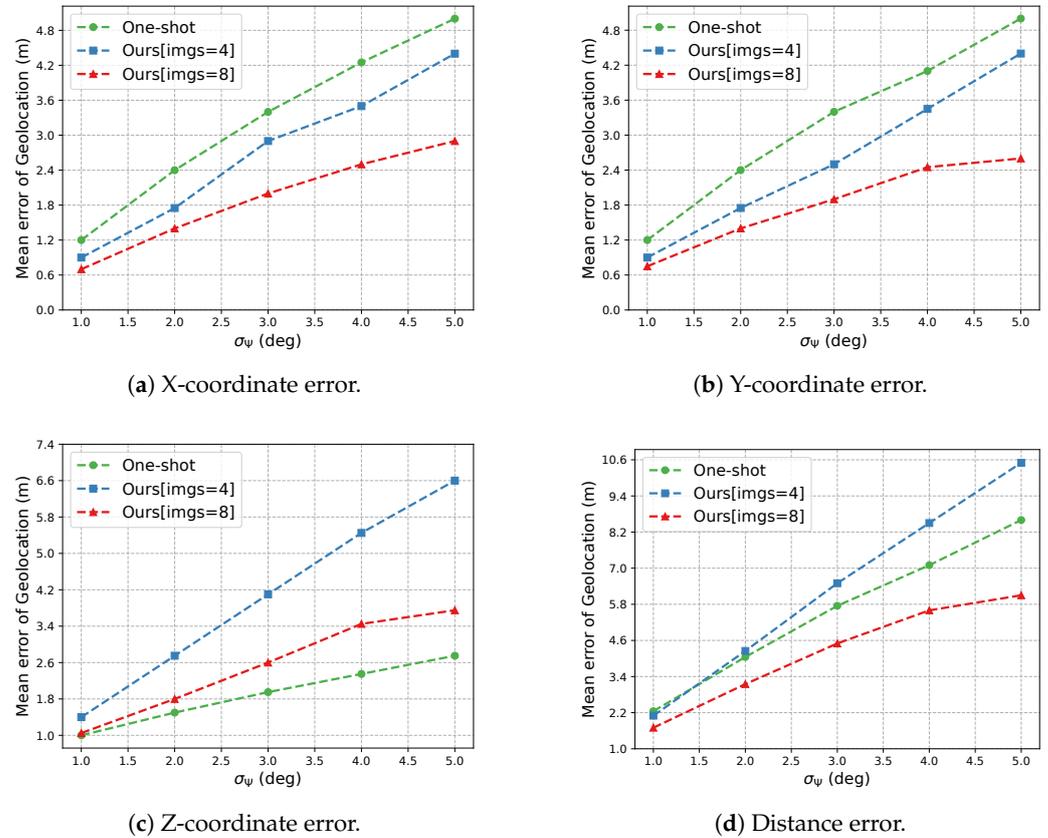


Figure 9. Geolocation errors under different measurement errors of attitude.

Then, we assumed that only the position has Gaussian measurement error in the UAV's navigation state, and then compared the performance of the proposed framework and the one-shot method. Specifically, we set $\sigma_z = 1m, \dots, 5m$, respectively, and $\sigma_x = \sigma_y = \sigma_z$. In the comparisons, we compare the three-dimensional geolocation errors and the distance error, as shown in Figure 10. The same conclusion can be obtained; that is, our method can greatly mitigate the impact of Gaussian measurement errors on the geolocation results due to the using of the historical measurement data. The difference is that our method with eight images is better than the one-shot method in terms of z coordinate. There are two reasons for this phenomenon. On the one hand, the influence of UAV's position measurement error is greater than that of attitude measurement error for the one-shot method. On the other hand, the proposed framework can mitigate the negative influence of Gaussian measurement error by using weighted least-squares. In terms of distance error, our method is superior to the one-shot method, no matter whether four images or eight images are used.

After that, we assume that both the position and the attitude angle have Gaussian measurement error in the UAV's navigation state, and then compare the performance of the proposed framework and the one-shot method. Specifically, we set $\sigma_z = 1m, \dots, 5m$, $\sigma_\psi = 1^\circ, \dots, 5^\circ$, respectively, and $\sigma_x = \sigma_y = \sigma_z$, $\sigma_\psi = \sigma_\theta = \sigma_\phi$. It can be seen in Figure 11 that our method with eight images has better mitigation effects on the negative influence of various Gaussian measurement errors. In this experiment, the assumption of measurement errors is consistent with most application scenarios. Even if the yaw-angle bias is not considered, the proposed framework is better than the one-shot method in practical applications.

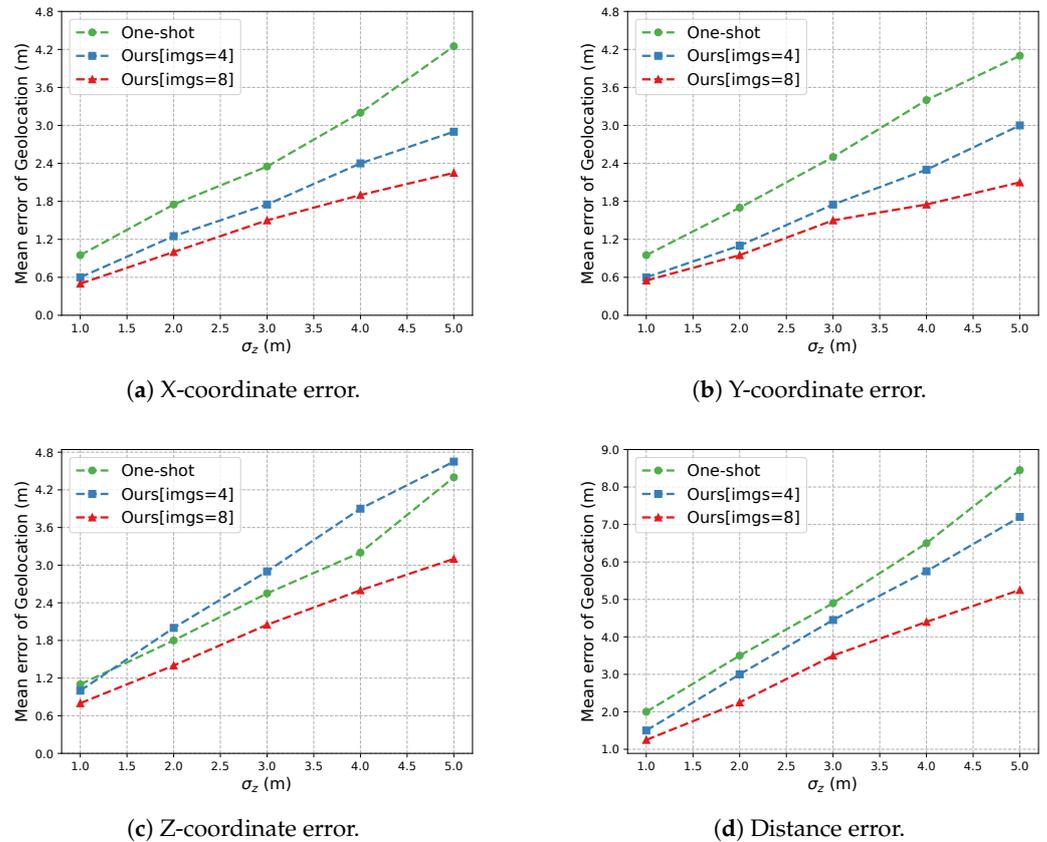


Figure 10. Geolocation errors under different measurement errors of position.

In addition to the Gaussian measurement errors, the yaw-angle biases provided by the low-quality sensors also affect the geolocation results. Therefore, we assume that the navigation state has both Gaussian measurement errors and yaw-angle biases, and compare the performances of the proposed method and the one-shot method at different yaw-angle biases. In detail, we set $\sigma_\psi = \sigma_\theta = \sigma_\phi = 2^\circ$ and $\sigma_x = \sigma_y = \sigma_z = 3m$. The yaw-angle biases $\delta\psi_L, \delta\psi_R$ of the left and right UAVs are set as (34). We set $\delta\psi_0$ to 10, 15, 20, 25, and 30, respectively, and then compared the performances of the two methods. It is shown in Figure 12 that the geolocation accuracy of the one-shot method is greatly reduced when there is bias in yaw-angle measurement, and as the yaw-angle measurement bias increases, the geolocation accuracy of the one-shot method becomes lower and lower. Compared with the one-shot method, the accuracy of our method does not change significantly as the yaw-angle measurement bias increases. This is because our method can estimate the yaw-angle bias through the iterations between the estimation of the target's altitude and the parameter regression to avoid the negative impact of the yaw-angle measurement bias on geolocation accuracy. However, in terms of z-coordinate error, the accuracy of both methods has no obvious change as yaw-angle bias increases. There are two reasons for this phenomenon. On the one hand, the one-shot method assumes that the distance between the UAV and the moving target is known and the yaw-angle measurement error has little impact on the target's altitude estimation. On the other hand, the proposed framework can eliminate the negative influence of the yaw-angle biases by utilizing the iterations between the processes of altitude estimation and parameter regression.

In conclusion, we first verified the effectiveness of the proposed corresponding-point-matching method, target altitude estimation method, and yaw-angle bias estimation method in the simulation environment. In addition, we conducted a lot of simulation experiments to verify the effectiveness and robustness of the proposed framework, and the simulation experiments showed that the geolocation results of the proposed framework are more accurate and stable under Gaussian measurement errors and yaw-angle biases compared with the commonly used one-shot method.

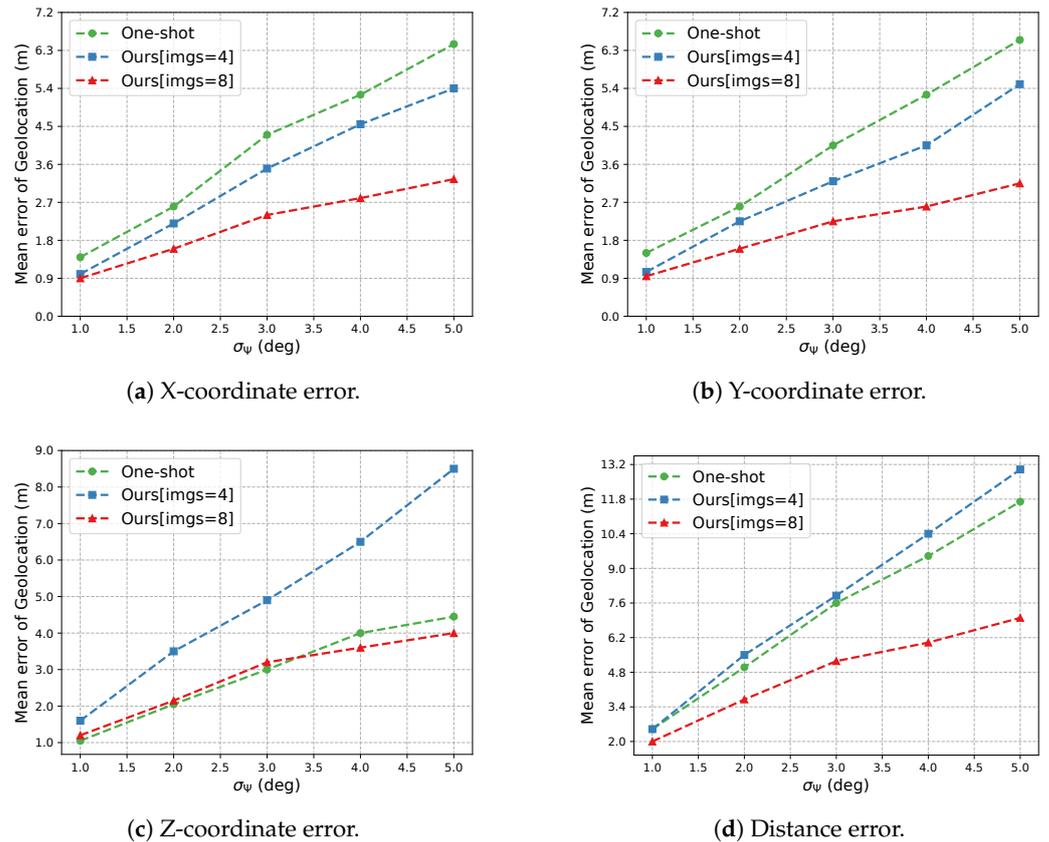


Figure 11. Geolocation errors under different measurement errors of attitude and position.

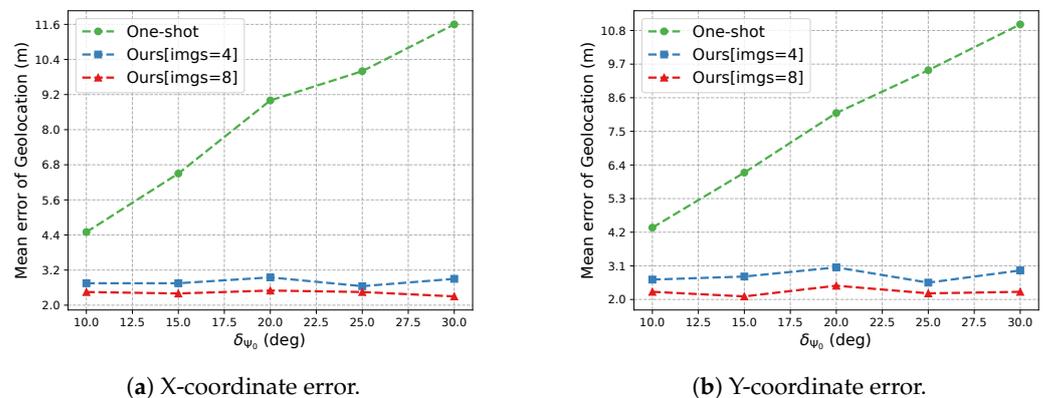


Figure 12. Cont.

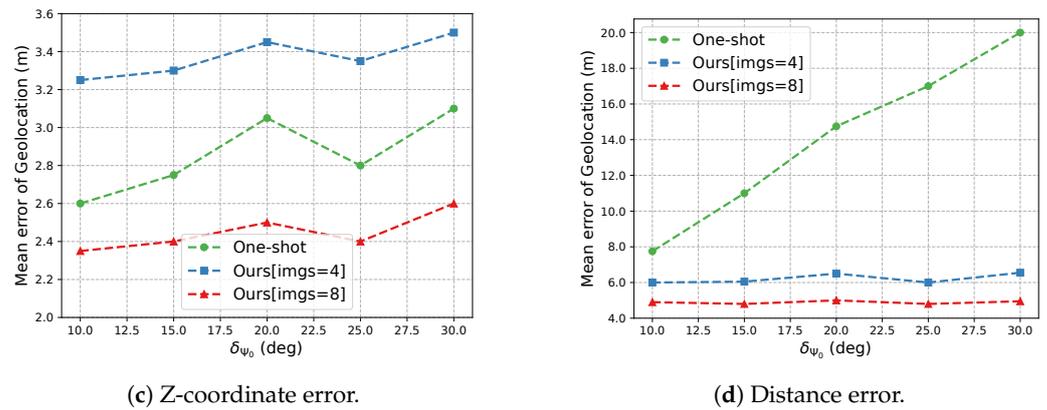


Figure 12. Geolocation errors under different yaw-angle biases.

3.2. Evaluation in Real Environment

A real indoor experiment has also been performed to further validate the proposed framework. The UAVs used in this experiment were a laboratory product designed by our group, as shown in Figure 13a. A camera was installed vertically downward on each UAV, so the attitude of the camera could be known from the attitude of the UAV. The two UAVs tracked the moving target, as shown in Figure 13b, and transmitted their pose information and target images to the ground station in real time. The two UAVs and the ground station were run in the ROS system, and the precise position and attitude information of the UAVs were provided by VICON (a motion capture system).

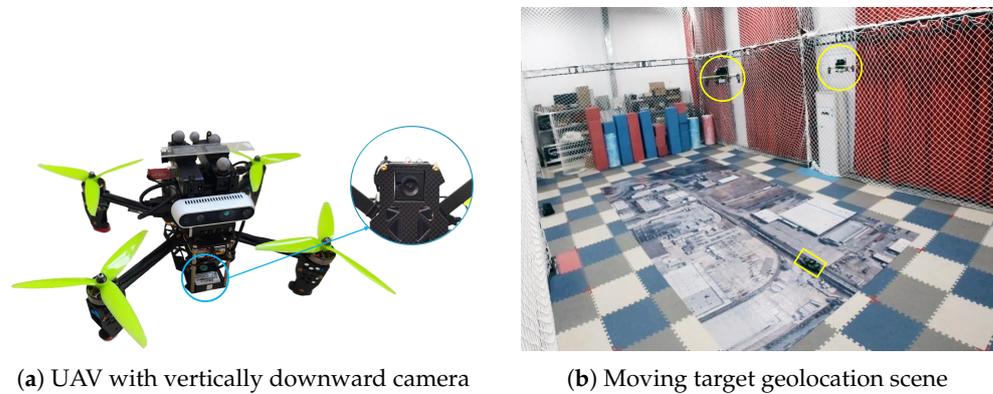


Figure 13. Realistic experimental environment.

The important experimental parameters are shown in Table 5. The results of the realistic experiment are shown in Table 6. A total of 147 geolocation estimates were performed while the UAVs were tracking the moving target. Each geolocation needed to use eight images, among which, the distance between two adjacent images should satisfy $T_{re} \geq 0.25$ m. This shows that our method successfully achieved the geolocation of the moving target, and the absolute mean errors of the three coordinates were 0.035, 0.034, and 0.159 m, respectively. When taking advantage of eight images for geolocation, the FPS was 22, which still meets the real-time requirements.

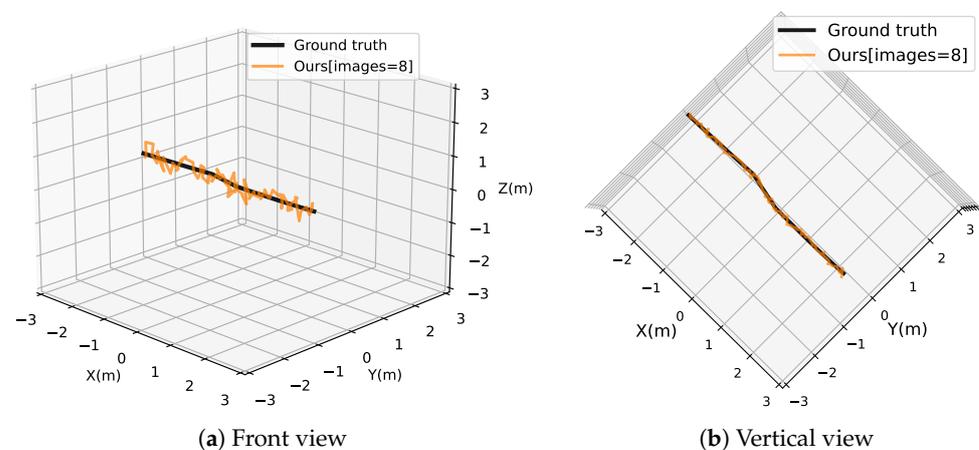
Table 5. Experimental parameters.

Number of Images	Flight Altitude	Flight Speed	σ_ψ	σ_z	$\delta\psi_L$	$\delta\psi_R$
8	2.8 m	0.26 m/s	2°	0.2 m	10°	−10°

Table 6. Geolocation results.

Coordinate	MAE (m)	STD (m)	MAX (m)
X	0.035	0.029	0.125
Y	0.034	0.025	0.104
Z	0.159	0.113	0.446

The geolocation path of the realistic is shown in Figure 14. The black path represents the actual position of the moving target obtained from VICON. The yellow path represents the position of the moving target obtained from our method. It can be seen that the path obtained by our method is undulating due to the influence of Gaussian measurement error of the UAV's navigation state. The above-mentioned simulation experiments have demonstrated that our method can mitigate the effects of Gaussian measurement errors by utilizing the historical measurements.

**Figure 14.** Paths of geolocation.

The simulated and realistic experiments show that the proposed framework implements the function of geolocating the moving target using only UAV vision and does not rely on the accurate AHRS. Compared with the commonly used one-shot method, the proposed framework can mitigate the effects of measurement errors in a UAV's position and attitude by using multiple measurement data and estimating the yaw-angle biases.

4. Discussion

In order to solve the problem that Gaussian measurement errors and system measurement errors of UAV state will lead to a decline in moving-target geolocation accuracy, we proposed a moving-target geolocation framework based on two UAVs. Our method uses a two-step strategy to obtain the final result. The accuracy of the final result depends on three aspects: the matching accuracy for the corresponding points, the accuracy of estimating target altitude, and the accuracy of estimating yaw-angle biases. It can be seen in the experimental results shown in Figure 8 and Table 1 that the proposed corresponding-point-matching method and error-corresponding-point-filtering method can enable historical measurement data to be used to estimate the current position of a moving target. In the subsequent processing step, our method obtains the final result through multiple iterations between target altitude estimation and yaw-angle bias estimation. The experimental results shown in Tables 2 and 3 demonstrate that our method can estimate the target altitude and yaw-angle biases, and the more measurement data are used, the more accurate the estimated target altitude and yaw-angle biases will be. Even so, due to the influence of Gaussian measurement errors, it is still difficult for the proposed framework to accurately estimate the yaw-angle biases, which will be one of the directions of our future exploration.

After that, we evaluated the geolocation performance of the proposed method. For Gaussian measurement errors, obtaining more measurement data is the most effective method to mitigate the influence of Gaussian measurement errors. However, we do not simply increase the number of UAVs to obtain more observation data. If so, the cost will be higher and the system will be extremely complex. Our strategy is to have historical measurement data used to estimate the current position of the moving target through corresponding point matching. From the experimental results in Figures 9–11, we can see that our method can mitigate the influence of Gaussian measurement errors very well. In contrast, the one-shot method is easily affected by the Gaussian measurement error because it can only use one measurement datum. For yaw-angle biases, our strategy is to eliminate the negative impact of yaw-angle biases through multiple iterations between target altitude estimation and yaw-angle bias estimation. The experimental results in Figure 12 show that our method can greatly mitigate the impact of yaw-angle biases on geolocation results compared with the one-shot method, which does not consider yaw-angle biases. It is worth noting that Zhang et al. [24] also considered the influence of yaw-angle biases on target geolocation, but their method assumes that the target is stationary. Unlike this, our method can be applied to moving-target geolocation.

In practical applications, the state measurement error of UAV is an important factor affecting the geolocation accuracy of moving-target tracking. The experimental results show that the proposed method can effectively mitigate the influence of UAV state measurement error on moving-target geolocation accuracy. However, the limitation of our method is that corresponding point matching requires the use of image background information, which is difficult to achieve in some scenes with simple background information (such as the geolocation of a moving target on the sea). Therefore, our next research plan is to develop a more general geolocation method.

5. Conclusions

In this paper, we explored the problem of moving-target geolocation based on a UAV platform with low-quality navigation state sensors. Our aim was to mitigate the influences of Gauss measurement errors and yaw-angle measurement bias from low-quality sensors on the moving-target geolocation accuracy. For this challenge, we proposed a vision-based moving-target geolocation framework using dual UAVs, which can use historical measurement data to estimate the yaw-angle biases of two UAVs and mitigate the impact of Gaussian measurement errors. The results of the simulations and realistic experiments demonstrate that the proposed framework has higher geolocation accuracy than the traditional one-shot based method.

The framework proposed in this paper is only applicable to single target geolocation scenarios. However, multi-target geolocation based on the UAV platform is also one of the urgent needs in practical applications. Therefore, in future work, we will develop a more general geolocation method.

Author Contributions: Conceptualization, B.D. and J.G.; methodology, T.P.; software, T.P.; validation, J.G.; formal analysis, H.D.; investigation, H.D. and B.Z.; resources, T.P.; data curation, J.G.; writing—original draft preparation, T.P.; writing—review and editing, J.G. and H.D.; visualization, T.P.; supervision, B.D.; project administration, B.D.; funding acquisition, B.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 61472095 and 61902423.

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

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

References

1. Potena, C.; Khanna, R.; Nieto, J.; Siegwart, R.; Nardi, D.; Pretto, A. AgriColMap: Aerial-ground collaborative 3D mapping for precision farming. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1085–1092. [[CrossRef](#)]
2. Yuan, Y.; Fang, F.; Zhang, G. Superpixel-based seamless image stitching for UAV images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 1565–1576. [[CrossRef](#)]
3. Zhang, F.; Yang, T.; Liu, L.; Liang, B.; Bai, Y.; Li, J. Image-only real-time incremental UAV image mosaic for multi-strip flight. *IEEE Trans. Multimed.* **2020**, *23*, 1410–1425. [[CrossRef](#)]
4. Gautam, D.; Lucieer, A.; Bendig, J.; Malenovský, Z. Footprint determination of a spectroradiometer mounted on an unmanned aircraft system. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3085–3096. [[CrossRef](#)]
5. Zhao, X.; Pu, F.; Wang, Z.; Chen, H.; Xu, Z. Detection, tracking, and geolocation of moving vehicle from uav using monocular camera. *IEEE Access* **2019**, *7*, 101160–101170. [[CrossRef](#)]
6. Tang, D.; Fang, Q.; Shen, L.; Hu, T. Onboard detection-tracking-localization. *IEEE/ASME Trans. Mechatronics* **2020**, *25*, 1555–1565. [[CrossRef](#)]
7. Deng, F.; Zhang, L.; Gao, F.; Qiu, H.; Gao, X.; Chen, J. Long-range binocular vision target geolocation using handheld electronic devices in outdoor environment. *IEEE Trans. Image Process.* **2020**, *29*, 5531–5541. [[CrossRef](#)] [[PubMed](#)]
8. Yang, X.; Lin, D.; Zhang, F.; Song, T.; Jiang, T. High accuracy active stand-off target geolocation using UAV platform. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–4.
9. Cai, Y.; Ding, Y.; Zhang, H.; Xiu, J.; Liu, Z. Geo-location algorithm for building targets in oblique remote sensing images based on deep learning and height estimation. *Remote Sens.* **2020**, *12*, 2427. [[CrossRef](#)]
10. Fabian, A.J.; Klenke, R.; Truslow, P. Improving UAV-Based Target Geolocation Accuracy through Automatic Camera Parameter Discovery. In Proceedings of the AIAA Scitech 2020 Forum, Orlando, FL, USA, 6–10 January 2020; p. 2201.
11. Liu, C.; Liu, J.; Song, Y.; Liang, H. A novel system for correction of relative angular displacement between airborne platform and UAV in target localization. *Sensors* **2017**, *17*, 510. [[CrossRef](#)] [[PubMed](#)]
12. El Habchi, A.; Moumen, Y.; Zerrouk, I.; Khiati, W.; Berrich, J.; Bouchentouf, T. CGA: A new approach to estimate the geolocation of a ground target from drone aerial imagery. In Proceedings of the 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 21–23 October 2020; pp. 1–4.
13. Namazi, E.; Mester, R.; Lu, C.; Li, J. Geolocation estimation of target vehicles using image processing and geometric computation. *Neurocomputing* **2022**, *499*, 35–46. [[CrossRef](#)]
14. Xu, C.; Huang, D.; Liu, J. Target location of unmanned aerial vehicles based on the electro-optical stabilization and tracking platform. *Measurement* **2019**, *147*, 106848. [[CrossRef](#)]
15. Bai, G.; Liu, J.; Song, Y.; Zuo, Y. Two-UAV intersection localization system based on the airborne optoelectronic platform. *Sensors* **2017**, *17*, 98. [[CrossRef](#)] [[PubMed](#)]
16. Carniglia, P.; Balaji, B.; Rajan, S. Geolocation of mobile objects from multiple UAV optical sensor platforms. In Proceedings of the 2018 IEEE SENSORS, New Delhi, India, 28–31 October 2018; pp. 1–4.
17. Xu, C.; Yin, C.; Han, W.; Wang, D. Two-UAV trajectory planning for cooperative target locating based on airborne visual tracking platform. *Electron. Lett.* **2020**, *56*, 301–303. [[CrossRef](#)]
18. Qiao, C.; Ding, Y.; Xu, Y.; Xiu, J. Ground target geolocation based on digital elevation model for airborne wide-area reconnaissance system. *J. Appl. Remote Sens.* **2018**, *12*, 016004. [[CrossRef](#)]
19. Gao, F.; Deng, F.; Li, L.; Zhang, L.; Zhu, J.; Yu, C. MGG: Monocular Global Geolocation for Outdoor Long-Range Targets. *IEEE Trans. Image Process.* **2021**, *30*, 6349–6363. [[CrossRef](#)] [[PubMed](#)]
20. Zhu, J.; Fang, Y. Learning Object-Specific Distance From a Monocular Image. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3838–3847. [[CrossRef](#)]
21. Wang, X.; Qin, W.; Bai, Y.; Cui, N. Cooperative target localization using multiple UAVs with out-of-sequence measurements. *Aircr. Eng. Aerosp. Technol.* **2017**, *89*, 112–119. [[CrossRef](#)]
22. Xu, C.; Yin, C.; Huang, D.; Han, W.; Wang, D. 3D target localization based on multi-unmanned aerial vehicle cooperation. *Meas. Control* **2021**, *54*, 895–907. [[CrossRef](#)]
23. Pachter, M.; Ceccarelli, N.; Chandler, P.R. Vision-based target geolocation using micro air vehicles. *J. Guid. Control Dyn.* **2008**, *31*, 597–615. [[CrossRef](#)]
24. Zhang, L.; Deng, F.; Chen, J.; Bi, Y.; Phang, S.K.; Chen, X.; Chen, B.M. Vision-based target three-dimensional geolocation using unmanned aerial vehicles. *IEEE Trans. Ind. Electron.* **2018**, *65*, 8052–8061. [[CrossRef](#)]
25. Zhang, L.; Deng, F.; Chen, J.; Bi, Y.; Phang, S.K.; Chen, X. Trajectory planning for improving vision-based target geolocation performance using a quad-rotor UAV. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *55*, 2382–2394. [[CrossRef](#)]

26. Quigley, M.; Conley, K.; Gerkey, B.; Faust, J.; Foote, T.; Leibs, J.; Wheeler, R.; Ng, A.Y. ROS: An open-source Robot Operating System. In Proceedings of the ICRA Workshop on Open Source Software, Kobe, Hyogo Japan, 12–17 May 2009; Volume 3, p. 5.
27. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788. [[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.