

Article

Position and Attitude Estimation Method Integrating Visual Odometer and GPS

Yu Yang ¹, Qiang Shen ^{1,2,*}, Jie Li ¹, Zilong Deng ¹, Hanyu Wang ¹ and Xiao Gao ¹

¹ The School of Mechatronic Engineering, Beijing Institute of Technology, Beijing 100081, China

² Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China

* Correspondence: bit82shen@bit.edu.cn

Received: 17 March 2020; Accepted: 3 April 2020; Published: 9 April 2020

Abstract: The monocular visual odometer is widely used in the navigation of robots and vehicles, but it has defects of the unknown scale of the estimated trajectory. In this paper, we presented a position and attitude estimation method, integrating the visual odometer and Global Position System (GPS), where the GPS positioning results were taken as a reference to minimize the trajectory estimation error of visual odometer and derive the attitude of the vehicle. Hardware-in-the-loop simulations were carried out; the experimental results showed that the positioning error of the proposed method was less than 1 m, and the accuracy and robustness of the attitude estimation results were better than those of the state-of-art vision-based attitude estimation methods.

Keywords: attitude and position estimation; visual odometer; GPS

1. Introduction

The position and attitude are essential measurements in guidance and control of an Unmanned Aircraft Vehicle (UAV), and they are usually obtained by Inertial Navigation Sensors (INS) [1,2] and being corrected by the GPS (Global Position System) [3–5], magnetic sensors [6], and visual odometers [7,8], etc. However, these methods require the initial alignment of INS, leading to the problem that the vehicle needs to stay stationary at the beginning when aligning the INS [9,10]. This problem can be solved by transfer alignment with external attitude measurement equipment [11–13], but it would increase the cost and weight on the navigation system.

Monocular cameras are widely used in the visual navigation of UAVs because they have the advantages of being inexpensive and small. For a UAV equipped with a monocular camera, the attitude estimation can be realized without INS, and the methods of estimating the attitude with pure computer vision are studied by many researchers, while a variety of methods have been proposed, such as horizon detection method [14,15] and vanishing points method [16–18]. Timotheatos proposed a vehicle attitude estimation method based on horizon detection [19], which detects the horizon line based on a canny edge and a hough detector along with an optimization step performed by a Particle Swarm Optimization (PSO) algorithm. The roll angle of the vehicle is determined from the angle formed by the slope of the horizon, and the pitch angle is computed using the current and the initial horizon positions. This method is further studied by Dumble [20], who extracts the horizon more accurately based on the spatial correlation and removes the straight horizon line assumption. Additionally, Hwangbo proposed an attitude estimation method based on vanishing points [21]. The attitude of the vehicle is estimated based on the assumption that the outlines of buildings are either vertical or parallel to the horizon. However, all these methods take external references to estimate attitude, and they may be inefficient when there are no references or the references are not identical to their assumptions.

In addition to using references to achieve attitude estimation, the attitude of the vehicle can be obtained by another vision-based attitude estimation method, Visual Odometer (VO), also known as the Simultaneous Localization And Mapping (SLAM) method [22]. SLAM has been a hot research topic in the last two decades, and has been developed by many researchers, such as ORB-SLAM series proposed by Raul Mur-Artal [23], SVO series proposed by Christian Forster [24], LSD-SLAM series and DSO series proposed by Jakob Engel [25,26]. However, there are defects of the monocular visual odometer that the scale of the estimated trajectory is unknown [27,28], so it is impossible to obtain the positioning results from pure VO with a monocular camera. Through the combination of a low-cost monocular camera and a low-cost GPS receiver, one could get higher positioning accuracy than the GPS receiver and eliminate the scale uncertainty brought by the monocular camera, and the attitude of the vehicle can be extracted from VO as well.

In this paper, we proposed a position and attitude estimation method, integrating the Visual Odometer and GPS, namely GVO (Visual Odometer integrated with GPS). In GVO, the relative position and attitude were estimated by a monocular visual odometer, and the positioning results obtained by GPS were taken as references to eliminate the scale uncertainty of the pose of camera, where the position and attitude of the vehicle could be obtained simultaneously. The application of this method would not be limited by other factors except the general requirements of GPS and visual odometer. For a long-endurance UAV with a monocular vision-based navigation system, when the attitude is estimated by the INS, it is usually used with GPS to improve the robustness and reduce the error of the system. In the long-distance flight of UAV, when the system suddenly loses power in an emergency and restarts in the air, the alignment of INS is difficult, while the GVO method could still work and help align the INS. Furthermore, with the development of image processing hardware, the proposed algorithm could be implemented on the image processing system of the onboard camera without additional costs. So, as an assistant, backup, and even alternative technology to the INS, the proposed method could greatly improve the robustness and stability of the navigation system with promising advantages.

The structure of this paper is given as follows. In Section 2, the proposed method is described in detail, and the scale factor and rotation matrix between VO and GPS are derived, the optimization of estimation results is studied as well. In Section 3, the proposed method is verified by hardware-in-the-loop simulations, and the experimental results show that the proposed method performs better than traditional visual methods, in which attitude estimation error is less than 2° , and the positioning error is less than 1 m. At last, conclusions are drawn in Section 4.

2. The Position and Attitude Estimation Method of GVO

2.1. The Principle of GVO

The principle of the proposed method is shown in Figure 1, where the trajectories of the vehicle are simplified from 3D (3-Dimensional) space to 2D (2-Dimensional) space to help understand where the height information in 3D space is omitted here, and the trajectories are represented on a 2D plane. In Figure 1, the light blue lines indicate the trajectories measured by GPS, the red lines represent the motion trajectories estimated by VO, and the broken lines in yellow and dark blue are the attitude estimated by VO. The trajectory of the vehicle is estimated by GPS and Visual Odometer, represented by \mathbf{x}_n and \mathbf{x}_n^{vo} , respectively, where n represents the index of points. As shown in Figure 1, firstly, the GPS and visual odometer estimate the flight trajectories separately. Secondly, the trajectories estimated by visual odometer and GPS are transformed into the same coordinate system. At last, we assume that the error of GPS measurements \mathbf{x}_n is unbiased white noise [29] and that there is an accumulative error in the estimated position of VO \mathbf{x}_n^{vo} . So, the GPS positioning results are taken as measurements of the true position to correct the estimations of VO, and the scale factor between \mathbf{x}_n^{vo} and \mathbf{x}_n is derived. Considering that the pose of the camera obtained by VO contains the attitude information, the attitude of the vehicle can be estimated as well.

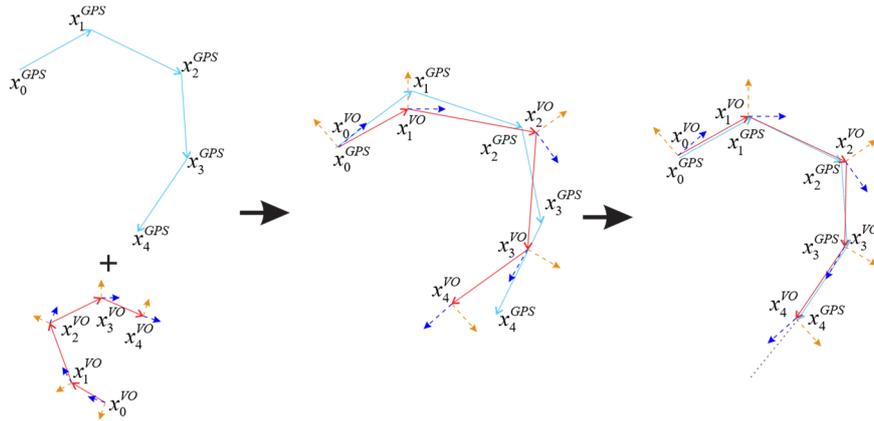


Figure 1. The diagram of integrating Visual Odometry and Global Position System (GPS).

2.2. The Position and Attitude Estimation

The update frequency of GPS is usually 5–10 Hz. The computational frequency of VO is 25–30 Hz, which is several times of the update frequency of GPS. So, it is necessary to have a hardware alignment or software alignment method to realize the synchronization between the pose T_n obtained by VO and x_n estimated by GPS.

Let x_k be the position measurement of GPS, which has been synchronized with VO at frame k , and let T_k be the transform matrix estimated by VO between frame $k-1$ and k . For the sake of brevity, 3D coordinates like $x = (x, y, z)$ are automatically transformed into matrix $x = [x \ y \ z \ 1]^T$ when multiplied with transform matrix T_k . T_k is given as follows,

$$T_k = \begin{bmatrix} R_k & t_k \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (1)$$

where $T_0 = I \in \mathbb{R}^{4 \times 4}$, $R_k \in SO(3)$ is the rotation matrix, $SO(3)$ is the special orthogonal group $SO(3) = \{R_k \in \mathbb{R}^{3 \times 3} | R_k R_k^T = I, \det(R_k) = 1\}$, $t_k \in \mathbb{R}^3$ is the translation matrix.

The position measurements of GPS are usually indicated by geographic coordinates, longitude L_k , latitude B_k , and height H_k . In order to obtain the attitude angles of pitch and roll, the East-North-Up (ENU) coordinate system is selected to be the intermediate coordinate system to fuse the measurements of GPS and VO, where the transformations between geographic coordinates and ENU frame are given in [30]. After time alignment between GPS and VO, the coordinates of GPS are transformed into the ENU frame $x_k^{enu} = (x_k^{enu}, y_k^{enu}, z_k^{enu})$ to integrate with VO.

As the relationship between the initial local frame of VO and ENU frame is certain, the transformation matrix of them is constant. The position and attitude of the camera in the ENU frame can be obtained by using the poses estimated by VO. The transformation between measurements of VO and GPS in the ENU frame can be represented by,

$$R(x_k^{vo} - x_{k-1}^{vo}) = K(x_k^{enu} - x_{k-1}^{enu}), \quad x_k^{vo} = \prod_{j=0}^k T_j x_0^{vo}, \quad (2)$$

where x_k^{vo} is the position in the local frame of VO at frame k , $x_0^{vo} = (0 \ 0 \ 0)$ is the origin in the local frame of VO, and $x_0^{enu} = (0 \ 0 \ 0)$ is the origin of GPS measurements in ENU frame. R is the rotation matrix from the local frame of VO to ENU frame, K is a diagonal matrix, and it represents the scale factor from ENU frame to the local frame of VO, in which the three row vectors represent the scale factors in east, north, and up directions, respectively.

To simplify Equation (2), t_k and d_k are defined as,

$$t_k = x_k^{vo} - x_{k-1}^{vo} = [t_x^k \ t_y^k \ t_z^k]^T, \quad (3)$$

$$d_k = x_k^{enu} - x_{k-1}^{enu} = [d_x^k \ d_y^k \ d_z^k]^T. \quad (4)$$

Thus, obtaining (5),

$$\mathbf{R}\mathbf{t}_k = \mathbf{K}\mathbf{d}_k,$$

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} t_x^k \\ t_y^k \\ t_z^k \end{bmatrix} = \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix} \begin{bmatrix} d_x^k \\ d_y^k \\ d_z^k \end{bmatrix}. \quad (5)$$

For the convenience of calculation, the elements in (5) are moved to the left side as given below,

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & -k_1 & 0 & 0 \\ r_{21} & r_{22} & r_{23} & 0 & -k_2 & 0 \\ r_{31} & r_{32} & r_{33} & 0 & 0 & -k_3 \end{bmatrix} \begin{bmatrix} t_x^k \\ t_y^k \\ t_z^k \\ d_x^k \\ d_y^k \\ d_z^k \end{bmatrix} = 0, \quad (6)$$

However, there is usually no analytic solution of \mathbf{R} and \mathbf{K} due to the white noise in GPS measurements and the accumulative error of VO. So, the least square method is used to solve this equation. To obtain \mathbf{R} and \mathbf{K} , define the solution vector \mathbf{m} as follows,

$$\mathbf{m} = [r_{11} \ r_{12} \ r_{13} \ r_{21} \ r_{22} \ r_{23} \ r_{31} \ r_{32} \ r_{33} \ -k_1 \ -k_2 \ -k_3]^T. \quad (7)$$

When measurements of GPS and VO from frame 1 to frame k are introduced into (6), it can be rewritten as,

$$\begin{bmatrix} \mathbf{t}_k^T & & & d_x^k & & \\ & \mathbf{t}_k^T & & & d_y^k & \\ & & \mathbf{t}_k^T & & & d_z^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{t}_1^T & & & d_x^1 & & \\ & \mathbf{t}_1^T & & & d_y^1 & \\ & & \mathbf{t}_1^T & & & d_z^1 \end{bmatrix} \mathbf{m} = 0. \quad (8)$$

As shown in (6), each pair of positioning results of GPS and VO gives three formulas, while there are 12 unknown variables. So, this over-determined equation could be solved by using more than four pairs of positioning results of GPS and VO, and the solution vector \mathbf{m} can be obtained after the fourth frame of the input image sequence. Then, the estimated matrices $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{K}}$ are given as follows,

$$[\tilde{\mathbf{R}} \quad -\tilde{\mathbf{K}}] \begin{bmatrix} t_x^k \\ t_y^k \\ t_z^k \\ d_x^k \\ d_y^k \\ d_z^k \end{bmatrix} = 0, \quad (9)$$

where $\tilde{\mathbf{R}}$ is the estimated nonorthogonal matrix that approximates the orthogonal matrix \mathbf{R} . Since $\tilde{\mathbf{R}}$ does not always meet orthogonality condition, a further process of $\tilde{\mathbf{R}}$ needs to be done, and (9) is divided by $|\tilde{\mathbf{R}}|$, as shown in (10).

$$\hat{\mathbf{R}} = \tilde{\mathbf{R}}/|\tilde{\mathbf{R}}|, \quad \mathbf{K} = \tilde{\mathbf{K}}/|\tilde{\mathbf{R}}|. \quad (10)$$

Here, the scale factor \mathbf{K} is obtained in this equation. But $\hat{\mathbf{R}}$ may still not be orthogonal, and we have to find an orthogonal matrix, which is the most approximate to $\hat{\mathbf{R}}$, where the Singular Value Decomposition (SVD) method is used to solve this problem, and it is given as follows,

$$\hat{\mathbf{R}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad \mathbf{R} = \mathbf{U}\mathbf{V}^T, \quad (11)$$

where \mathbf{U} , \mathbf{V} are the orthogonal matrices, $\mathbf{\Sigma}$ is a diagonal matrix with positive elements.

In this way, \mathbf{R} is calculated by using more than four pairs of positioning results of GPS and VO, where the selected points must not be on a straight line. Because the linear relationship between the

points would produce infinite solutions, the selection of the points can be conducted by using the positioning results of GPS; as long as they are not on a straight line, they can be used to calculate \mathbf{R} . When there are more data involved in the calculation after a period of flight, the results can be more accurate. This is one of the limitations of the proposed method that the positioning results of GPS and VO used to calculate \mathbf{R} and \mathbf{K} must not be on a straight line. Another limitation is the defect of VO that the ground must appear in the view of the camera, and if there are only ocean and sky, it will lead to large errors in the estimations of VO.

Then, the orientation \mathbf{R}_k in the pose \mathbf{T}_k of VO can be used to derive the attitude of the vehicle \mathbf{a}_k in the ENU frame, which is given by

$$\mathbf{a}_k = \mathbf{R}\mathbf{R}_k. \quad (12)$$

Through the equations above, the rotation matrix and scale factor between GPS and VO are obtained, and the attitude of the vehicle is derived as well.

2.3. The Optimization of GVO

However, the rotation matrix and scale factor between GPS and VO derived in the last section are not optimal solutions because there are accumulative errors in the poses obtained by VO, which would lead to a large error in the estimated rotation matrix and scale factor. In order to reduce the error of estimated position and attitude as much as possible, the optimization process has to be done.

According to (5), the position measurements of VO in the ENU frame are given as follows,

$$\mathbf{x}_k^{\text{ve}} - \mathbf{x}_{k-1}^{\text{ve}} = \mathbf{K}^{-1}\mathbf{R}(\mathbf{x}_k^{\text{vo}} - \mathbf{x}_{k-1}^{\text{vo}}), \quad (13)$$

where $\mathbf{x}_k^{\text{ve}} \in \mathbb{R}^3$ is the position estimated by VO, which has been transformed to the ENU frame.

After the transformation in (13), the optimization process is carried out, and the proposed optimization algorithm is given in Algorithm 1.

Algorithm 1 Optimization of GVO

Require: Positioning results of GPS $\mathbf{x}_k^{\text{enu}}$ and poses of camera \mathbf{T}_k .

1: Optimize \mathbf{R} and \mathbf{K} by minimizing error between GPS measurements $\mathbf{x}_k^{\text{enu}}$ and the positions estimated by VO \mathbf{x}_k^{ve} with initial values obtained by (10) and (11).

2: Optimize the poses \mathbf{T}_k of camera by minimizing error between GPS measurements $\mathbf{x}_k^{\text{enu}}$ and the positions estimated by VO \mathbf{x}_k^{ve} .

3: Use BA (Bundle Adjustment) to optimize the camera poses \mathbf{T}_k again by minimizing the image reprojection error between the matched 3D points in ENU frame and their keypoints in the image.

4: **if** The optimization converges **then**

5: **return** The optimized \mathbf{R} , \mathbf{K} and \mathbf{T}_k .

6: **else**

7: Repeat step1 to step 6

8: **end if**

There are three major steps of the optimization process. Firstly, taking \mathbf{R} and \mathbf{K} obtained by (10) and (11) as the initials to help accelerate the convergence of optimization. They are optimized by minimizing errors between GPS measurements and the positions estimated by VO, which is given by,

$$\{\mathbf{R}, \mathbf{K}^{-1}\} = \underset{\mathbf{R}, \mathbf{K}^{-1}}{\operatorname{argmin}} \sum_k \|\mathbf{x}_k^{\text{enu}} - \mathbf{x}_k^{\text{ve}}\|_2^2 = \underset{\mathbf{R}, \mathbf{K}^{-1}}{\operatorname{argmin}} \sum_k \|\mathbf{x}_k^{\text{enu}} - (\mathbf{x}_{k-1}^{\text{ve}} + \mathbf{K}^{-1}\mathbf{R}(\mathbf{x}_k^{\text{vo}} - \mathbf{x}_{k-1}^{\text{vo}}))\|_2^2. \quad (14)$$

Let \mathbf{T}_k^{ve} be the poses of VO, which have been transformed to the ENU frame,

$$\mathbf{T}_k^{\text{ve}} = \begin{bmatrix} \mathbf{R}_k^{\text{ve}} & \mathbf{t}_k^{\text{ve}} \\ \mathbf{0}^T & 1 \end{bmatrix} = \mathbf{K}^{-1}\mathbf{R}\mathbf{T}_k. \quad (15)$$

After the optimization of \mathbf{R} and \mathbf{K} , the poses of the camera \mathbf{T}_k^{ve} at each frame are optimized by minimizing the errors between the GPS positioning results and the poses estimated by VO, which is represented by,

$$\mathbf{T}_{k,0}^{ve} = \prod_{j=0}^k \mathbf{T}_j^{ve}, \mathbf{x}_k^{ve} = \mathbf{T}_{k,0}^{ve} \mathbf{x}_0^{ve}, \quad (16)$$

$$\{\mathbf{T}_{k,0}^{ve}\} = \underset{\mathbf{T}_{k,0}^{ve}}{\operatorname{argmin}} \sum_k \|\mathbf{x}_k^{\text{enu}} - \mathbf{T}_{k,0}^{ve} \mathbf{x}_0^{ve}\|_2^2. \quad (17)$$

At last, we use BA (Bundle Adjustment) to optimize the camera poses by minimizing the image reprojection error between the matched 3D points in ENU frame and their key points in the image, where the method is referenced in [31]. The position of the ground mark $\mathbf{x}_{\text{mark}}^{ve}$ in the ENU frame is given by

$$\mathbf{x}_{\text{mark}}^{ve} = \mathbf{x}_0^{ve} + \mathbf{K}^{-1} \mathbf{R}(\mathbf{x}_{\text{mark}}^{\text{vo}} - \mathbf{x}_0^{\text{vo}}), \quad (18)$$

where $\mathbf{x}_{\text{mark}}^{\text{vo}}$ is the position of ground mark in the local frame of VO.

After the optimization process above, the system would judge whether the optimization result converges or not. If it converges, the optimization result will be output. Otherwise, perform the optimization process again until it converges.

3. Hardware-in-the-Loop Simulation

3.1. The Experimental Environment

Hardware-in-the-loop simulations were conducted to verify the proposed method, and the block diagram of the simulation system is shown in Figure 2.

As shown in Figure 2, the simulation system consisted of three parts—the simulation computer, PX4 embedded flight controller, and Nvidia Jetson TX2 embedded image processing board. The computer was used to simulate the visual environment, the PX4-embedded flight controller was used to control the attitude of aircraft, and the Nvidia Jetson TX2-embedded image processing board was used to run the GVO method. The CPU of the computer was Intel i7-7700K 4.20 GHz, the GPU of the computer was Nvidia GTX 1080 8 G, and the RAM of the computer was 32 GB. The integral proportional guidance law was added to the PX4-embedded flight controller with existing codes to help guidance control, where the codes could be downloaded on the website [32]. The size and weight of the TX2 module were 50×87 mm and 85 g, respectively. The power of TX2 when running the algorithm was 20 W and the cost of it was \$399.

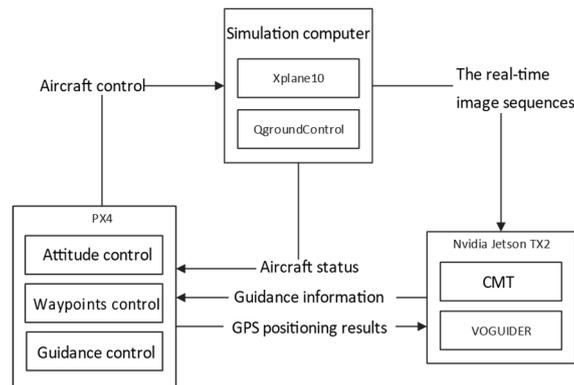


Figure 2. Block diagram of the simulation system.

The block diagram of the working process of the simulation system is shown in Figure 3. First, the 3D environment was rendered by Xplane10 on the computer. The real-time image sequences were captured during flight by setting the installation and intrinsics of the camera. Then, the image sequences would be transmitted to the NVIDIA image processing board to run the proposed GVO method and get the estimated position and attitude of UAV, where the GPS positioning results were obtained from the PX4 flight controller. The estimated position and attitude were fed back to PX4 so that it could realize attitude control using the preset waypoints of the Qground Control software on

the computer and give the attitude control message to Xplane10. When Xplane10 received the attitude control message from PX4, it changed the flight status of the aircraft, while the dynamic model of the aircraft and the windy model of the environment were taken into account. In addition, the flight status of the UAV would be fed back to PX4 to form a closed loop, which would assist PX4 to control the attitude of the UAV during flight. The update frequency of the PX4 controller, the image output frequency of XPlane 10, and the image sampling frequency of NVIDIA Jetson were all 50 Hz.

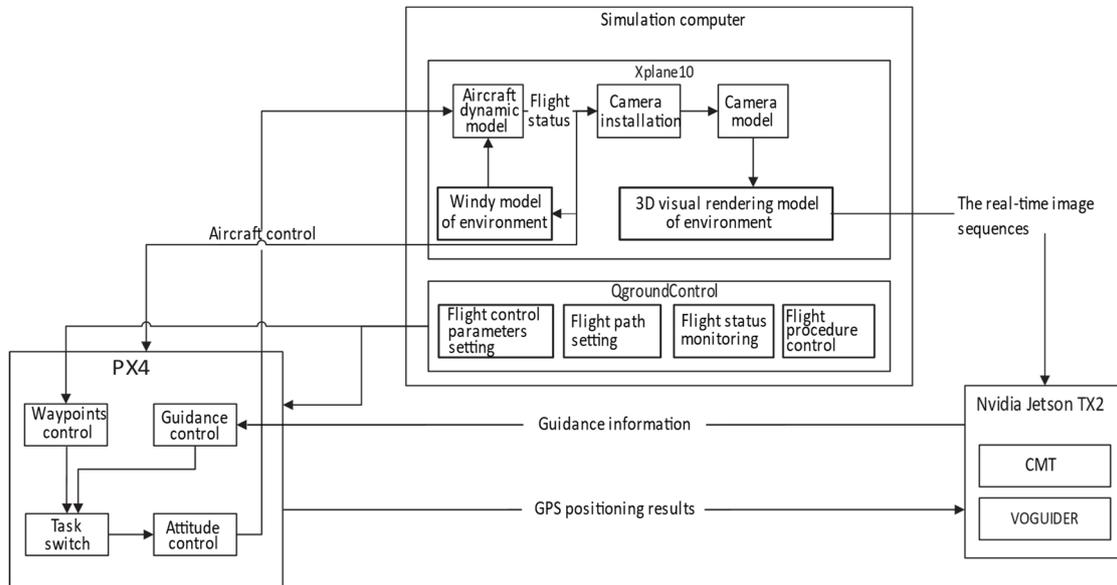


Figure 3. Block diagram of the working process of the simulation system.

3.2. Positioning Results of GVO

To verify the positioning accuracy of GVO in different conditions, multiple heights were employed to conduct simulations, and the simulation conditions are shown in Table 1.

Table 1. The simulation conditions.

Number	Flying Height	Flying Speed	Field Angle of Camera	Resolution of Camera
1	100 m	30 m/s	45° × 45°	1024 × 1024
2	200 m	30 m/s	45° × 45°	1024 × 1024
3	300 m	30 m/s	45° × 45°	1024 × 1024
4	400 m	30 m/s	45° × 45°	1024 × 1024
5	500 m	30 m/s	45° × 45°	1024 × 1024

The flying speed in Table 1 is the cruising speed of MyTwinDream, a fixed-wing Unmanned Aircraft Vehicle (UAV), manufactured by FPVmodel in Xiamen China. Due to the disturbance of wind, the speed of UAV was not fixed during flight. The camera parameters here were the parameters of a commonly used onboard camera.

The positioning results of GVO and GPS are shown in Figure 4, where the blue lines are estimated by GVO, and the red lines are measurements of GPS with a standard variance of 5 m.

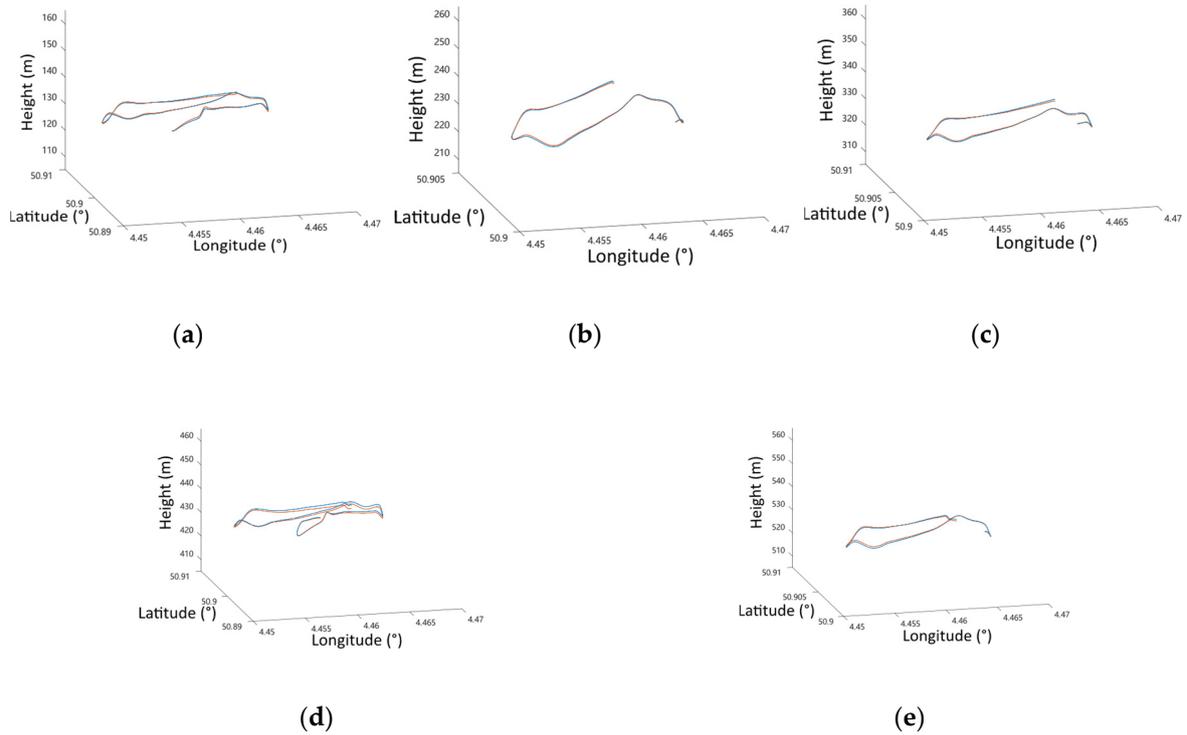


Figure 4. The trajectory estimation results. (a) 100 m height. (b) 200 m height. (c) 300 m height. (d) 400 m height. (e) 500 m height.

As we could see from Figure 4, the trajectories estimated by GVO basically overlapped with the original GPS measurements. The statistics of the positioning errors of GVO are given in Table 2, and the corresponding bar diagram is shown in Figure 5.

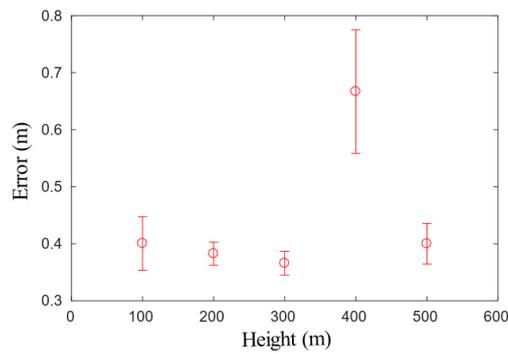


Figure 5. The bar diagram of positioning error.

Table 2. The aircraft positioning error.

Height /m	Mean Deviation /m	Deviation Variance /m ²
100	0.41	0.13
200	0.37	0.08
300	0.36	0.08
400	0.67	0.17
500	0.41	0.11

As we could see from Table 2 and Figure 5, the mean deviation of positioning results was less than 1 m, and the standard variance of GVO was less than 0.7 m, which was much smaller than that of pure GPS with a standard variance of 5 m.

The convergence property of the proposed method was tested as well. The positioning errors of GVO are shown in Figure 6, and they were convergent in all cases, while the slowest one converged after the 250th frame at the frame rate of 25 frames per second. It showed that the system converged in a short time, which was within 10 s by using the proposed method.

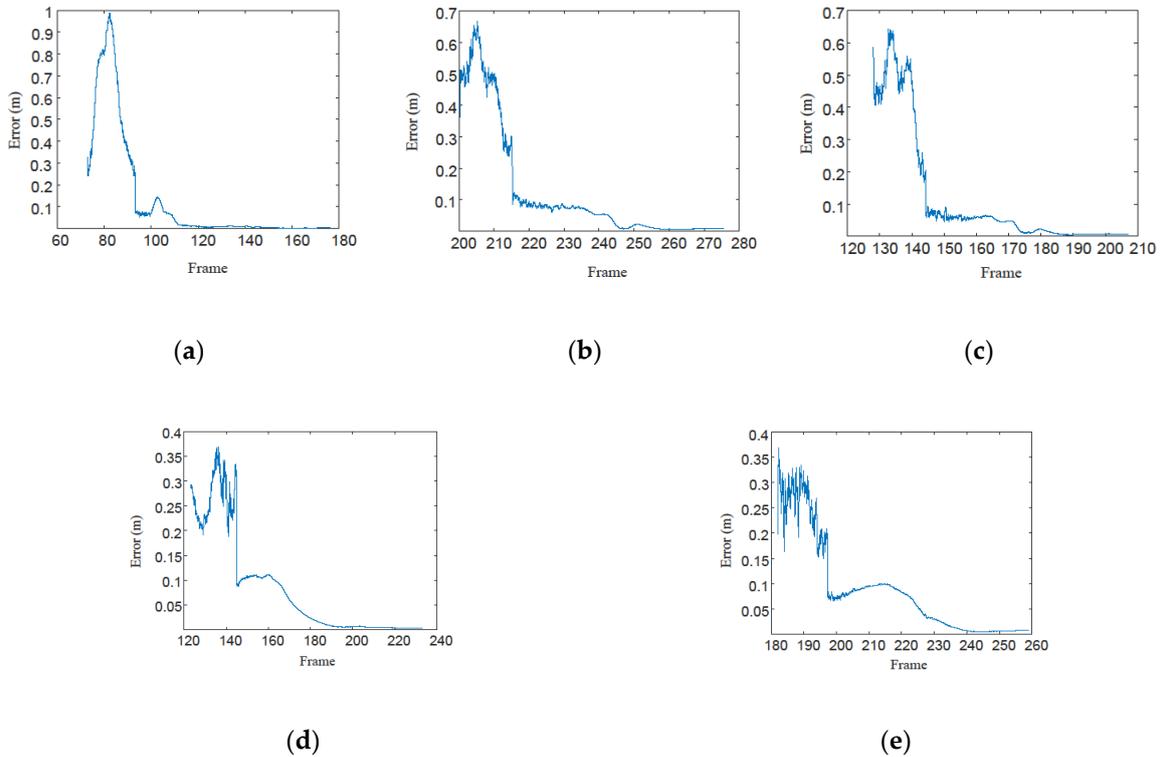


Figure 6. The positioning errors of GVO (Visual Odometer integrated with GPS). (a) 100 m height. (b) 200 m height. (c) 300 m height. (d) 400 m height. (e) 500 m height.

3.3. Attitude Estimation Results

To assess the robustness and accuracy of GVO on attitude estimation, this section has compared GVO with two state-of-art monocular visual attitude estimation methods—the horizon detection method, and the vanishing points method.

The premise of the horizon detection method for attitude estimation is that the horizon must appear in the view of the camera. This algorithm estimated the roll angle of the camera by detecting the slope of the horizon, and the pitch angle was estimated from a distance between the center of the horizon and the central axis of the camera, which is shown in Figure 7a. The vanishing points method estimated attitude when there were a large number of artificial constructions. The attitude was estimated by the vanishing points of outlines of the buildings when the outlines were supposed to be parallel or vertical to the horizon, as shown in Figure 7b.



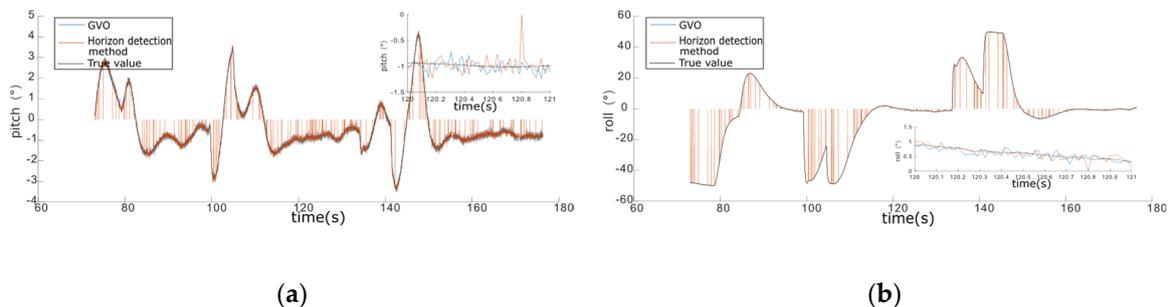
Figure 7. The attitude estimation method. (a) The horizon detection method. (b) The vanishing points method.

Simulations were carried out in multiple environments like the plain, mountain, and city to compare the proposed method with different state-of-art attitude estimation methods, and the experimental conditions are given in Table 3.

Table 3. Experimental conditions.

Number	Area	Height	Algorithms to Compare
1	Plain	100	Horizon detection method
2	Plain	200	Horizon detection method
3	Plain	300	Horizon detection method
4	Mountain	100	Horizon detection method
5	Mountain	200	Horizon detection method
6	Mountain	300	Horizon detection method
7	City	100	Vanishing points method
8	City	200	Vanishing points method

The horizon detection method was compared with the proposed method in plain and mountain areas, where the experimental results are given in Figures 8 and 9. As shown in Figure 8, the attitude estimation results of GVO were very close to those of the horizon detection method in the plain area, but the horizon detection method failed occasionally, which was also at the time when the estimated values jumped to zero. The horizon detection method judged whether the detected lines met the length threshold of the horizon to ensure the horizon detection accuracy, and it failed when there were no detected lines that could meet the requirements of the horizon assumption. The GVO method was not affected by the environment, so there were no failures of GVO in the simulations.



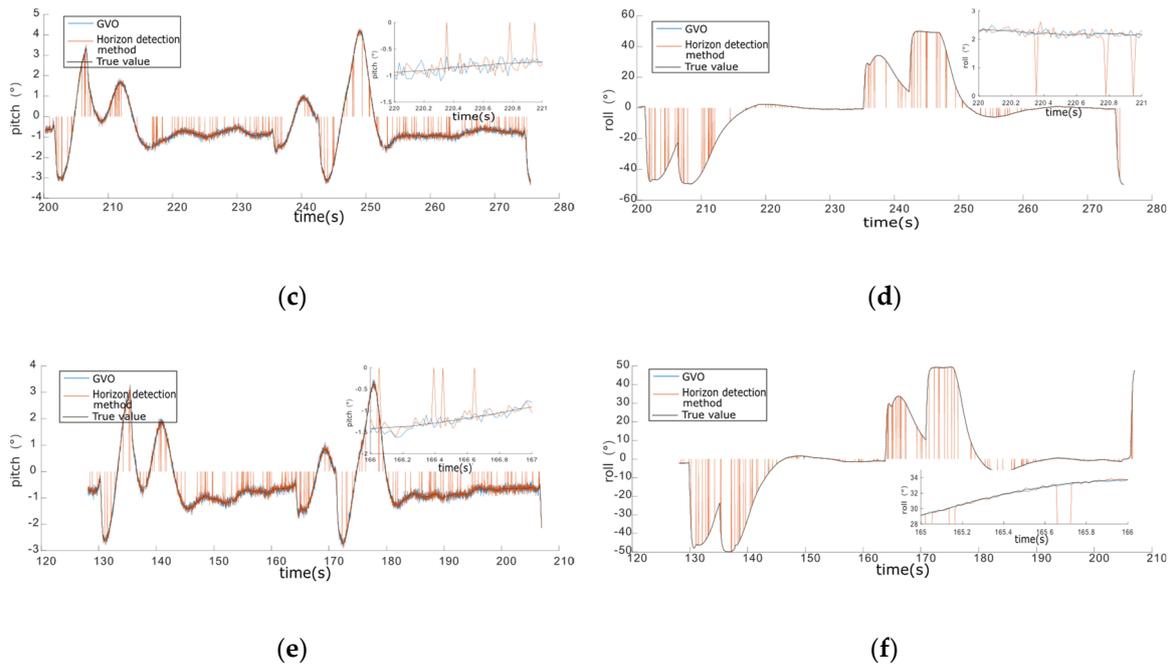
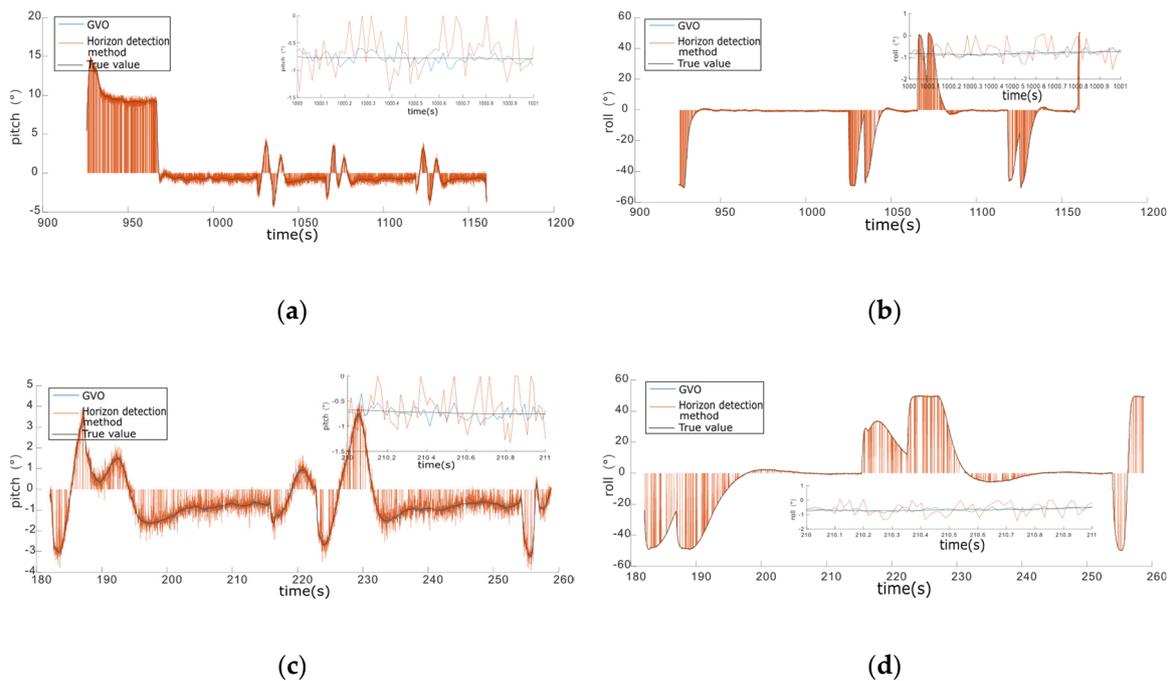


Figure 8. The simulation results of GVO and horizon detection method in the plain area. (a) Pitch angle of 100 m height. (b) The roll angle of 100 m height. (c) Pitch angle of 200 m height. (d) The roll angle of 200 m height. (e) Pitch angle of 300 m height. (f) The roll angle of 300 m height.

The simulation results of GVO and horizon detection method in the mountainous area are shown in Figure 9. As shown in Figure 9, the accuracy and success rate of GVO were obviously higher than that of the horizon detection method. Compared with the plain area, the horizon in the mountainous area was more difficult to be detected. We could see from Figure 9 that the failure frequency of the horizon detection method was much larger than that of the plain area, and the attitude estimation accuracy in the mountainous area was worse. Meanwhile, the GVO still worked properly, which proved that the advantages of GVO were more prominent in the mountainous areas.



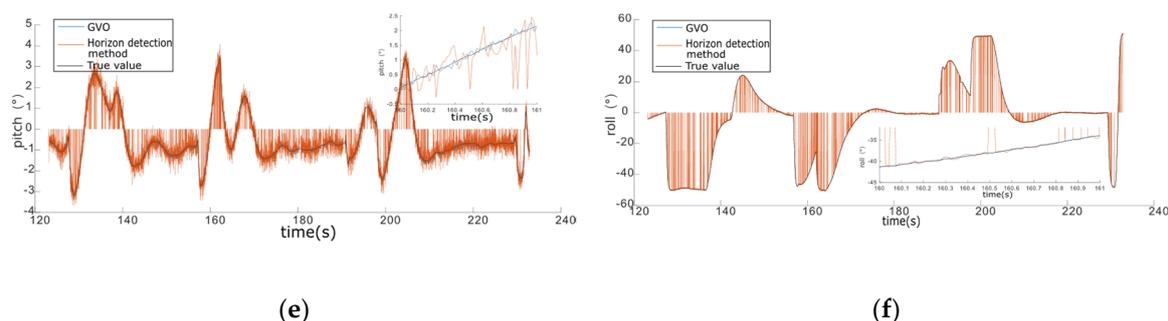


Figure 9. The simulation results of GVO and horizon detection method in the mountainous area. (a) Pitch angle of 100 m height. (b) The roll angle of 100 m height. (c) Pitch angle of 200 m height. (d) The roll angle of 200 m height. (e) Pitch angle of 300 m height. (f) The roll angle of 300 m height.

The statistics of GVO and horizon detection method in plain and mountain environments are given in Table 4. As shown in Table 4, the accuracy of GVO was a little higher than that of the horizon detection method, but the success rate of GVO was much larger, especially in the mountainous area. As long as VO and GPS could work properly, the success rate of GVO was always 100%, which proved that the proposed method could estimate the attitude efficiently without being influenced by the external environment. So, the GVO method had the advantages of higher accuracy and better robustness than the horizon detection method.

Table 4. Comparison between GVO (Visual Odometer integrated with GPS) and horizon detection method.

Algorithm	Area	Height /m	Success Rate	Average Error /°
GVO	Plain	100	100%	1.2
Horizon detection	Plain	100	97%	2.0
GVO	Plain	200	100%	1.1
Horizon detection	Plain	200	98%	2.1
GVO	Plain	300	100%	1.3
Horizon detection	Plain	300	97%	2.0
GVO	Mountain	100	100%	1.2
Horizon detection	Mountain	100	83%	2.5
GVO	Mountain	200	100%	1.1
Horizon detection	Mountain	200	85%	2.3
GVO	Mountain	300	100%	1.1
Horizon detection	Mountain	300	91%	2.2

As shown in Figure 10, simulations in the city were conducted to compare GVO with the vanishing points method, and the statistics are given in Table 5. As we could see from Figure 10 and Table 5, the success rate of the vanishing points method decreased sharply with the increase of height, while the success rate of the GVO method was still 100%. When the height increased, the outlines of the buildings got shorter, and the detection accuracy of the vanishing points method would be reduced when the lengths of the lines were too short to meet the detection threshold. The GVO was not affected by the variation of height, and it had higher attitude estimation accuracy with an average error of 1.2°, while the average error of the vanishing point method was 2.9°.

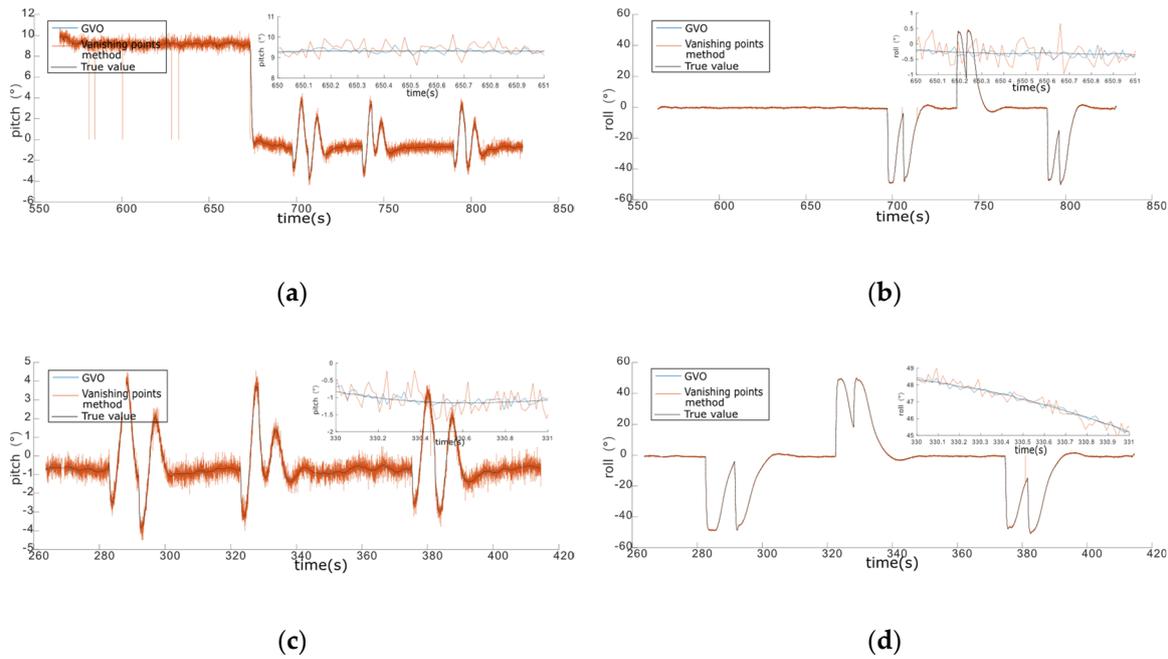


Figure 10. The simulation results of the GVO method and the vanishing points method in the city area. (a) Pitch angle of 100 m height. (b) The roll angle of 100 m height. (c) Pitch angle of 200 m height. (d) The roll angle of 200 m height.

Table 5. Comparison between GVO and vanishing points method.

Algorithm	Height	Success Rate	Average Error /°
GVO	100	100%	1.3
Vanishing points	100	93%	3.1
GVO	200	100%	1.1
Vanishing points	200	82%	2.7

Overall, GVO showed better performance of accuracy and robustness in attitude estimation than the state-of-art methods. From the simulations above, the success rate of GVO in attitude estimation was always 100% in spite of the variation of environment and height. In comparison, the average success rates of the horizon detection method and vanishing points method were 91.3% and 87.5%, respectively. The attitude estimation error of GVO ranged from 1.1° to 1.3°, while the error of horizon detection method ranged from 2.0° to 2.1° in plain area and 2.2° to 2.5° in the mountainous area. The average error of vanishing points method ranged from 2.7° to 3.1° in the city.

4. Conclusions

In this paper, an algorithm, integrating the Visual Odometer and GPS, was proposed to estimate the position and attitude of UAV. The integration method of VO and GPS was described, and the estimations of the proposed method were optimized as well. Hardware-in-the-loop simulations were conducted; the TX2 image processing board that we used was released in 2016, and there are alternative image processing chips with smaller sizes, such as Huawei Hi3559A with the size of 38 × 38 mm, manufactured by Hisilicon in Shanghai China, which is more suitable to be applied on UAV. As such chips enable this method to be applied in engineering. The experimental results showed that the mean deviation of positioning results of GVO was less than 1 m, and the standard variance was less than 0.7 m. Simulations compared with two state-of-art visual attitude estimation methods were carried out to verify the proposed method on attitude estimation, where the results showed that the GVO method performed better in both accuracy and robustness.

Author Contributions: Conceptualization, Y.Y. and J.L.; Data Curation, Y.Y.; Formal Analysis, Y.Y.; Funding Acquisition J.L.; Investigation, J.L.; Methodology, Y.Y. and J.L.; Project Administration, Q.S.; Resources, Q.S.; Software, Y.Y.; Supervision, Q.S.; Validation, Z.D.; Visualization, Z.D.; Writing-Original Draft Preparation, Z.D.; Writing-Review and Editing, H.W. and X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61973033.

Acknowledgments: We would like to thank the National key laboratory of science and technology on electromechanical dynamic characteristics at the Beijing Institute of Technology for their support during the writing of this paper.

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

References

1. Savage, P.G. Strapdown Inertial Navigation Integration Algorithm Design Part 2: Velocity and Position Algorithms. *J. Guid. Control. Dyn.* **1998**, *21*, 208–221.
2. Hao, Y.; Xiong, Z.; Gao, W.; Li, L. Study of strapdown inertial navigation integration algorithms. In Proceedings of the 2004 International Conference on Intelligent Mechatronics and Automation, Chengdu, China, 26–31 August 2004; pp. 751–754.
3. Christian, E.; Lasse, K.; Heiner, K. Real-time single-frequency GPS/MEMS-IMU attitude determination of lightweight UAVs. *Sensors* **2015**, *15*, (10), 26212–26235.
4. Rhudy, M.; Gross, J.; Gu, Y.; Napolitano, M. Fusion of GPS and Redundant IMU Data for Attitude Estimation. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Minneapolis, MN, USA, 13–16 August 2012; p. 5030.
5. Ahmad, I.; Benallegue, A.; El Hadri, A. Sliding mode based attitude estimation for accelerated aerial vehicles using GPS/IMU measurements. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 3142–3147.
6. Zhang, J.; Jin, Z.-H.; Tian, W.-F. A suboptimal Kalman filter with fading factors for DGPS/MEMS-IMU/magnetic compass integrated navigation. In Proceedings of the Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, Shanghai, China, 12–15 October 2003; pp. 1229–1234.
7. Zhang, J.; Singh, S. INS Assisted Monocular Visual Odometry for Aerial Vehicles. In *Field and Service Robotics*. Springer: Berlin/Heidelberg, Germany, 2015, 183–197.
8. Sirtkaya, S.; Seymen, B.; Alatan, A. A. Loosely coupled Kalman filtering for fusion of Visual Odometry and inertial navigation. In Proceedings of 16th International Conference on Information Fusion (FUSION), Istanbul, Turkey, 9–12 July 2013; pp. 219–226.
9. Chouaib, I.; Wainakh, B.M.; Khalaf, C.W. Robust self-corrective initial alignment algorithm for strap-down INS. In Proceedings of the 2015 10th Asian Control Conference (ASCC); Institute of Electrical and Electronics Engineers (IEEE), Kota Kinabalu, Malaysia, 31st May–3rd June, 2015; pp. 1–6.
10. Tan, C.; Zhu, X.; Su, Y.; Wang, Y.; Wu, Z.; Gu, D. A New Analytic Alignment Method for a SINS. *Sensors* **2015**, *15*, 27930–27953.
11. Zhao, L.; Gao, W.; Li, P.; Nie, Q. The study on transfer alignment for SINS on dynamic base. In Proceedings of the 2005 IEEE International Conference on Mechatronics and Automation, Niagara Falls, Ont., Canada, 29 July–1 August 2006; Vol. 3, pp. 1318–1322.
12. Hao, Y.; Xiong, Z. W.; Wang. Rapid transfer alignment based on unsented Kalman filter. In Proceedings of the American Control Conference 2006, Minneapolis, MN, USA, 14–16 June.
13. Chen, Y.; Zhao, Y. New rapid transfer alignment method for SINS of airborne weapon systems. *J. Syst. Eng. Electron.* **2014**, *25*, 281–287.

14. Dusha, D.; Boles, W.; Walker, R. Attitude Estimation for a Fixed-Wing Aircraft Using Horizon Detection and Optical Flow. In Proceedings of the 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007), Glenelg, Australia, 3–5 December 2007; pp. 485–492.
15. Boroujeni, N.S.; Etemad, S.A.; Whitehead, A. Robust Horizon Detection Using Segmentation for UAV Applications. In Proceedings of the 2012 Ninth Conference on Computer and Robot Vision; Toronto, ON, Canada, 2012; pp. 346–352.
16. Shabayek, A.E.R.; Demonceaux, C.; Morel, O.; Fofi, D. Vision Based UAV Attitude Estimation: Progress and Insights. *J. Intell. Robot. Syst.* **2011**, *65*, 295–308.
17. Rother, C. A New Approach for Vanishing Point Detection in Architectural Environments. *Image Vision Comput.* **2002**, *20*, 647–655.
18. Denis, P.; Elder, J.H.; Estrada, F.J. Efficient Edge-Based Methods for Estimating Manhattan Frames in Urban Imagery. In Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 12–18 October 2008; Vol. 5303, pp. 197–210.
19. Timotheatos, S.; Piperakis, S.; Argyros, A.; Trahanias, P. Vision Based Horizon Detection for UAV Navigation. In Proceedings of the 27th International Conference on Robotics in Alpe-Adria-Danube Region, Patras, Greece, 6–8 June 2018; pp. 181–189.
20. Dumble, S.J.; Gibbens, P.W. Horizon Profile Detection for Attitude Determination. *J. Intell. Robot. Syst.* **2012**, *68*, 339–357.
21. Hwangbo, M.; Kanade, T. Visual-inertial UAV attitude estimation using urban scene regularities. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 2451–2458.
22. Nister, D.; Naroditsky, O.; Bergen, James R. Visual odometry. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; DOI: 10.1109/CVPR.2004.1315094.
23. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.; R., M.-A.; J.D., T. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163.
24. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 15–22.
25. Engel, J.; Schöps, T.; Cremers, D. LSD-SLAM: Large-Scale Direct Monocular SLAM. In Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 834–849.
26. Engel, J.; Koltun, V.; Cremers, D. Direct Sparse Odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 611–625.
27. Esteban, I.; Dorst, L.; Dijk, J. Closed Form Solution for the Scale Ambiguity Problem in Monocular Visual Odometry. In Proceedings of Third International Conference Intelligent Robotics and Applications, Shanghai, China, 10–12 November; pp. 665–679.
28. Choi, S.; Park, J.; Yu, W. Resolving scale ambiguity for monocular visual odometry. In Proceedings of the 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Jeju, Korea, 30 October–2 November 2013; pp. 604–608.
29. Kaplan, E.D.; Christopher, J.H. *Understanding GPS: Principles and Applications*, 2nd ed.; Artech House: Norwood, MA, USA, 2005; ISBN 978-1-58053-894-7.
30. Cai, G., Chen, B.M., Lee, T.H., Eds.; *Advances in Industrial Control*; Springer: London, UK, 2011; pp. 23–34. ISBN 0857296345.

31. Mur-Artal, R.; Tardos, J. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262.
32. PX4/Firmware. Available online: <https://github.com/PX4/Firmware/tree/v1.8.0> (accessed on 19 Jun 2018).



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).