Artificial Marker and MEMS IMU-Based Pose Estimation Method to Meet Multirotor UAV Landing Requirements

Yibin Wu, **Xiaoji Niu** *, **Junwei Du, Le Chang, Hailiang Tang and Hongping Zhang**

GNSS Research Center, Wuhan University, No. 129 Luoyu Road, Wuhan 430079, China;
ybwu@whu.edu.cn (Y.W.); dujunwei@whu.edu.cn (J.D.); changlesgg@whu.edu.cn (L.C.); thl@whu.edu.cn (H.T.);
hpzhang@whu.edu.cn (H.Z.)
* Correspondence: xjniu@whu.edu.cn

Abstract: The fully autonomous operation of multirotor unmanned air vehicles (UAVs) in many applications requires support of precision landing. Onboard camera and fiducial marker have been widely used for this critical phase due to its low cost and high effectiveness. This paper proposes a six-degrees-of-freedom (DoF) pose estimation solution for UAV landing based on an artificial marker and a micro-electromechanical system (MEMS) inertial measurement unit (IMU). The position and orientation of the landing maker are measured in advance. The absolute position and heading of the UAV are estimated by detecting the marker and extracting corner points with the onboard monocular camera. To achieve continuous and reliable positioning when the marker is occasionally shadowed, IMU data is fused by an extended Kalman filter (EKF). The error terms of the IMU sensor are modeled and estimated. Field experiments show that the positioning accuracy of the proposed system is at centimeter level, and the heading error is less than 0.1 degrees. Comparing to the marker-based approach, the roll and pitch angle errors decreased by 33% and 54% on average. Within five seconds of vision outage, the average drifts of the horizontal and vertical position were 0.41 and 0.09 m, respectively.

Keywords: multirotor UAV; precision landing; artificial marker; pose estimation; sensor fusion

---

1. Introduction

With the rapid development of micro-electromechanical systems (MEMSs) and high-performance flight control processors, applications and research of multirotor unmanned air vehicles (UAVs) have received increasing attention worldwide. Due to the unique structure and stability, multirotor UAVs are widely used in precision agriculture [1,2], rescue [3], surveillance [4,5], etc. These applications also pose new challenges for the precision landing of UAVs [6]. To land safely and precisely, the localization of UAV requires high accuracy, continuity, and reliability.

Global navigation satellite system/inertial navigation system (GNSS/INS)-integrated navigation [7] technologies are the most widely used navigation methods for various UAVs [8,9]. However, the precision of the integrated navigation positioning is completely dependent on the GNSS. Although decimeter-level positioning results can be obtained by a consumer-grade dual-frequency GNSS receiver with sufficient satellites [10,11], in areas with low satellite observation such as city canyons and forests, positioning errors will increase. Moreover, it is totally unavailable for indoor applications. To ease these problems, vision-based pose estimation approaches for UAV landing have attracted extensive research due to their low cost and ability to sense the environment [12,13]. A survey of the vision-based landing techniques for UAVs is presented in [14]. In most applications, e.g., precision agriculture, we

can determine the landing point of UAVs where we are allowed to place a customized pad in advance to take advantage of the pertaining features.

Abundant works concerning the design of the fiducial marker system have been proposed in literature, such as ARTag [15], AprilTag [16], and ArUco [17]. The authors in [18,19] use AprilTag markers as their landing pad. In [20], ARTag markers are used for UAV localization. Researchers also have proposed specialized markers to meet precision landing requirements of UAVs.

The authors in [21] use video from the onboard camera and the orientation and altitude of the ground vehicle to calculate the UAV position with respect to (w.r.t.) the target. In [22], an integration algorithm of monocular and stereo vision is exploited to estimate the approach angle and relative height of a UAV by Hough transform and vanishing-line geometry. In [23], the authors introduce a dynamic fiducial marker that can change its appearance according to the requirements of the UAV. A landing marker which consists of the letter "H" inside a black circle is presented in [24]. A landing pad composed of markers of different size is utilized to permit detection from high as well as very low altitude in [25]. In [26], the authors describe a pattern comprising several concentric white rings on a black background. The rings are identified by the unique ratio between their inner and outer border radius. The authors in [27] present a landing marker consisting of three inner circles, each of which is evenly divided into eight fan areas. The area between the inner circles is filled by an even distribution of black and white. They use the template-matching strategy for marker detection in daytime, and a morphology technique named the hit-and-miss algorithm at nighttime. In 2018, the authors extended this work by introducing a method to exploit the trained features extracted from a convolutional neural network (CNN) for marker tracking, which can be performed at very high altitude [28]. The authors in [29] propose a graphics processing unit (GPU)-based pose estimation method based on a marker composed of circles and use experimental covariance measurements to estimate the camera pose during occlusions and vibrations. A comparison of the state of the art on the autonomous landing of UAVs using artificial markers is also shown in this work, regarding the marker shape, positioning error, frame rate, etc. Authors in [27] classify the previous vision-based landing systems for UAVs into two categories (namely passive and active methods) and compare their strength and weakness.

In the above marker-based positioning methods, when the marker is severely occluded, the system will not work. Furthermore, these approaches do not take into consideration the absolute heading angle, although some of them do retrieve the relative heading. To increase the positioning rate as well as cover the short-period marker occlusions, inertial measurements can be introduced to fuse with image data.

The authors in [24] use projective geometry to estimate the five-degrees-of-freedom (DOF) pose from the elliptic projection of the marker. Then the gravity vector estimated by the inertial measurement unit (IMU) is incorporated to resolve the remaining geometric ambiguity. In [25], the authors perform a homography-based method to obtain UAV pose from the marker and integrate it with the IMU data in a loosely coupled way. Their work focus on the landing of UAVs on a moving platform. However, the horizontal angles, which are essential for flight control, are not estimated in the system. Moreover, the IMU sensor errors are not modeled, which can result in significant position error accumulation with time.

In this work, a six-DoF state estimation system based on an artificial marker, a visible light monocular camera, and a MEMS IMU to meet the requirements of multirotor UAV landing is introduced. The IMU used in our system is at dollar-level, which is widely used for commercial smartphones. Referring to the prior works about marker design and detection systems, an improved methodology applicable to UAV landing is proposed. The inner corners of the marker are used for pose estimation when the UAV is too close to the ground and the whole marker is out of the image frame. IMU data is fused by an extended Kalman filter (EKF) to increase the navigation result rate to permit aggressive maneuvers and to cope with short-time marker occlusion. This also results in more smoothing results for both position and attitude. Unlike previous visual/inertial fusion system for UAV landing, models of the error components of IMU are considered and roll and pitch angles are also

predicted by INS and updated in the EKF. Field tests illustrate the magnificent performance of our system. It worth mentioning that this work only focuses on the pose estimation for the UAV landing instead of on the whole closed-loop control system.

This paper is organized as follows. In Section 2, the design and detection of the marker as well as the relative state estimation method with the extracted marker is presented. Then the fusion algorithm of pose information from the marker and IMU data is described. Experimental evaluations of the proposed system are provided and discussed in Section 3. Finally, Section 4 draws some conclusions.

2. Methods

2.1. Marker Detection

Similar to previous works, the marker proposed in this work is a big square consisting of small black and white squares. They are distributed on a plane according to a certain rule with the advantages of large information capacity and good expandability.

The marker is constituted by $N \times N$ black and white squares of the same size. The number and size of the squares are dependent on the size of the dataset, as well as the amount of encoded information. The border of the marker is black with several black squares properly filled inside. To avoid the central symmetry of the marker, which leads to ambiguous heading definition, it must be ensured that the ultimate marker is not centrosymmetric when determining internal black squares. Our marker is specifically designed for UAV landing, reducing the computational cost while limiting false positives. Unlike the aforementioned markers, the structure inside the border of the proposed marker is much simple. To meet the requirement of estimating pose at different heights along the landing path especially when the UAV is close to the ground, a black square is set at the marker center. At the beginning, the UAV detects the whole marker and adjust its horizontal position and heading while landing. Once it descends to a certain height where the marker is too large to be detected, the center square can be used for guidance. Thus, N should be an odd number.

A marker database is consequently built for different landing points. Every marker has its own specific number and involves a different command for the UAV. We use a binary bit stream to represent the coding matrix of the marker for fast and accurate matching, with values of 0 for black and 1 for white. The $(N-2) \times (N-2)$ matrix uniquely represents the encoding information of the marker. Figure 1 shows examples of two different markers ($N = 7$) and their encoding matrices.

Figure 1. Examples of the proposed marker and binary encoding matrix.

The main process of pose estimation by the marker is shown in Figure 2. First, we perform histogram equalization and morphological processing on the captured image to reduce the noise and then execute binarization. The Otsu [30] algorithm is employed since it provides the optimal image threshold given that the pixel distribution is bimodal, which is true in this scenario. We use the Canny [31] operator to extract the contours in the image and the Douglas–Peucker algorithm to fit the contours to a polygon. After removing the unreasonable contours, all the quadrilateral contours are extracted, and the internal pixels are extracted and clustered to match the encoding matrix in the database. After determining the unique matrix, the vertices of the marker on the image are obtained, and all corners inside the marker are extracted [32]. If the entire marker cannot be extracted, the

corners are tracked by calculating the optical flow [33]. When the UAV is too close to the marker, the corners of the central black square are extracted. Since the positions of the corner points are known, the position and orientation of the camera can be estimated by a set of correspondences between real world points and their projection on image. This is introduced in Section 2.2. Figure 3a–f presents the image processing results of the proposed system.
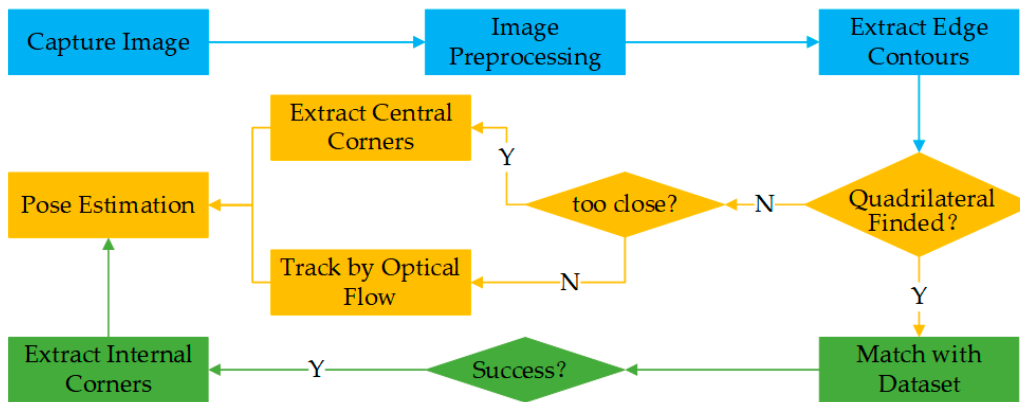

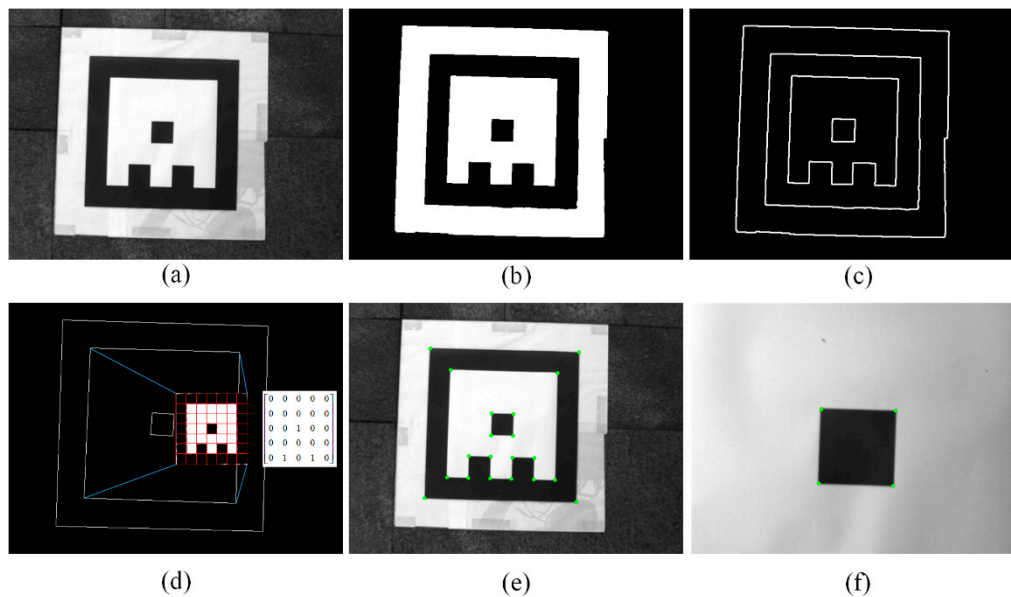**Figure 2.** Flow chart of pose estimation by detecting the marker.

Figure 3. Image processing results. (**a**) Original image; (**b**) binary image after Otsu binarization; (**c**) polygon contours extracted in (**b**); (**d**) quadrilateral contours and the matched marker; (**e**) the corner points extracted on the marker; (**f**) the extracted corner points of the central marker when the unmanned air vehicle (UAV) is very close to the ground.

2.2. Pose Estimation by Marker

The relative pose between two coordinate systems can be retrieved by matching the corresponding points. The world coordinate system (w-frame) is established with the center of the marker as the origin. The camera coordinate system (c-frame) is established with the optical center of the camera as the origin, x-axis pointing to the right, y-axis pointing down. The pixel coordinate system (p-frame) is established with the upper left corner of the image as the origin and (u, v) parallel to (x, y) of the c-frame, respectively. Figure 4 shows the coordinate systems. The camera projection model is shown as Equation (1). The marker can be laid arbitrarily at the landing point, but we have to measure its

heading angle in advance for UAV absolute heading estimation. For easy understanding, we assume that the *x*-axis and *y*-axis of the w-frame points to the north and east, respectively.
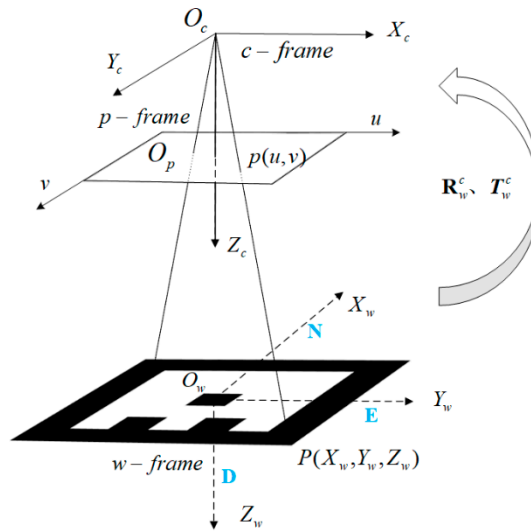
Figure 4. Definition of coordinate systems and camera projection model. O_c is the optic center of the camera and $O_c - X_cY_cZ_c$ denotes the c-frame; $O_p - uv$ denotes the p-frame; and $O_w - X_wY_wZ_w$ denotes the w-frame.

The transformation among coordinate systems is given as

$$\boldsymbol{P}^c = \mathbf{R}_w^c \boldsymbol{P}^w + \boldsymbol{T}_w^c, \tag{1}$$

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{K}\boldsymbol{P}^c, \tag{2}$$

where $\boldsymbol{P}^c$ and $\boldsymbol{P}^w$ are the coordinates of the corner points in c-frame and w-frame, respectively; (u, v) are the coordinates of the corner points in p-frame; $\mathbf{R}_w^c$ and $\boldsymbol{T}_w^c$ are the rotation matrix and translation vector from w-frame to c-frame, respectively; $\mathbf{K}$ is the intrinsic parameter matrix of the camera, which can be calibrated by the chessboard method proposed in [34]; (c_x, c_y) is the camera principal point in p-frame; and f_x, f_y represent the focal length of the camera in terms of pixel dimensions in the *x* and *y* direction, respectively. According to the matching pairs of the corners in w-frame and p-frame, the estimation of the UAV pose $(\mathbf{R}_w^c, \boldsymbol{T}_w^c)$ is a perspective-n-point (PNP) problem that can be solved by [35]. The random sample consensus (RANSAC) method [36] is introduced to exclude corner outliers to ensure the robustness of the visual positioning. Then, we regard it as the initial value and perform the Levenberg–Marquardt [37] approach to optimize it according to the minimum reprojection error criterion, shown as

$$\min_{\mathbf{R}_w^c, \boldsymbol{T}_w^c} \sum_{i \in n} \rho_i \left\| \boldsymbol{P}_i - \frac{1}{Z_c} \mathbf{K}(\mathbf{R}_w^c \boldsymbol{P}^w + \boldsymbol{T}_w^c) \right\|, \tag{3}$$

where $i \in n$ denotes the *i*th point of the all *n* points; ρ_i indicates the weight factor. To reduce the influence of a corner observation with a large error on the result, we use the Cauchy loss function (Equation (4)) to reduce the weight of this observation [38].

$$\rho(s) = \log(1 + s) \tag{4}$$

The relative position of the camera w.r.t. the w-frame can be transformed to the absolute pose in the e-frame by the following equations,

$$\delta r^e = \mathbf{D}\delta r^n, \tag{5}$$

$$\mathbf{D} = \begin{bmatrix} R_M + h & 0 & 0 \\ 0 & (R_N + h)\cos\varphi & 0 \\ 0 & 0 & -1 \end{bmatrix}, \tag{6}$$

where δr^e is the position vector from the origin of w-frame to the camera center projected in the e-frame; δr^n is the camera coordinates in the n-frame; R_M is the meridian radius of curvature; R_N is the radius of curvature in the prime vertical; φ is the geodetic latitude; and h is the altitude.

2.3. Visual/Inertial Fusion

INS is able to calculate UAV pose from the initial state by IMU measurements alone. By detecting the marker, the relative position and heading of the UAV in the w-frame is retrieved which provides reliable external observations for INS to correct its errors. Since the corner points are on the same plane that is parallel to the horizontal plane, the observability of the horizontal angle (i.e., roll and pitch angle) is weak for PNP. For INS, the external position assist and gravity constraint make the horizontal angle observable. Moreover, the IMU data rate is higher than that of the camera in general. INS mechanization can predict the UAV state during two consecutive image frames. Due to the magnificent complementarity, it is valuable to fuse the two sensors' data. The algorithm overview is presented in Figure 5.
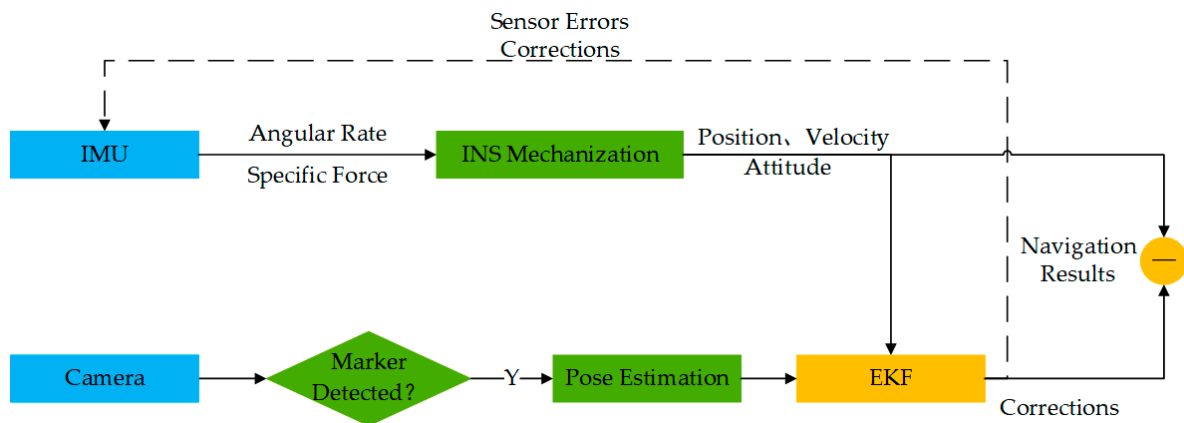
Figure 5. System overview of the camera/INS (inertial navigation system) integrated navigation. The vehicle pose is retrieved by the two sensors independently and then fused by error-state extended Kalman filter (EKF), namely loosely coupled.

2.3.1. INS Mechanization and System Model

The main procedure of the INS mechanization can be divided into three steps:

1. Integrate the angular rate to update the vehicle attitude;
2. Transform the acceleration to the n-frame by the updated attitude, then integrate it to obtain the vehicle velocity;
3. Integrate the velocity to calculate the vehicle position.

Refer to [39] for the detailed update equations.

The demands for an accurate estimation of the vehicle pose necessitate the modeling and online estimation of the error components of the sensor. The state vector of the integrated navigation system constructed in this paper can be expressed as

$$x(t) = \begin{bmatrix} \delta r^n & \delta v^n & \phi & b_g & b_a & s_g & s_a \end{bmatrix}, \tag{7}$$

where δr^n is the position error in the navigation frame (n-frame), which has its origin coinciding with that of the b-frame, with its x-axis pointing north, z-axis pointing down, and y-axis completing a

right-hand orthogonal frame, i.e., the north–down–east (NED) system; δv^n is the velocity error in the n-frame; $\boldsymbol{\phi}$ is the attitude error, composed of roll error, pitch error, and heading error; $\boldsymbol{b}_g$ and $\boldsymbol{b}_a$ are the residual bias error of the gyroscope and the accelerometer, respectively; $\boldsymbol{s}_g$ and $\boldsymbol{s}_a$ are the residual scale factor error of the gyroscope and the accelerometer, respectively.

The continuous time system is modeled as,

$$\dot{\boldsymbol{x}}(t) = \mathbf{F}(t)\boldsymbol{x}(t) + \mathbf{G}(t)\boldsymbol{w}(t), \tag{8}$$

where, $\mathbf{F}(t)$ is the system dynamics matrix; $\mathbf{G}(t)$ is the noise distribution matrix of the system; and $\boldsymbol{w}(t)$ is the vector describing the system noise.

Due to the uncertainty in the sensors and the gravity field, the navigation parameters obtained from the INS mechanization equation contain errors. The phi-angle error model [40] is used here. Thus, the differential equation of the inertial navigation error is as follows:

$$\dot{\boldsymbol{\phi}} = -\boldsymbol{\omega}_{in}^n \times \boldsymbol{\phi} + \delta\boldsymbol{\omega}_{in}^n - \mathbf{C}_b^n \delta\boldsymbol{\omega}_{ib}^b, \tag{9}$$

$$\delta\dot{\boldsymbol{v}}^n = \mathbf{C}_b^n \delta\boldsymbol{f}^b + \mathbf{C}_b^n \boldsymbol{f}^b \times \boldsymbol{\phi} + \boldsymbol{v}^n \times (2\delta\boldsymbol{\omega}_{ie}^n + \delta\boldsymbol{\omega}_{en}^n) - (2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n) \times \delta\boldsymbol{v}^n + \delta\boldsymbol{g}_l^n, \tag{10}$$

$$\delta\dot{\boldsymbol{r}}^n = -\boldsymbol{\omega}_{en}^n \times \delta\boldsymbol{r}^n + \delta\boldsymbol{\theta} \times \boldsymbol{v}^n + \delta\boldsymbol{v}^n, \tag{11}$$

$$\delta\boldsymbol{\theta} = \begin{bmatrix} \delta r_E / (R_N + h) \\ -\delta r_N / (R_M + h) \\ -\delta r_E \tan\varphi / (R_N + h) \end{bmatrix} \tag{12}$$

where, b denotes the body frame (b-frame), which has its axes coinciding with the IMU's body; $\delta\boldsymbol{\omega}_{ib}^b$ and $\delta\boldsymbol{f}^b$ are the gyroscope measurement error and the accelerometer measurement error, respectively; $\boldsymbol{\omega}_{ie}^n$ is the rotation rate vector of the e-frame with respect to the inertial frame (i-frame) projected to the n-frame; $\boldsymbol{\omega}_{en}^n$ is the rotation rate vector of the n-frame with respect to the e-frame projected to the n-frame; $\delta\boldsymbol{g}_l^n$ is the gravity vector projected to the n-frame; $\mathbf{C}_b^n$ is the rotation matrix from the body frame (b-frame) to the n-frame; δr_N and δr_E are the position error on north and east direction of the n-frame, respectively; The variables $\boldsymbol{b}_g$, $\boldsymbol{b}_a$, $\boldsymbol{s}_g$, and $\boldsymbol{s}_a$ are modeled by a first-order Gauss–Markov process.

$$\begin{aligned} \dot{x} &= -\tfrac{1}{T}x + w \\ x_{k+1} &= e^{-\Delta t_{k+1}/T} x_k + w_k \end{aligned} \tag{13}$$

where T is the correlation time of the process, and w is the driving white noise process.

2.3.2. Linearized Measurement Model

Since the IMU and the camera cannot be installed at the same place with parallel axes, the pose retrieved from the marker is different from that derived from the INS, which is known as the lever-arm effect and installation angle effect, as described in Equations (14) and (15), respectively. These misalignments have to be calibrated in advance or they will cause significant errors in both position and attitude. For easy understanding, it is assumed that the b-frame coincides with the UAV coordinate system (v-frame).

The installation relationship between the camera and IMU is depicted in Figure 6.

The camera position derived from the IMU and camera is expressed as follows,

$$\begin{aligned} \hat{\boldsymbol{r}}_c^e &= \hat{\boldsymbol{r}}_{IMU}^e + \mathbf{C}_n^e \hat{\mathbf{C}}_b^n \boldsymbol{l}_c^b \\ &= \boldsymbol{r}_{IMU}^e + \delta\boldsymbol{r}_{IMU}^e + \mathbf{C}_n^e [\mathbf{I} - (\boldsymbol{\phi}\times)] \mathbf{C}_b^n \boldsymbol{l}_c^b \\ &= \boldsymbol{r}_c^e + \delta\boldsymbol{r}_{IMU}^e + \mathbf{C}_n^e (\mathbf{C}_b^n \boldsymbol{l}_c^b \times) \boldsymbol{\phi} \\ \widetilde{\boldsymbol{r}}_c^e &= \boldsymbol{r}_c^e + \mathbf{C}_n^e \boldsymbol{n}_{rc} \end{aligned} \tag{14}$$

where $\hat{r}_c^e$ is the camera position calculated by the IMU-derived position and lever arm; $\widetilde{r}_c^e$ is the camera position estimated by marker detection; l_c^b is the lever arm vector from the IMU center to the camera center in the b-frame; n_{rc} is the position error of visual estimation; and []× represents the antisymmetric matrix of a vector.

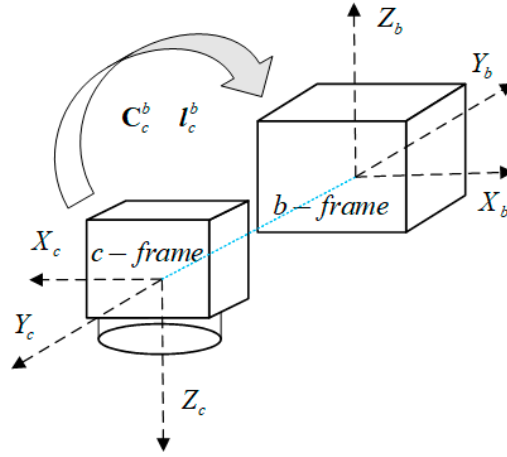**Figure 6.** Installation relationship between the camera and inertial measurement unit (IMU).

Since the landing maker can provide a reliable heading reference, it can also be exploited to correct INS errors as an aiding source. The camera attitude derived from the IMU and camera are expressed as follow,

$$\begin{aligned} \hat{\mathbf{C}}_c^n &= \hat{\mathbf{C}}_b^n\mathbf{C}_c^b = [\mathbf{I} - (\boldsymbol{\phi}\times)]\mathbf{C}_b^n\mathbf{C}_c^b \\ \widetilde{\mathbf{C}}_c^n &= [\mathbf{I} + (\boldsymbol{v}\times)]\mathbf{C}_c^n \end{aligned}, \tag{15}$$

$$\begin{aligned} \hat{\mathbf{C}}_c^n &= [\mathbf{I} - (\boldsymbol{\rho}\times)]\widetilde{\mathbf{C}}_c^n \\ [\mathbf{I} - (\boldsymbol{\phi}\times)] &= [\mathbf{I} - (\boldsymbol{\rho}\times)][\mathbf{I} + (\boldsymbol{v}\times)] \end{aligned}, \tag{16}$$

$$\begin{aligned} \boldsymbol{\rho}\times &= \mathbf{I} - \hat{\mathbf{C}}_c^n\widetilde{\mathbf{C}}_c^{nT} \\ \boldsymbol{\rho} &= \boldsymbol{\phi} + \boldsymbol{v} \end{aligned}, \tag{17}$$

where $\hat{\mathbf{C}}_c^n$ and $\widetilde{\mathbf{C}}_c^n$ are the camera attitude matrices derived from the INS and recovered from the visual solution, respectively; $\mathbf{C}_c^b$ is the rotation matrix from the c-frame to the b-frame; v is the heading error of the visual estimation; and ρ is the attitude difference between the vehicle attitude calculated by INS and vision. Hence, the observation vector can be expressed as a combination of the position error observation and heading error observation as follows:

$$\mathbf{Z} = \begin{bmatrix} z_{rc}^n \\ \rho_{heading} \end{bmatrix}, \tag{18}$$

where, $\rho_{heading}$ is the third element of ρ and z_{rc}^n can be written as

$$\begin{aligned} z_{rc}^n &= \mathbf{C}_e^n\left(\hat{r}_c^e - \widetilde{r}_c^e\right) \\ &= \mathbf{C}_e^n\left(\delta r_{IMU}^e + \mathbf{C}_n^e\left(\mathbf{C}_b^n l_c^b\times\right)\boldsymbol{\phi} - \mathbf{C}_n^e n_{rc}\right) \\ &= \delta r^n + \left(\mathbf{C}_b^n l_c^b\times\right)\boldsymbol{\phi} - n_{rc} \end{aligned}. \tag{19}$$

## 3. Experimental Results Discussion

3.1. Experiments Description

To demonstrate the performance of the proposed system, we carried out field tests and analyzed the results from two aspects: the retrieved pose errors and reprojection errors of the corner points. The GNSS and high-precision fiber-optic IMU (POS320, MAP Space Time Navigation Technology Co., LTD, Wuhan, China) integrated navigation systems were used as references for the pose. The camera and consumer grade MEMS IMU (ICM20602)-integrated navigation system was the system under test. The camera used in the experiments was a global shutter camera with a 644×484 resolution (Mako G-030C, Allied Vision, Stadtroda, Germany). The IMU data rate was 50 fps, and the image frame rate was 20 fps. Figure 7 shows the experimental equipment platform. The spatial displacement of the IMU and the camera was calibrated in advance [41]. Table 1 provides the technical parameters of the two IMUs. The position and heading of the marker in the e-frame were previously obtained by GNSS post processed kinematic (PPK) static measurements. The size of the marker used in the tests was 29 cm $\times$ 29 cm.

Table 1. Technical parameters of the micro-electromechanical system (MEMS) IMU and POS320.

| Parameters	MEMS IMU (System under Test)	POS320 (Reference System)
Gyro bias stability $(°/h)$	200	0.5
Angle random work $(°/\sqrt{h})$	0.24	0.05
Accelerometer bias stability (m/s^2)	0.01	0.00025
Velocity random work $(m/s/\sqrt{h})$	3	0.1
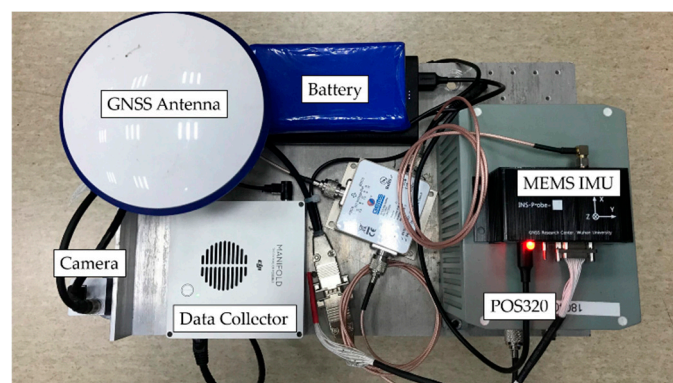
Figure 7. Experimental equipment.

Without the expensive motion capture system, it is too difficult, if not impossible, to obtain a reliable vehicle pose as ground truth with light load. A machine with four stepper motors, allowing precise movement of the camera in 3D space, was designed and built in [29] to evaluate the estimated position of the camera w.r.t. the maker. Although the high-precision GNSS/INS integrated systems can provide six-DOF pose references, they are usually too heavy (more than 10 kg) to be mounted on the UAV (like the one used in our experiments). Therefore, we hold the experimental device with reference system by hand to mimic the motion and trajectory of the UAV when landing to evaluate the pose accuracy convincingly. The GNSS/INS pose reference system was fixed on a rigid platform with the camera and MEMS IMU. The position and angle misalignment between the two IMUs was calibrated previously to project the ground truth to the MEMS IMU center. The position of the antenna phase center relative to the reference system was also measured in advance. The camera, IMU, and GNSS data were collected for postprocessing. We performed post processed kinematic (PPK) for the GNSS data process, then the smoothing results of the GNSS/INS integrated navigation were calculated

as ground truth. The experiments were executed five times on Friendship Square at Wuhan University on 28 February 2019. The trajectory and attitude variation of the test platform in the fifth test is shown in Figures 8 and 9. The experimental environment is shown in Figure 10.
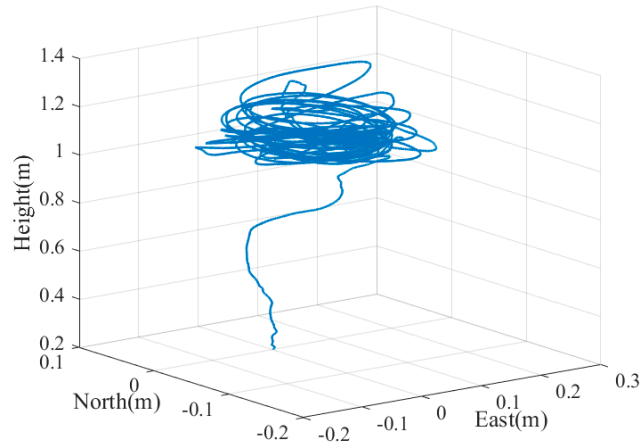
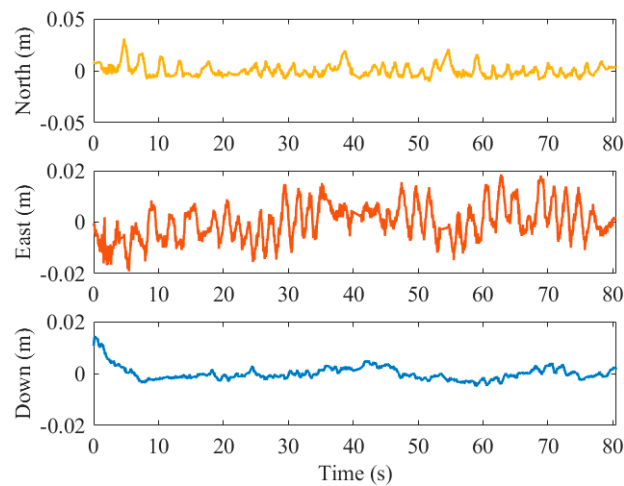**Figure 8.** Trajectory of the test platform in test 05.
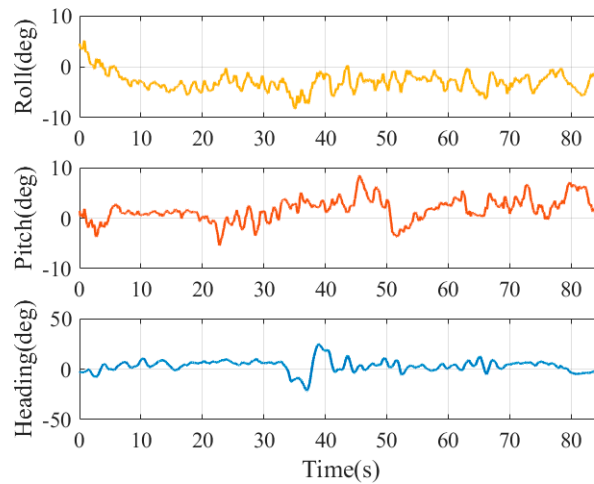
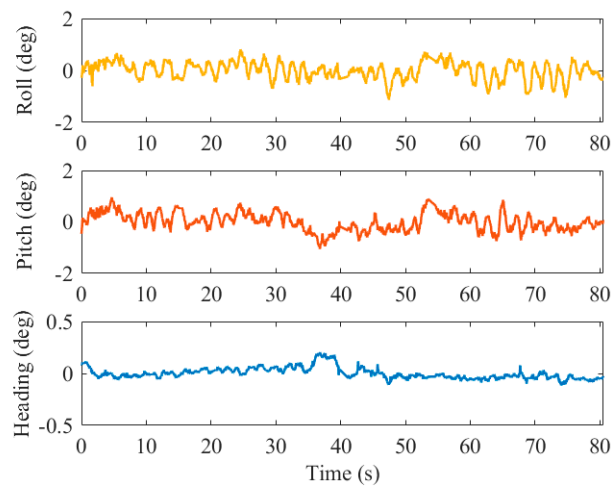**Figure 9.** Attitude variation of the test platform in test 05.
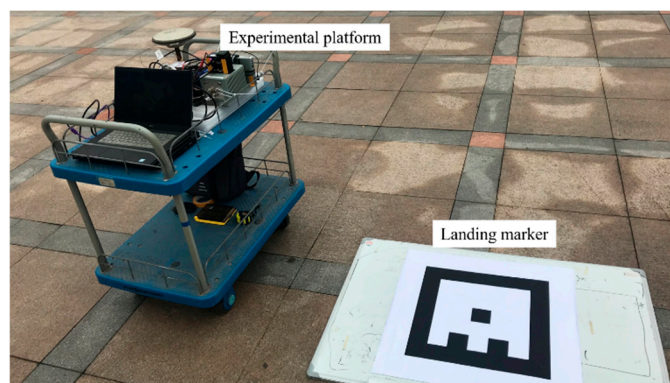
Figure 10. Experimental environment.

3.2. Results Discussion

The position and attitude errors of the results by the pure visual solution in the third test are shown in Figures 11 and 12, respectively.

Figure 11. Position errors of the visual solution in test 03.

Figure 12. Attitude errors of the visual solution in test 03.

The statistical results of the position and attitude errors of the visual solution for five experiments are shown in Tables 2 and 3, respectively. The maximum and mean values of the error are calculated by the absolute error sequence over the whole article.

Table 2. Statistical results of the position errors of the visual solution.

Statistics		01	02	03	04	05
	RMSE	0.007	0.006	0.008	0.006	0.006
North Error (m)	MAX	0.024	0.029	0.025	0.021	0.031
	MEAN	0.006	0.004	0.007	0.005	0.004
	RMSE	0.008	0.009	0.010	0.008	0.007
East Error (m)	MAX	0.025	0.025	0.024	0.034	0.019
	MEAN	0.007	0.007	0.008	0.007	0.006
	RMSE	0.003	0.003	0.004	0.004	0.003
Down Error (m)	MAX	0.013	0.010	0.013	0.012	0.014
	MEAN	0.003	0.003	0.003	0.003	0.002

RMSE: root mean square error; MAX: maximum.

Table 3. Statistical results of the attitude errors of the visual solution.

Statistics		01	02	03	04	05
	RMSE	0.45	0.53	0.54	0.38	0.34
Roll Error (°)	MAX	1.45	1.58	1.75	1.47	1.14
	MEAN	0.37	0.42	0.41	0.31	0.27
	RMSE	0.37	0.49	0.55	0.39	0.35
Pitch Error (°)	MAX	1.12	1.33	1.76	1.52	1.05
	MEAN	0.29	0.41	0.45	0.30	0.28
	RMSE	0.05	0.08	0.09	0.05	0.05
Yaw Error (°)	MAX	0.15	0.28	0.29	0.13	0.20
	MEAN	0.04	0.05	0.06	0.04	0.04

°: degrees.

The tables show that the positioning error of the visual solution does not exceed 0.01 m in any of the three directions. It can be observed that the positioning results on horizontal direction are less accurate and more unstable than that on vertical direction. Given that the available satellite number was beyond 20 and the carrier-phase ambiguities were resolved to their integer values in our experiments, the accuracy of PPK smoothing positioning could be thought to be at the centimeter level [42]. Meanwhile, it is obvious that the noise influences of visual positioning and PPK positioning are independent of each other, hence the absolute positioning accuracy of the proposed visual solution is equivalent to that of the smoothed PPK.

For the attitude errors, the root mean square error (RMSE) of the roll angle was 0.45° (1σ), the pitch angle error was 0.43° (1σ), and the heading angle error was 0.1° (1σ). Since the corners of the marker are all on the horizontal plane (almost parallel to the p-frame), we can calculate the heading angle very accurately based on the position of the corresponding points on the image, but the roll and pitch estimations are less stable because of the lack of observations in the vertical direction.

The statistical results of the position and attitude errors of the visual/inertial fusion solution in five experiments are shown in Tables 4 and 5, respectively.

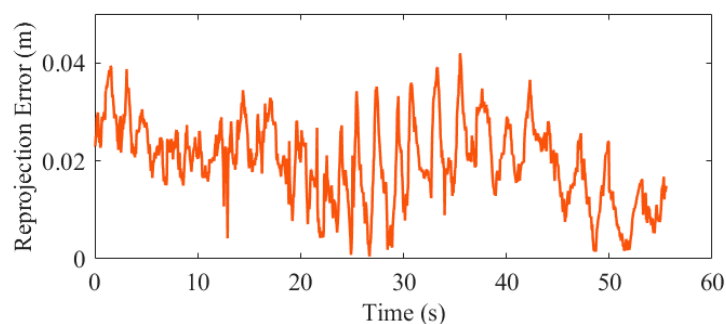**Table 4.** Statistical results of the position errors by the visual/inertial solution.

Statistics		01	02	03	04	05
	RMSE	0.007	0.006	0.009	0.007	0.006
North Error (m)	MAX	0.025	0.180	0.029	0.022	0.034
	MEAN	0.006	0.005	0.007	0.005	0.005
	RMSE	0.009	0.009	0.010	0.008	0.008
East Error (m)	MAX	0.033	0.031	0.025	0.034	0.023
	MEAN	0.007	0.007	0.008	0.007	0.006
	RMSE	0.004	0.004	0.005	0.004	0.004
Down Error (m)	MAX	0.013	0.010	0.015	0.010	0.012
	MEAN	0.003	0.003	0.003	0.004	0.003

The statistical results show that the positioning errors of the visual/inertial integrated solution are not more than 0.01 m in any of the three directions, which is consistent with the vision solution.

In addition, the reprojection errors was calculated to prove the pose accuracy. The corner points on the image that are not used for estimation (rejected by the RANSAC) were projected into the w-frame by the estimated pose, and the errors comparing to the real 3D corners on the horizontal direction were calculated. The results of the first test are shown in Figure 13, and the statistical results of the five experiments are shown in Table 6.

Table 5. Statistical results of the attitude errors by the visual/inertial solution.

Statistics		01	02	03	04	05
	RMSE	0.08	0.44	0.34	0.21	0.43
Roll Error (°)	MAX	0.24	1.66	1.50	0.88	1.86
	MEAN	0.06	0.24	0.20	0.17	0.28
	RMSE	0.17	0.14	0.24	0.27	0.17
Pitch Error (°)	MAX	0.83	0.58	0.80	1.03	0.70
	MEAN	0.11	0.08	0.19	0.21	0.12
	RMSE	0.06	0.13	0.07	0.05	0.03
Yaw Error (°)	MAX	0.13	0.36	0.32	0.12	0.15
	MEAN	0.05	0.10	0.05	0.04	0.02

Figure 13. Corner points reprojection errors on the horizontal plane in test 01.

Table 6. Statistical results of the corner reprojection error.

Statistics	01	02	03	04	05
RMSE (m)	0.022	0.023	0.021	0.020	0.024
MAX (m)	0.042	0.049	0.049	0.033	0.044
MEAN (m)	0.020	0.020	0.018	0.019	0.023

The statistical results show that the corner reprojection error is 0.022 m (1σ), which also verifies that the position accuracy of the proposed system is at the centimeter level.

For the attitude errors, it can be observed from the statistical results in Table 5 that the roll angle error is 0.30° (1σ), the pitch angle error is 0.20° (1σ), and the heading angle error is 0.07° (1σ). Comparing to the pure visual solution, the roll and pitch angle accuracies have increased by 33% and 54% on average, respectively. Additionally, Figure 14 shows that the pitch angle and roll angle estimated by the integrated navigation are more stable and smooth than those of the visual solution. Due to the excellent sensibility to vehicle motion of the IMU, the horizontal angle vibration caused by the maneuver is alleviated by the fusion with INS.
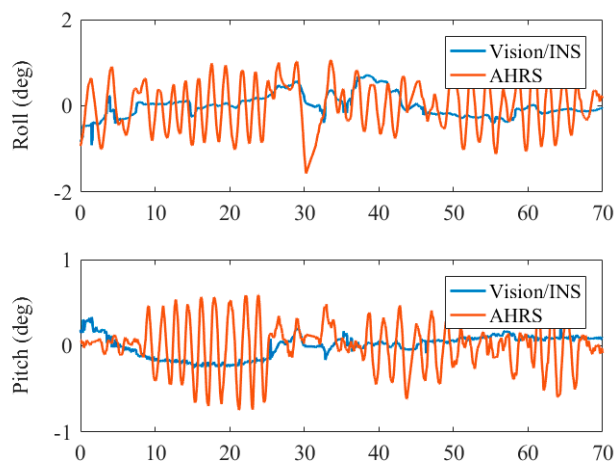
The attitude and heading reference systems (AHRS) [43,44] is widely used for continuous attitude estimation for flight control of UAVs since it is independent of GNSS. It exploits the IMU for pitch and roll angle estimation and magnetometer augmentation for improving heading stability. Because there was no magnetometer in our test device, we used the IMU data for AHRS and compared the accuracy and stability of the horizontal angles. The Mahony AHRS algorithm [45] was performed in the test. Figure 15 illustrates the roll and pitch errors of the two systems in the fourth test.

It can be observed that the roll and pitch angle estimation of the proposed system is much more accurate and stable than that of the AHRS. The AHRS provides pitch and roll angles relative to the earth gravity vector, thus, the accuracy of horizontal angles suffers from the vehicle acceleration since it cannot be separated by IMU only. In the tests, we held the platform to move in circles, therefore, the

direction-alternating acceleration resulted in the sinusoidal curve of the angles error of the AHRS. As for the horizontal angles estimated by the proposed system, the error of INS-derived position was related to the attitude error, hence the difference between the high-precision position obtained by marker detection and that of INS could correct the error of roll and pitch angles by the filter.
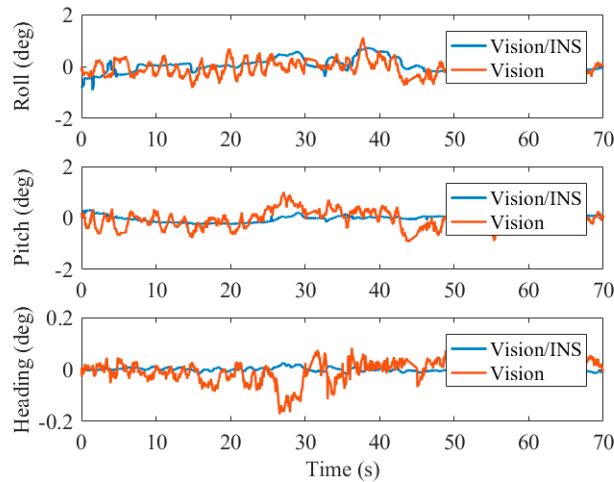
Figure 14. Comparison of roll and pitch errors by the visual solution and visual/inertial solution in test 04.

Figure 15. Comparison of roll and pitch errors by the attitude and heading reference system (AHRS) solution and visual/inertial solution.

As aforementioned, INS can provide continuous positioning results when the external aiding source is short-term unavailable. To illustrate that our system is robust to failure in marker extraction, we artificially interrupted the visual correction in data processing and calculated the errors of the positioning results. We set five outages for each test, and each outage was set to 5 s. Figures 16 and 17 are the position and reprojection errors of the simulation experiments in the fifth test, respectively.

The position error in the horizontal and vertical direction for the 25 outages in all the tests is 0.41 and 0.09 m (1σ), respectively, indicating that the proposed system can provide stable and reliable pose estimation when the marker is invalid occasionally. However, the navigation error of INS is accumulated with time as shown in Figure 17, hence long-term failure of marker detection will decay the performance.

To verify the robustness of the proposed system when the UAV is at a relative higher altitude or large-angle maneuver, we held the test platform 8 m away to capture the marker with no slant and with 40 degrees tilting, respectively. As the distance increases, the marker in the image becomes smaller and

the field of view becomes larger, hence objects of different shapes around the marker came into view which caused the debris in the figures. Only the polygon containing correct grayscale information would be matched. Figure 18 presents the contours extraction and marker detection results in the two situations. The positioning errors (RMS) on vertical direction for these two tests are 0.12 and 0.23 m, respectively. The errors on horizontal direction are 0.21 and 0.30 m, respectively. It can be observed that the long distance and large tilt angle did not affect the marker detection. However, the positioning error increased since the marker was small on the image and the geometric condition of corner points was poor. Because the precision landing of the UAV depends primarily on the pose accuracy when it is close to the landing pad, this positioning results at such a distance are enough for the UAV to approach the target.
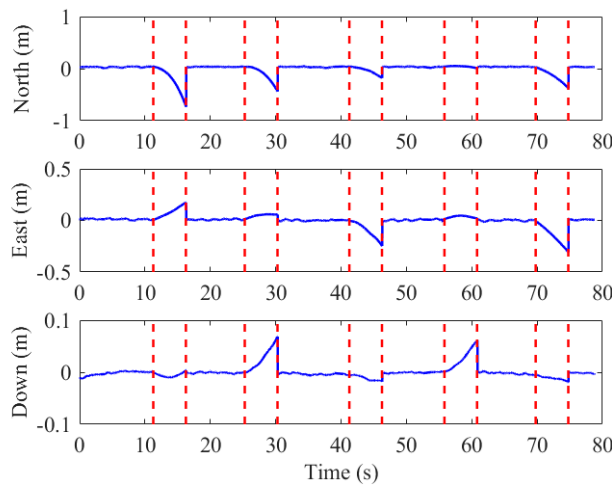
Figure 16. Position errors of the visual-inertial solution in the vision outage simulation test.
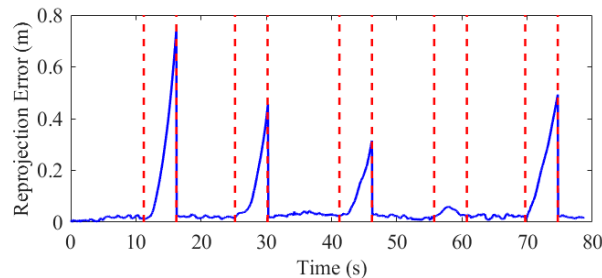
Figure 17. Reprojection error in the vision outage simulation test.
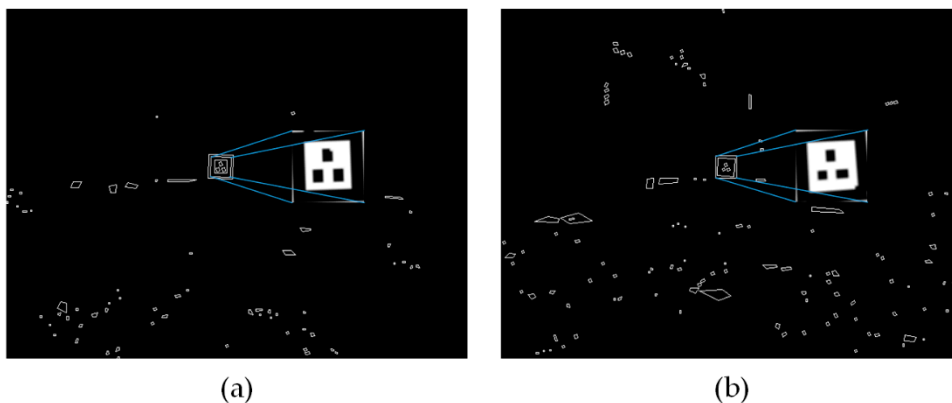
Figure 18. Edge extraction and marker detection results in challenging situations. (**a**) Detected marker at a distance of 8 m with no tilt; (**b**) detected marker at a distance of 8 m with tilt of 40 degrees.

Finally, we record the time spent on each module of the system over 1000 trials on the Tegra K1 developer kit (NVIDIA, Santa Clara, CA, USA), including a 4-Plus-1 ARM Cortex-A15 CPU. The results are shown in Table 7.

Table 7. Running time statistics of different modules of the system.

Module	MAX (ms)	MEAN (ms)
Marker detection	36.15	25.21
Pose estimation by marker	9.01	1.71
Filtering update	1.21	0.25
Total	70.34	32.26

It can be observed form the time consumption in Table 7 that the system runs efficiently. The most time-consuming part is marker detection, which requires a large number of pixel operations, while the other parts take relatively shorter times since they only need to process a small amount of low-dimension data. Overall, the average time consumption of the entire system is 32.26 ms, which is able to meet the real-time control requirements of UAVs.

4. Conclusions

In this paper, an artificial marker and MEMS IMU-based six-DoF pose estimation solution to meet the UAV landing requirements is proposed. The marker is specially designed for UAV landing due to its simplicity to generate and effectiveness to detect. An onboard monocular camera is used to recognize the marker and recover the absolute position and attitude of the UAV in the world frame. Then, this information is fused with IMU data by an EKF, thus our system still works in the case of visual failure, which is a fatal problem for previous systems. Improving upon the previous visual/inertial landing systems, the error components of the inertial sensor are modeled, and the horizontal angles are also estimated. Experiments show that the position accuracy is at centimeter level, the heading error is less than 0.1°, and the roll and pitch angle errors are less than 1° with the UAV at a height of about 1.4 m. The visual/inertial fused system improves the roll and pitch accuracy by 33% and 54%, respectively. Moreover, vision outage simulation tests show our system can provide relatively reliable position estimation (0.41 m drift in the horizontal direction and 0.09 m in the vertical direction in 5 s) when the maker occasionally occluded. Tests in challenging situations prove that the marker can be detected at a distance of 8 m and tilt of 40 degrees and the positioning accuracy is at decimeter level. The statistics of time consumption of the proposed system in an onboard computer show that it is able to meet the requirement of real-time control. The proposed system also applies to marker-based state estimation of indoor ground vehicles.

Future works will extend pose estimation to the entire route of UAV flight. Natural features are going to be used to estimate the UAV pose, when there is no marker in the field of vision. The marker can be used to correct the accumulated navigation error as long as it is detected.

References

1. Grenzdörffer, G.; Teichert, B. The photogrammetric potential of low-cost UAVs in forestry and agriculture. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *31*, 1207–1214.
2. Herwitz, S.; Johnson, L.F.; Dunagan, S.E.; Higgins, R.G.; Sullivan, D.; Zheng, V.J.; Lobitz, B.M.; Leung, J.G.; Gallmeyer, B.A.; Aoyagi, M.; et al. Imaging from an unmanned aerial vehicle: Agricultural surveillance and decision support. *Comput. Electron. Agric.* **2004**, *44*, 49–61. [CrossRef]
3. Waharte, S.; Trigoni, N. Supporting search and rescue operations with UAVs. In Proceedings of the International Conference on Emerging Security Technologies, Canterbury, UK, 6–7 September 2010; pp. 142–147.
4. Barmpounakis, E.N.; Vlahogianni, E.I.; Golias, J.C. Unmanned Aerial Aircraft Systems for transportation engineering: Current practice and future challenges. *Int. J. Transp. Sci. Technol.* **2016**, *5*, 111–122. [CrossRef]
5. Liu, Y.; Wang, Q.; Zhuang, Y.; Hu, H. A Novel trail detection and scene understanding framework for a quadrotor UAV with monocular vision. *IEEE Sens. J.* **2017**, *17*, 6778–6787. [CrossRef]
6. Merz, T.; Duranti, S.; Conte, G. Autonomous landing of an unmanned helicopter based on vision and inertial sensing. In *Experimental Robotics*, 9th ed.; Springer: Berlin, Germany, 2006; pp. 343–352.
7. Groves, P.D. *Principles of GNSS, Inertial, Multisensor Integrated Navigation Systems*; Artech House: Norwood, MA, USA, 2013.
8. Nemra, A.; Aouf, N. Robust INS/GPS sensor fusion for UAV localization using SDRE nonlinear filtering. *IEEE Sens. J.* **2010**, *10*, 789–798. [CrossRef]
9. Falco, G.; Gutiérrez, M.C.C.; Serna, E.P.; Zacchello, F.; Bories, S. Low-cost real-time tightly-coupled GNSS/INS navigation system based on carrier-phase double-differences for UAV applications. In Proceedings of the 27th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2014), Tampa, FL, USA, 8–12 September 2014; p. 841857.
10. Dabove, P.; Di Pietra, V. Single-baseline RTK positioning using dual-frequency GNSS receivers inside smartphones. *Sensors* **2019**, *19*, 4302. [CrossRef] [PubMed]
11. Robustelli, U.; Baiocchi, V.; Pugliano, G. Assessment of dual frequency GNSS observations from a Xiaomi Mi 8 Android smartphone and positioning performance analysis. *Electronics* **2019**, *8*, 91. [CrossRef]
12. Balamurugan, G.; Valarmathi, J.; Naidu, V.P.S. survey on UAV navigation in GPS denied environments. In Proceedings of the International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Odisha, India, 3–5 October 2016; pp. 198–204.
13. Shen, S.; Mulgaonkar, Y.; Michael, N.; Kumar, V. Vision-based state estimation for autonomous rotorcraft MAVs in complex environments. In Proceedings of the IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 1758–1764.
14. Gautam, A.; Saripalli, S.P.B.S. A survey of autonomous landing techniques for UAVs. In Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS), Orlando, FL, USA, 27–30 May 2014; pp. 1210–1218.
15. Fiala, M. ARTag, a fiducial marker system using digital techniques. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 590–596.
16. Olson, E. AprilTag: A robust and flexible visual fiducial system. In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3400–3407.
17. Garrido-Jurado, S.; Muñoz-Salinas, R.; Madrid-Cuevas, F.J.; Marín-Jiménez, M.J. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* **2014**, *47*, 2280–2292. [CrossRef]
18. Ling, K. Precision Landing of a Quadrotor UAV on a Moving Target Using Low-Cost Sensors. Master's Thesis, University of Waterloo, Waterloo, ON, Canada, September 2014.
19. Kyristsis, S.; Antonopoulos, A.; Chanialakis, T.; Stefanakis, E.; Linardos, C.; Tripolitsiotis, A.; Partsinevelos, P. Towards autonomous modular UAV missions: The detection, geo-location and landing paradigm. *Sensors* **2016**, *16*, 1844. [CrossRef] [PubMed]
20. Jayatilleke, L.; Zhang, N. Landmark-based localization for Unmanned Aerial Vehicles. In Proceedings of the IEEE International Systems Conference (SysCon), Orlando, FL, USA, 15–18 April 2013; pp. 448–451.

21. Millet, P.; Ready, B.; McLain, T. Vision-based precision landings of a tailsitter UAV. In Proceedings of the AIAA Guidance, Navigation, Control Conference, Chicago, IL, USA, 10–13 August 2009.

22. Pan, X.; Ma, D.; Jin, L.; Jiang, Z. Vision-Based Approach Angle and Height Estimation for UAV Landing. In Proceedings of the Congress on Image and Signal Processing, Hainan, China, 27–30 May 2008; pp. 801–805.

23. Acuna, R.; Willert, V. Dynamic Markers: UAV landing proof of concept. In Proceedings of the Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), Joao Pessoa, Brazil, 6–10 November 2018; pp. 496–502.

24. Yang, S.; Scherer, S.A.; Zell, A. An onboard monocular vision system for autonomous takeoff, hovering and landing of a micro aerial vehicle. *J. Intell. Robot. Syst.* **2013**, *69*, 499–515. [CrossRef]

25. Araar, O.; Aouf, N.; Vitanov, I. Vision based autonomous landing of multirotor UAV on moving platform. *J. Intell. Robot. Syst.* **2016**, *85*, 369–384. [CrossRef]

26. Lange, S.; Sunderhauf, N.; Protzel, P. A vision based onboard approach for landing and position control of an autonomous multirotor UAV in GPS-denied environments. In Proceedings of the International Conference on Advanced Robotics, Munich, Germany, 22–26 June 2009; pp. 1–6.

27. Nguyen, P.H.; Kim, K.W.; Lee, Y.W.; Park, K.R. Remote marker-based tracking for UAV landing using visible-light camera sensor. *Sensors* **2017**, *17*, 1987. [CrossRef] [PubMed]

28. Nguyen, P.H.; Arsalan, M.; Koo, J.H.; Naqvi, R.A.; Truong, N.Q.; Park, K.R. LightDenseYOLO: A fast and accurate marker tracker for autonomous UAV landing by visible light camera sensor on drone. *Sensors* **2018**, *18*, 1703. [CrossRef] [PubMed]

29. Benini, A.; Rutherford, M.J.; Valavanis, K.P. Real-time, GPU-based pose estimation of a UAV for autonomous takeoff and landing. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 3463–3470.

30. Otsu, N. A Threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]

31. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [CrossRef]

32. Shi, J.; Tomasi, C. Good features to track. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 593–600.

33. Bouguet, J.-Y. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corp.* **2001**, *5*, 4.

34. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]

35. Lepetit, V.; Moreno-Noguer, F.; Fua, P. EPnP: An accurate O(n) solution to the PnP problem. *Int. J. Comput. Vis.* **2009**, *81*, 155. [CrossRef]

36. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]

37. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In Proceedings of the International Workshop on Vision Algorithms, Corfu, Greece, 21–22 September 1999; pp. 298–372.

38. Agarwal, S.; Mierle, K. Ceres Solver. Available online: http://ceres-solver.org (accessed on 6 November 2019).

39. Shin, E.H. Estimation Techniques for Low-Cost Inertial Navigation. Ph.D. Thesis, Department of Geomatics Engineering, The University of Calgary, Calgary, AB, Canada, May 2005.

40. Shin, E.-H. Accuaracy Improvement of Low Cost INS/GPS for Land Applications. Master's Thesis, Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada, December 2001.

41. Furgale, P.; Rehder, J.; Siegwart, R. Unified temporal and spatial calibration for multi-sensor systems. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 1280–1286.

42. Teunissen, P.; Montenbruck, O. *Springer Handbook of Global Navigation Satellite Systems*; Springer: Berlin, Germany, 2017.

43. Geiger, W.; Bartholomeyczik, J.; Breng, U.; Gutmann, W.; Hafen, M.; Handrich, E.; Huber, M.; Jackle, A.; Kempfer, U.; Kopmann, H.; et al. MEMS IMU for AHRS applications. In Proceedings of the IEEE/ION Position, Location and Navigation Symposium, Portland, OR, USA, 20–23 April 2008; pp. 225–231.

44. Madgwick, S.O.H.; Harrison, A.J.L.; Vaidyanathan, R. Estimation of IMU and MARG orientation using a gradient descent algorithm. In Proceedings of the IEEE international conference on rehabilitation robotics, Toronto, ON, Canada, 24–28 June 2011; pp. 1–7.
45. Mahony, R.; Hamel, T.; Pflimlin, J. Nonlinear complementary filters on the special orthogonal group. *IEEE Trans. Autom. Control* **2008**, *53*, 1203–1218. [CrossRef]