Vision-Aided Inertial Navigation for Small Unmanned Aerial Vehicles in GPS-Denied Environments

Abstract This paper presents a vision-aided inertial navigation system for small unmanned aerial vehicles (UAVs) in GPS-denied environments. During visual estimation, image features in consecutive frames are detected and matched to estimate the motion of the vehicle with a homography-based approach. Afterwards, the visual measurement is fused with the output of an inertial measurement unit (IMU) by an indirect extended Kalman filter (EKF). A delay-based approach for the measurement update is developed to introduce the visual measurement into the fusion without state augmentation. This method supposes that the estimated error state is stable and invariant during the second half of one visual calculation period. Simulation results indicate that delay-based navigation can reduce the computational complexity by about 20% compared with general augmented Vision/INS (inertial navigation system) navigation, with almost the same estimate accuracy. Real experiments were also carried out to test the performance of the proposed navigation system by comparison with the augmented filter method and a referential GPS/INS navigation.

Keywords Small UAVs, Visual Navigation, Indirect EKF, Motion Estimation, Delay-Based Measurement Update

1. Introduction

Small UAVs typically use a low-cost IMU integrated with a GPS receiver to build a navigation system due to a limited on-board computational capability and payload capacity [1, 2]. Although the IMU is able to track sudden motions both in angular velocity and linear acceleration of a short duration, it is subject to a growing unbounded error caused by low frequency bias. The GPS can bind the accumulated error by estimating the absolute velocity and position of the vehicle directly. However, it is easily disturbed by the weather and the flight environment. Under these GPS-denied circumstances, general GPS/INS navigation is no longer reliable. Therefore, other sensors are required for autonomous flight.

Computer vision has played an important role in both the control and navigation of small UAVs over the past decade [3]. Visual measurement can be used in navigation [4], terrain following [5], obstacle avoidance [6] and urban canyon flight [7]. According to the selection of the camera, there are two main methods for visual applications: stereo and monocular. Stereovision systems have been successfully applied in small UAVs for years.
Map-based approaches for the visual navigation of small UAVs focus on the SLAM (simultaneously localization and mapping) algorithm [12, 13]. This estimates the state of the vehicle and builds a map at the same time. Another method using mosaics as environment representations is proposed in [10]. Optical flow (OF) is a typical mapless method for small/mini UAVs [14, 15]. This method runs at a very high frequency, but is limited to domains where the motion between consecutive frames is expected to be small. OF usually requires an inertial measurement to compensate for the effect of the rotation on the optical flow field. Visual odometry (VO) is considered as another popular mapless approach for the visual navigation of small UAVs [16]. In this system, the visual sensor is in charge of the position and orientation estimation by tracking ground template and the attitude is measured by the on-board IMU. A major defect of the original VO algorithm is the accumulated error of the state estimate, as it fails to distinguish the translation from the rotation without inertial compensation.

As to the above defect, methods that can estimate the motion of the vehicle in six degrees of freedom have also been applied in visual navigation. An EKF-based visual and inertial navigation system with an L-M-based nonlinear estimation method has been proposed [17, 18]. The state of the EKF was expanded by including the states of the vehicle at two consecutive moments. A method to combine inertial and visual measurements by an unscented Kalman filter (UKF) was developed in [19]. The state was augmented so that the homography estimated by the visual sensor could be directly used as the measurement of the filter. Moreover, a delay state EKF that allowed loose coupling of the inertial estimation and the VO-based visual measurement was described in [20]. The delayed state was created by appending the position and orientation of the vehicle at the last VO frame. Other methods have been proposed to build vision-aided inertial navigation systems and they also use state augmented approaches to introduce visual measurement into the filter [21, 22]. In the above systems, the visual motion estimate is usually taken as a relative measurement, so the most intuitive method is state augmentation. One major defect is that the augmented state requires extra calculations for the filter state and the state covariance.

This paper describes the whole process of the vision-aided inertial navigation for small UAVs, from feature detection to the fusion of inertial and visual measurements. A scale invariant feature transform (SIFT) approach is adopted to detect and describe the features from aerial images. Then the matched features in consecutive frames are used to estimate the motion of the vehicle in six degrees of freedom with a homography-based method. We propose a new approach to combine the inertial and visual measurements to estimate the state of the vehicle based on an indirect EKF. The main contribution of this paper is the derivation of a delay-based measurement update method for the EKF to fuse the visual and inertial measurements without any state augmentation. A simulation environment will be set up to compare the proposed and the general augmented methods by both computational complexity and estimate accuracy. The performance of this navigation system will also be verified by real experiments.

2. Homography-based visual navigation

The visual navigation that estimates the motion of the vehicle according to the visual information is described in this section. Stable features are detected and matched to estimate the homography between consecutive frames. Afterwards, the homography is decomposed into the translational and rotational motions of the vehicle.

2.1 Feature detection and match

In this paper, SIFT [23] is used for feature detection. This method searches for features not only in the geometry space, but also in the scale space established by the Gaussian difference of the image in different scales. A descriptor, which is invariant to rotation, translation and scale, is established to describe the corresponding feature. Since thousands of features might be extracted even from a small image, a fuzzy-based threshold adjustment method is proposed to stabilize the number of features from one image regardless of the change in the image sequence. This method can reduce the calculation of the image processing without affecting the accuracy of the posterior visual estimation. Further details of the proposed adaptive method are available in [24].

Matching the features in consecutive frames is implemented by the calculation of Euclidean distance between associated descriptors, as all the descriptors have been normalized in the process of feature detection. A fast nearest-neighbour algorithm is used to match the features quickly. A matched pair is accepted if and only if the distance between the two descriptors is the shortest, less than a threshold and shorter than 0.8 times the distance of the second nearest neighbour. A bi-directional match method is applied to improve the robustness of the match.

2.2 Robust homography calculation

Homography is modelled to indicate the transformation between two images, including scale, rotation and translation. The homography model is defined as follows:
where \([u_p, v_p, 1]^T\) and \([u_q, v_q, 1]^T\) are homogeneous positions for the corresponding points of two consecutive frames in the pixel coordinate, \(H = (h_{11} \ h_{12} \ h_{13}) \ \ h_{21} \ \ h_{22} \ \ h_{23} \ \ h_{31} \ \ h_{32} \ \ h_{33}\) is the homography and \(\lambda\) is a scale factor. The homography has eight degrees of freedom, as it is defined up to the scale factor.

A random sample consensus (RANSAC) approach is applied to eliminate the erroneous matches as not all the pairs are matched correctly. A bucketing technique [25] is adopted in the step of the subset selection in RANSAC to avoid a dramatically enlarged error of the estimate when features from one subset are too close. Singular value decomposition (SVD) is used to estimate the homography by the matched feature pairs that pass the RANSAC filter. An M-estimator algorithm is applied to add weights to the matched pair according to the match accuracy to improve the robustness of the estimate.

2.3 Homography-based motion estimation

Suppose a small UAV with an on-board downward-looking camera is flying above plane \(\Pi\), \(m_1\) and \(m_2\) are two projections in the camera coordinate of a fixed point \(p\) as shown in Figure 1. \(\tilde{m}_1\) and \(\tilde{m}_2\) are defined as the homogeneous points corresponding to the projections in the pixel coordinate. Their relationships are given as:

\[
\begin{align*}
    x_1 \tilde{m}_1 &= A m_1 \\
    x_2 \tilde{m}_2 &= A m_2 \\
    \lambda \tilde{m}_2 &= H \tilde{m}_1
\end{align*}
\]

where \(A\) is the camera intrinsic matrix, which can be calibrated accurately before the flight and \(z_1\) and \(z_2\) are the positions in the \(z\) direction of the two projected points in the camera coordinate.

Without a loss of generality, it is considered that the world coordinate is fixed to the camera coordinate of position 1. Then \(R_{12}\) is defined as the rotation matrix and \(t_{12}\) is the translation vector. They are both described in the camera coordinate of position 1 to express the motion of the vehicle. The relationship between the two projections of the fixed point \(p\) is expressed as:

\[
m_2 = R_{12}(m_1 + t_{12})
\]

Since point \(p\) is on plane \(\Pi\), the above equation is rewritten as:

\[
m_2 = R_{12}(I + \frac{t_{12}n^T}{d}) m_1
\]

where \(d\) is the Euclidean distance between position 1 and plane \(\Pi\), \(n\) is the unit normal vector of the plane and \(I\) is a \(3 \times 3\) identity matrix. Define the scale factor as \(\lambda = z_1/z_2\) and the combination of Equations (2), (3) and (5) gives:

\[
H = AR_{12}(I + \frac{t_{12}n^T}{d})A^{-1}
\]

The calibrated homography is defined as:

\[
H_c = A^{-1}HA = R_{12}(I + \frac{t_{12}n^T}{d})
\]

Equations (6) and (7) indicate the relationship between the homography and the camera motion. The rotation and the translation are obtained by the SVD of the calibrated homography. Further details on the calibrated homography decomposition are available in [26].

3. Fusion of inertial and visual estimates

In this section, the EKF is employed to establish a Vision/INS navigation system to improve the state estimation of the vehicle in GPS-denied environments. A continuous-discrete formulation is used in this filter. This means that the state estimation is propagated based on continuous nonlinear system dynamics, whereas the measurement is updated at discrete time iterations. The state update is propagated by the output of the IMU at 50Hz. The measurement update is driven each time when the visual computer finishes the motion estimation. A novel delay-based measurement update method is proposed, which can introduce visual measurement into the navigation state filter without any state augmentation.

3.1 State description

The system state of the filter contains 16 parameters and is defined as follows:

\[
X = (p^n, \nu^n, q, b_\alpha, b_\omega)^T
\]

where \(p^n = (x, y, z)\) is the position of the vehicle in the world coordinate, \(\nu^n = (\nu_x, \nu_y, \nu_z)\) is the velocity of the
vehicle in the world coordinate, which means the velocity in north, east and down directions respectively, \( q = (q_0, q_1, q_2, q_3) \) is the attitude in quaternions and \( b_a = (a_{bx}, a_{by}, a_{bz}) \) and \( b_g = (\omega_{bx}, \omega_{by}, \omega_{bz}) \) are the biases of the three-axis accelerometers and three-axis gyroscopes respectively.

The world coordinate is defined as a North, East and Down (NED) relative coordinate system with the origin at the initial location of the vehicle. The navigation coordinate is the same as the world coordinate in this paper. Quaternions are adopted to avoid the singularity, which may occur in Euler angles. The bias in the state can provide a continuous estimate of the current drift of IMU to compensate for the time-varying bias effects. Suppose that the installation of the IMU and GPS antenna has been compensated, the inertial navigation process model in continuous state-space form is written as:

\[
X(t) = f(X(t), U(t), t) + G(X(t), t)w(t) \tag{9}
\]

where

\[
G = \begin{bmatrix}
I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\
0_{3 \times 3} & -C_0^B(q) & 0_{3 \times 3} & 0_{3 \times 3} \\
o_{3 \times 3} & 0_{4 \times 3} & -I_2^B(q) & 0_{4 \times 3} \\
o_{3 \times 3} & o_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3}
\end{bmatrix}
\]

\[
f = \begin{bmatrix}
f_B^B(q)\bar{a}^b + g^b \\
\frac{1}{2}\bar{\Omega}(\bar{\omega})\bar{\omega} \\
o_{3 \times 1} \\
o_{3 \times 1}
\end{bmatrix}
\]

\[
\bar{a}^b = a_m^b - b_a, \quad \bar{\omega} = \omega_m^b - b_\omega
\]

Here \( I_{m \times m} \) means the \( m \) order identity matrix \( O_{m \times n} \) is the zero matrix with \( m \) rows and \( n \) columns, \( C_0^B \) is the direction cosine matrix (DCM) from the body coordinate to the navigation (world) coordinate, \( g^B \) is the gravity vector in the world coordinate, \( U(t) \) is the system input from the inertial sensors, \( a_m^b \) and \( \omega_m^b \) are the outputs of the accelerometers and gyroscopes respectively and \( w(t) \sim N(0, W) \) represents the process noise of the system.

The attitude equation in the form of a quaternion is written as:

\[
\dot{q} = \frac{1}{2} \Omega(\omega)b^q = \frac{1}{2} Z(q)b^\omega \tag{10}
\]

where:

\[
\Omega(\omega) = \begin{bmatrix}
0 & -\omega_3 & -\omega_2 & -\omega_1 \\
\omega_3 & 0 & -\omega_1 & \omega_2 \\
\omega_2 & \omega_1 & 0 & -\omega_3 \\
\omega_1 & \omega_2 & \omega_3 & 0
\end{bmatrix}
\]

\[
Z(q) = \begin{bmatrix}
-q_1 & -q_2 & -q_3 & 0 \\
q_0 & -q_3 & q_2 & 0 \\
q_3 & q_0 & -q_1 & 0 \\
-q_2 & q_1 & 0 & 0
\end{bmatrix}
\]

3.2 Error state propagation

The navigation state, as shown in Equation (10), should be linearized and discretized before the state propagation. The error state vector is applied in an indirect extended Kalman filter, which is defined as:

\[
\delta X(t) = X(t) - \hat{X}(t) = (\delta v^B, \delta q, \delta b_a, \delta b_\omega)^T
\]

where \( \hat{X}(t) \) is the estimate of the state. The state model is linearized by the first order Taylor expansion and discretized as:

\[
\delta \dot{X}(t) = F(t)\delta X(t) + G(t)w(t) \tag{12}
\]

\[
\delta X_k = \Phi_{k,k-1}\delta X_{k-1} + W_{k-1} \tag{13}
\]

where:

\[
F(t) = \frac{\partial f(X(t))}{\partial X(t)} \bigg|_{X(t)=X(t)}
\]

\[
\Phi_{k,k-1} \approx I + F(t_{k-1})\Delta t = I + F_{k-1}\Delta t
\]

and \( \Delta t \) is the sampling time interval of the IMU. The covariance matrix of the equivalent white noise sequence \( W_k \) is derived as:

\[
Q_k = \bar{Q}T + (F_{k,k-1}\bar{Q}F_{k,k-1}^T)\Delta t^2/2
\]

\[
\approx (\bar{Q} + \Phi_{k,k-1}\bar{Q}\Phi_{k,k-1}^T)\Delta t/2
\]

where \( \bar{Q} = GWG^T \). Then the error state-based propagation of the discretized nonlinear extended Kalman filter is derived as:

\[
\tilde{X}_{k,k-1} = \tilde{X}_{k-1} + f(\tilde{X}_{k-1})\Delta t \tag{15}
\]

\[
p_{k,k-1} = \tilde{p}_{k,k-1} + F_{k,k-1}\tilde{p}_{k,k-1}^T + Q_{k-1} \tag{16}
\]

3.3 Visual measurement update

The direct output of the visual process is the rotation matrix \( R_{12} \) and the translation vector \( t_{12} \) of the camera positions between two consecutive frames in the camera coordinates. As a relative measurement, the visual motion estimation is usually introduced into the system filter by the state augmentation. The augmented state of the filter contains the states of the vehicle at both of the two positions where the consecutive frames are taken and calculated. The visual measurement is treated as the motion between these two states of the vehicle. The major defect of this method is the unavoidable extra calculation required by the state augmentation, which has a great effect on the on-board navigation of small UAVs.

A novel measurement update method is proposed in this section. If the time interval of the visual process and the coordinate conversion between the camera and the vehicle are known, the transformation is converted to the motion of the vehicle in the body coordinate. Then, with
the help of the DCM from the body coordinate to the navigation coordinate provided by the filter estimate at position 1, the attitude, displacement and velocity can be obtained. Since the velocity measurement requires a delay process, the visual update is divided into two parts: attitude and position update as well as velocity update.

3.3.1 Attitude and position measurement update

The DCM from the body coordinate at position 2 to the navigation coordinate is derived as:

\[ C^n_{b2} = C^n_{b1}C^b_2R_{t2}C^c_2 \]

(17)

where \( C^n_{b1} \) is the DCM from the body coordinate at position 1 to the navigation coordinate, which can be calculated from the state of the EKF filter at that moment, \( C^b_2 \) is the transformation matrix from the camera coordinate to the body coordinate and \( C^c_2 \) is the transformation matrix from the body coordinate to the camera coordinate. Then the attitude of the vehicle \( q \) in the form of quaternion can be calculated from \( C^n_{b2} \).

The displacement of the UAV from position 1 to position 2 in the navigation coordinate is derived as:

\[ \Delta p_{12} = C^n_{b1}C^b_2R_{t2} \]

(18)

Suppose the position of the vehicle \( p^n_1 \) at position 1 has been estimated, then the position at position 2 is given as:

\[ p^n_2 = p^n_1 + \Delta p_{12} \]

(19)

The attitude and position calculated from the visual observation are the measurements for the undelayed filter. The measurement of the filter with a discrete model is written as:

\[ Z_{1,k} = H_1X_k + V_{1,k} \]

(20)

where

\[ Z_{1,k} = (p^n, q)' \quad V_{1,k} \sim N(0, R_{1,k}) \]

\[ H_1 = \begin{bmatrix} I_{3x3} & 0_{3x3} & 0_{3x4} & 0_{3x3} \end{bmatrix} \]

The error state-based attitude measurement update is derived as follows:

\[ K_k = P_{k,k-1}H_1^T[H_1P_{k,k-1}H_1^T + R_{1,k}]^{-1} \]

(21)

\[ \delta \tilde{X}_k = K_k(Z_{1,k} - H_1\tilde{X}_{k,k-1}) \]

(22)

\[ \tilde{X}_k = \tilde{X}_{k,k-1} + \delta \tilde{X}_k \]

(23)

\[ P_k = (I - K_kH_1)P_{k,k-1}(I - K_kH_1)^T + K_kR_{1,k}K_k^T \]

(24)

3.3.2 Velocity measurement update

Suppose an image is taken and calculated at \( t_{k-m} \) and another consecutive image is dealt after \( m \) steps of the state propagation at \( t_k \). The average velocity in the navigation coordinate during \( t_{k-m} \) and \( t_k \) is derived as:

\[ \bar{v}^n_{k-m,k} = C^n_{b1}C^b_2R_{t2}t_{12}/t_m \]

(25)

with \( t_m = t_k - t_{k-m} \). Since the velocity in the filter state is the instantaneous velocity, the above average velocity cannot be introduced into the filter directly. Thus, there is an assumption that during a short time interval in the flight, the velocity cannot fluctuate remarkably. Then, the average velocity during \( t_{k-m} \) and \( t_k \) can be considered as the instantaneous velocity at \( t_{k-m/2} \), as:

\[ v^n_{k-m/2} \approx \bar{v}^n_{k-m,k} \]

(26)

Figure 2 illustrates the relationship between the average velocity and the instantaneous velocity. In this example, the real velocity of the vehicle increases gradually. It can be seen that the above assumption is reliable and reasonable. The difference between the average velocity and the instantaneous velocity at \( t_{k-m/2} \) is much smaller than that at \( t_{k-m} \) and \( t_k \). This assumption might be not so accurate under some extreme circumstances, but it is not the focus of this paper. We suppose that the UAV is in a stable flight and the above assumption is always tenable.

The delay-based measurement model for the velocity estimation is written as:

\[ \bar{v}^n_{k-m,k} = C^n_{b1}C^b_2R_{t2}t_{12}/t_m \]

(27)
where:
\[ Z_{2k} = v^a \quad V_{2k} \sim N(0, R_{2k}) \]
\[ H_2 = [0_{3 \times 3} \quad I_{3 \times 4} \quad 0_{3 \times 3} \quad 0_{3 \times 3}] \]

The process of this visual velocity measurement update is shown in Figure 3. It consists of two steps: the delay error state estimation and the error state correction. The velocity measurement at time \( t_k \) is introduced into the error state estimation at time \( t_n \) with \( n = k - m/2 \). Then the estimated error state at time \( t_n \) is used to correct the state at time \( t_k \). The measurement update equations are as follows:

\[ K_n = P_{n,n-1}H_2^T(H_2P_{n,n-1}H_2^T + R_{2,n})^{-1} \]  \( \text{(28)} \)
\[ \delta \hat{X}_n = K_n(Z_{2k} - H_2 \hat{X}_{n,n-1}) \]  \( \text{(29)} \)
\[ \hat{X}_k = \hat{X}_{k,k-1} + \delta \hat{X}_n \]  \( \text{(30)} \)
\[ P_k = (I - K_nH_2)P_{k,k-1}(I - K_nH_2)^T + K_nR_{2,k}K_n^T \]  \( \text{(31)} \)

Figure 3. The process of the delay measurement update

4. Simulations

4.1 Simulation environment setup

A simulation environment is set up to verify the proposed navigation system. It is an image-in-loop simulation system with a satellite image taken from Google Earth as the whole terrain map during the flight. Then the ground is presumed to be planar and the origin of the world coordinate is set on the centre of the image. The intrinsic matrix of the “on-board camera” is designed based on the parameters of the real camera in experiments. The size of the aerial image taken by the camera is \( 300 \times 300 \) in the pixel coordinate. The aerial image is considered as a sub-image of the whole map. The sub image taken by the vehicle at 30m high and 0° of the roll, pitch and yaw are obtained directly by the segmentation of the satellite image, whereas those taken from other positions or at other attitudes are obtained by the image interpolation of the map. The simulation environment is shown in Figure 4.

The flowchart of the simulation is shown in Figure 5. At first, a trajectory is designed as the input into a navigation state generator. The outputs of the generator, such as velocity and attitude, are used to obtain a sub-image from the whole map at this moment. Then features are detected and matched with those in the last simulated image and the corresponding features are used to estimate the homography robustly. The motion of the vehicle is estimated with the homography and the relative distance provided by the state generator. Finally, the visual measurement is fused with the inertial output, which is also simulated by the navigation state generator, by the delay-based EKF. The output of the navigation is compared with the original result provided by the navigation state generator.

Figure 4. The simulation environment

Figure 5. The simulation flowchart

The parameters of the sensors are simulated according to the characteristics of the real ones in the experiment. The uncertainty of the image is one pixel. The noise intensity of the gyrosopes is set 0.3°/s with the bias of 0.3°/s. The noise intensity of the accelerometers is 0.1m/s² with a bias of 0.3m/s². The measurement is obtained by the homography decomposition and the conversion from the camera coordinate to the navigation coordinate, which is a rather complex process. Therefore in this paper, the measurement covariance is determined by a statistic approach in the simulation. Suppose that the state of the small UAV (position, velocity and attitude) was estimated accurately at an instant, the state at the next instant can be obtained by both the state generator, as shown in Figure 5 and the visual measurement from the decomposition of the calibrated homography can be calculated from consecutive images. The difference between these two methods is recorded and this process is repeated many times at different positions, velocities or attitude. The data are used for the determination of the variance for each item in the
measurement vector. The measurement covariance is set as $R_1 = \text{diag}(3,3,3,2,2,2)$ and $R_2 = \text{diag}(3,3)$ in the simulation. For the real experiment, the measurement covariance is first set to be the same as the simulation. After the real flight, the acquired sensor data are analysed offline and the measurement covariance is optimized to get a better performance from the filter.

The initialization of the bias in the inertial sensors is another important element in the filter. In the simulation, the biases are initialized by the following approach. The UAV is put on the ground without any movement. The visual measurement is replaced by information on the changeless position, zero velocity and changeless attitude. Then, after a few seconds, the bias of the inertial sensors will be well estimated. Figure 6 and 7 show the initialized process of the three-axis gyroscopes and the three-axis accelerometers. Both these kinds of bias can be initialized accurately.

4.2 The effect of the measurement delay on the filter

Firstly, the effect of the measurement delay on the EKF filter is analysed. In this filter, the attitude estimated by the camera is a real-time measurement, while the velocity measurement is set with a different delay, from 0 to 2 seconds with an interval of 0.2 seconds. The ratio of the standard deviation of the velocity estimate after the stabilization is calculated for the analysis of the delay effect on the filter. The ratio is defined as $r(t) = \text{std}(t)/\text{std}(0)$, where $\text{std}(t)$ is the standard deviation with a delay of $t$ seconds and $r(0) = 1$. The simulation runs several times at each delay step and the average is taken as the final result.

Figure 8 shows the delay effect on the velocity estimation. It can be seen that the standard deviations of $v_x$ and $v_y$ increase linearly with the measurement delay. When the delay reaches 2 seconds, the standard deviation is about four times larger than that estimated by the non-delayed measurement. Our on-board avionics run at about 2Hz for the visual measurement. It can be concluded from this figure that a measurement delay of 0.5 seconds results in an increase in the standard deviation of about 50%. The delay of the visual measurement has a non-negligible effect on the estimate accuracy of the filter. The delay effect on the estimate of the velocity in the down direction is not as evident as others. One reason is that $v_z$ does not change as much in the desired trajectory, compared with the velocities in the horizontal directions. Hence, there is no clear difference between the delayed and non-delayed measurements.

4.3 The effect of the time interval of the visual process

In the velocity measurement update, the time interval $t_m$ between two visual processes is considered as an important factor to the filter. The frequency of the measurement updates is determined by the time interval. For example, if $t_m = 0.5$ seconds, the measurement update runs at 2Hz. Moreover, the time interval affects the assumption in Equation (26). If $t_m$ is too long, this assumption might be invalid. So another simulation is carried out to test the effect of the time interval. Two kinds of measurements are simulated: one is the velocity generated under the assumption at Equation (26); another is the “real” instantaneous velocity. The output of the system is defined as $r(t) = \text{std}_a(t)/\text{std}_d(t)$, where
$std_v(t)$ and $std_w(t)$ are the standard deviations of the velocity estimates of the above two kinds respectively, with $t$ as the time interval.

![Figure 9. The effect of the time interval](image)

The effect of the time interval on the velocity estimation is shown in Figure 9. The time interval changes from 0.1 to 1.9 seconds, but the standard deviation does not increase accordingly. The biggest is only 1.24 times the referential, when $t_m$ equals 1.7 seconds. The time interval has some effect on the filter by the frequency of the measurement update. However, the velocity estimate is stable in spite of the change in the time interval. This means that the assumption is reasonable in this application and the velocity calculated by our proposed method can be used in the measurement update of the filter.

### 4.4 Comparison with the augmented navigation

The last simulation test is about the comparison of the proposed navigation and the general augmented system. Both of them take the homography-based method, which is described in Section 2, to obtain the visual measurement. In the augmented filter, the state is expanded by including the state of the vehicle when the previous visual measurement is calculated. The augmented EKF is built mainly according to [17]. Matrix partition is applied for the optimization of the augmented filter calculation.

![Figure 10. The comparison of the attitude errors in simulation](image)

![Figure 11. The comparison of the velocity errors in simulation](image)

The error comparisons of the attitude and the velocity are shown in Figure 10 and Figure 11, respectively. Both the methods can limit the attitude and velocity errors within $(−1^\circ, 1^\circ)$ and $(−1\text{m/s}, 1\text{m/s})$. They have similar performances after the filters are stable. Figure 12 shows the comparison of the position error. The lines labelled “integrated” are the differences between the real position and the position calculated by the estimated velocity integration of the proposed filter. The correction of the proposed filter can be seen by comparing it with the integrated. The error of the position estimate of the proposed method is generally smaller than that of the velocity integration. Some drift is expected in the position estimation by these mapless methods. However, the drift has been remarkably reduced compared to approaches using inertial sensors alone.

![Figure 12. The comparison of the position errors in simulation](image)
realized by Matlab on a laptop (Window7 32-bit, Intel(R) Core(TM) i7-2670QM, 2.20GHz, 4GB RAM). Since the image processing is completed in advance, the recorded time is only for the whole process of the filter (6187 loops for the state propagation). Despite the fact that the matrix partition is applied, the average computing time of the augmented method is still about 1.702 seconds. The proposed method requires only 1.423 seconds for the whole calculation. This means that our method can reduce the computational complexity by about 20%, while keeping almost the same performance as the augmented approach.

![Figure 13. The experimental helicopter test bed](image)

| Standard deviation | Augmented | Proposed |
|--------------------|-----------|----------|
| Time(s)            | 1.702     | 1.423    |
| x(m)               | 2.767     | 1.634    |
| y(m)               | 2.745     | 1.198    |
| z(m)               | 0.735     | 1.224    |
| vx(m/s)            | 0.325     | 0.314    |
| vy(m/s)            | 0.318     | 0.286    |
| vz(m/s)            | 0.102     | 0.082    |
| roll(°)            | 0.287     | 0.229    |
| pitch(°)           | 0.212     | 0.215    |
| yaw(°)             | 0.839     | 0.594    |

Table 1. The comparison of two navigation systems

5. Experiments

5.1 System description

The experimental test bed helicopter is shown in Figure 13. The on-board avionics includes an R20 differential GPS receiver, an ADIS16355 IMU, an MS5540 barometric altimeter, a three-axis magnetometer that is composed of HMC1052 and HMC1041, an MV-1300UC colour industrial digital camera and four on-board computers. DSP TI28335 runs the navigation filter and fuses the measurements of GPS, IMU and other sensors. ARM LPC1766 is equipped to guide and control the small UAV. FPGA EP2C is in charge of the flight data recording and the control switch between a human pilot and autopilot. PC-104 PCM362 is used as the vision computer to receive the sensor data from the FPGA and record the aerial images simultaneously. The on-board equipment is shown in Figure 14.

![Figure 14. The on-board avionics](image)

The proposed visual navigation system consists of three main parts: image processing, homography-based motion estimation and sensor fusion. Figure 15 shows a flowchart of the whole system. In image processing, SIFT is applied to detect the features from consecutive frames, with a fuzzy controller to adjust the threshold online. Then the motion of the vehicle is estimated by a homography model of corresponding features. Afterwards, the motion is transferred into the attitude and velocity of the vehicle and fused with the inertial estimation by an indirect EKF with a delay-based measurement update.

![Figure 15. The system flowchart](image)

5.2 Test of the vision-aided inertial navigation system

An experiment is carried out to test the performance of the proposed navigation system with the above test bed. A flat wasteland is chosen as the experimental field. The relative height, which is used to recover the translational motion of the vehicle, is estimated by the output of the barometric sensor compensated by the initial height measurement on the ground. This measurement method is reasonable since the image scene is supposed to be a ground plane during the flight. The relative height can also be measured by adding other sensors, such as ultrasonic or laser rangefinders. However, the discussion of the relative distance estimate methods is not the focus...
of this paper. During the experiment, a skilled human pilot operates the helicopter manually. The sensor data are recorded on-board and analysed offline. The data is acquired at a rate of 50Hz for the inertial sensors and 2Hz for the visual sensor. The performance of the proposed Vision/INS navigation system is evaluated by comparing it with the augmented and referential GPS/INS navigation.

The attitude comparison of the whole flight is shown in Figure 16. The visual measurement does not work during the take-off and landing, in which periods the visual estimation is not accurate enough due to the rapid change of the aerial images. Thus, the visual measurement takes effect from the 16th second to 107th second in this figure. The GPS signal is cut off when the visual measurement is available. The augmented filter is also included in this comparison. The error here means the difference between the estimate and the output of the GPS/INS navigation. Figure 17 and 18 illustrate the comparisons of velocity and position estimates in the navigation coordinates.

The effectiveness of the proposed Vision/INS method has been verified by this experiment. In the proposed system, the visual measurement is applied to fuse with the inertial sensors for the state estimation of the vehicle when GPS is unavailable. This filter is able to provide stable and reliable estimates of the attitude and velocity with high accuracy, while the position estimate is subject to drift. This is a well-known drawback for any mapless method. In this filter, the drift has been greatly limited by the visual measurement. The average error and standard deviation of the proposed method are shown in Table 2.

### Table 2. The average error and the standard deviation of the proposed Vision/INS navigation system

| Parameter | Average error | Standard deviation |
|-----------|---------------|-------------------|
| x(m)      | 7.055         | 3.385             |
| y(m)      | 6.807         | 2.774             |
| z(m)      | 7.718         | 3.635             |
| vx(m/s)   | 0.467         | 0.626             |
| vy(m/s)   | 0.590         | 0.698             |
| vz(m/s)   | 0.373         | 0.257             |
| roll(°)   | 0.665         | 0.807             |
| pitch(°)  | 0.696         | 0.760             |
| yaw(°)    | 0.809         | 0.740             |

6. Conclusion and future work

This paper presented a vision-aided inertial navigation system for small UAVs in GPS-denied environments. A delay-based measurement update method was proposed to introduce the homography-based visual measurement into the indirect EKF without any state augmentation. The effects of the delay measurement and the time interval of the visual process on the filter have been analysed by simulations. The simulation results indicated that the proposed delay-based method can reduce the computational complexity by about 20%, compared with the general augmented filter, with almost the same estimate accuracies. The effectiveness of the proposed
method has been verified by an experiment. The navigation system developed in this paper achieves its good performance by comparing with the referential GPS/INS navigation for small UAVs. In summary, the proposed vision-aided inertial navigation method can provide a reliable state estimate for small UAVs in GPS-denied environments.

Future work will be carried out in two aspects. Firstly, other denied environments. Secondly, as well as the attitude and the velocity, the position of the vehicle can also be estimated by the proposed navigation but suffers a drift. This drift is mainly due to the accumulation of the error from the velocity estimation. Additional observations are required if the full autonomy of the UAV in GPS-denied environment is considered. In the future, a monocular SLAM system will be developed based on this system to improve the position estimation and the autonomy of small UAVs.

7. Acknowledgments

This work was supported by the National High-Tech Research & Development Program of China (Grant No. 2011AA040202). The authors would like to thank Yi Zhou, Chenghao Xue, Cong Wang and Qingru Zeng for their great help during the experiments.

8. References

[1] Kingston D B, Beard A W (2004) Real-time Attitude and Position Estimation for Small UAVs using Low-cost Sensors. AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop and Exhibit. Chicago, USA. 6488-6496.
[2] Oh S M, Johnson E N (2006) Development of UAV Navigation System based on Unscented Kalman Filter. Proceedings of the AIAA Guidance, Navigation, and Control Conference. USA. 1-21.
[3] Ludington B, Johnson E, Vachtsevanos G (2006) Augmenting UAV Autonomy. IEEE Robot Autom Mag. 13: 63-71.
[4] Kaiser M K, Gans N R, Dixon W E (2010) Vision-Based Estimation for Guidance, Navigation, and Control of An Aerial Vehicle. IEEE T Aero Elec Sys. 46: 1064-1077.
[5] Garratt M A, Chahl J S (2008) Vision-based terrain Following for An Unmanned Rotorcraft. J Field Robot. 25: 284-301.
[6] Green W E, Oh P Y (2008) Optic-Flow-Based Collision Avoidance. IEEE Robot Autom Mag. 15: 96-103.
[7] Conroy J, Gremillion G, Ranganathan B, et al. (2009) Implementation of Wide-Field Integration of Optic Flow for Autonomous Quadrotor Navigation. Auton Robot. 27: 189-198.
[8] Z. Yu, K. Nonami, J. Shin, and D. Celestino (2007) 3D Vision Based Landing Control of a Small Scale Autonomous Helicopter. Int. J. Adv. Robot. Syst. 4:51-56.
[9] Kelly J, Saripalli S, Sukhatme G S (2007) Combined Visual and Inertial Navigation for An Unmanned Aerial Vehicle. Proceedings of the International Conference on Field and Service Robotics. Chamonix, France.
[10] Caballero F, Merino L, Ferruz J, et al. (2009) Unmanned Aerial Vehicle Localization Based on Monocular Vision and Online Mosaicking. J. Intell. Robot. Syst. 55: 323-343.
[11] Weiss S, Siegwart R. (2011) Real-time metric state estimation for modular vision-inertial systems. Proceedings of IEEE International Conference on Robotics and Automation. Shanghai, China. 4531-4537.
[12] Pinies P, Lupton T, Sukkarieh S, et al. (2007) Inertial Aiding of Inverse Depth SLAM using a Monocular Camera. Proceedings of IEEE International Conference on Robotics and Automation. Roma, Italy. 2797-2802.
[13] Achtelik M W, Lynen S, Weiss S, et al. (2012) Visual-inertial SLAM for a small helicopter in large outdoor environments. IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura, Algarve, Portugal. 2651-2652.
[14] Kendoul F, Nonami K, Fantoni I, et al. (2009) An Adaptive Vision-Based Autopilot for Mini Flying Machines Guidance, Navigation and Control. Auton Robot. 27: 165-188.
[15] Kendoul F, Fantoni I, Nonami K (2009) Optic Flow-Based Vision System for Autonomous 3D Localization and Control of Small Aerial Vehicles. Robot. Auton. Syst. 57: 591-602.
[16] Amidt O, Kanade T, Fujita K (1999) A Visual Odometer for Autonomous Helicopter Flight. Robot Auton Syst. 28: 185-193.
[17] Roumeliotis S I, Johnson A E, Montgomery J F (2002) Augmenting Inertial Navigation with Image-based Motion Estimation. IEEE International Conference on Robotics and Automation. Washington DC, USA. 4326-4333.
[18] Johnson A E, Matthis H L (1999) Precise Image-based Motion Estimation for Autonomous Small Body Exploration. Artificial Intelligence, Robotics and Automation in Space, Proceedings of the Fifth International Symposium. Noordwijk, Netherlands. 627-634.
[19] Andersen E D, Taylor C N (2007) Improving MAV Pose Estimation using Visual Information. IEEE/RSJ International Conference on Intelligent Robots and Systems. San Diego, USA. 3751-3756.
[20] Tardif J P, George M, Laverne M, et al. (2010) A New Approach to Vision-Aided Inertial Navigation. IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, Taiwan. 4161-4168.
[21] Voigt R, Nikolic J, Hurezler C, et al. (2011) Robust Embedded Egomotion Estimation. IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA. 2694-2699.
[22] Mourikis A I, Roumeliotis S I (2007) A Multi-State Constraint Kalman Filter for Vision-Aided Inertial Navigation. Proceedings of IEEE International Conference on Robotics and Automation. Roma, Italy. 3565-3572.

[23] Lowe D G (2004) Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision. 2: 91-110.

[24] Wang C L, Wang T M, Liang J H, et al. (2012) A Fuzzy-Based Threshold Method Applied in SIFT for Visual Navigation of Small UAVs. IEEE Conference on Industrial Electronics and Applications. Singapore. 807-812.

[25] Zhang Z (1997) Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting. Image Vision Comput. 15: 59-76.

[26] Tsai R, Huang T, Wei-Le Z. (1982) Estimating three-dimensional motion parameters of a rigid planar patch, II: Singular value decomposition. IEEE Trans. Acoust. Speech Signal Process. 30: 525-534.