Bearing-only Visual SLAM for Small Unmanned Aerial Vehicles in GPS-denied Environments

Chao-Lei Wang   Tian-Miao Wang   Jian-Hong Liang   Yi-Cheng Zhang   Yi Zhou

Robotics Institute, Beijing University of Aeronautics and Astronautics, Beijing 100191, China

Abstract: This paper presents a hierarchical simultaneous localization and mapping (SLAM) system for a small unmanned aerial vehicle (UAV) using the output of an inertial measurement unit (IMU) and the bearing-only observations from an onboard monocular camera. A homography based approach is used to calculate the motion of the vehicle in 6 degrees of freedom by image feature match. This visual measurement is fused with the inertial outputs by an indirect extended Kalman filter (EKF) for attitude and velocity estimation. Then, another EKF is employed to estimate the position of the vehicle and the locations of the features in the map. Both simulations and experiments are carried out to test the performance of the proposed system. The result of the comparison with the referential global positioning system/inertial navigation system (GPS/INS) navigation indicates that the proposed SLAM can provide reliable and stable state estimation for small UAVs in GPS-denied environments.

Keywords: Visual simultaneous localization and mapping (SLAM), bearing-only observation, inertial measurement unit, small unmanned aerial vehicles (UAVs), GPS-denied environment.

1   Introduction

The navigation system for small unmanned aerial vehicles (UAVs) is typically composed of low-cost inertial sensors and a global positioning system (GPS) receiver due to the limited onboard computational capability and payload capacity[1]. The GPS can bound the accumulated error of the inertial sensors by estimating the absolute velocity and position of the vehicle. However, in many aerial applications, the UAV is required to perform a task within the GPS-defined environment, such as indoors, urban canyons. GPS is also easy to be disturbed by the weather during the flight. For the autonomy of small UAVs in GPS-denied environments, other sensors are required to be fused with the inertial measurement unit (IMU) for determination of the vehicle state without any priori information of the flight environment. This is known as the simultaneous localization and mapping (SLAM) system, with which the vehicle can build an online map and estimate its location in 6 degrees of freedom in that map.

Vision seems a good alternative to the GPS for the autonomous navigation of small UAVs in terms of weight, cost and information. Visual sensors have been used in surveillance tasks of UAVs for years[2]. They have also played an important role in autonomous navigation and control of small UAVs[3]. Performing the SLAM with the visual measurement has received a lot of attention over the past few years[4, 5]. A visual 3-D SLAM was developed for UAVs in partially structured environments[6]. In this algorithm, the vehicle is modeled as a rigid body with uniform motion and the acceleration is considered as the system noise. Another proposed visual SLAM system takes only natural landmarks as observations[7]. In this system, a homography

based approach is used to estimate the motion of the UAV and a novel landmark initialization method is developed.

The fusion of IMU and visual sensors is usually used in autonomous robots[8−10]. In these implementations, the inertial measurement is usually used as the input of the process model in the filter. A robust inertial SLAM algorithm using bearing-only observations was developed[11,12]. This extended Kalman filter (EKF) based SLAM is able to estimate the location of the vehicle in 6 degrees of freedom and 3-D positions of the features in the environment. Monocular vision is usually taken as the visual sensor for the SLAM of small UAVs. For these bearing-only sensors, the initialization of features in 3 dimensions is considered as a difficult problem. Davison et al.[13] showed a delayed initialization method for bearing-only SLAM. It waits until the observations have enough parallax to determine the 3-D position of a feature and include it into the filter. An undelayed initialization method using the inverse depth of features relative to the camera locations from which they are first viewed was proposed[14]. In this algorithm, once the depth estimation of a feature is sufficiently accurate, the inverse depth form can be converted into the Euclidean form safely. The absolute scale of the monocular SLAM can also be estimated by fusion of inertial and visual measurements[15]. Besides, the observability and consistency of the EKF based SLAM were analyzed[16−18]. The research of the airborne visual SLAM also involves the multi-UAV SLAM[19] and visual SLAM for indoor aerial vehicle[20].

In this paper, we will provide a monocular visual SLAM for a small UAV in GPS-denied environments. A hierarchical structure is employed in this system. A homography based method is adopted to estimate the motion of the vehicle by a single onboard camera. Then, the velocity and the attitude are estimated by the fusion of the inertial and visual measurements. Another EKF takes the velocity as the input and estimates the positions of the UAV and landmarks. Since the monocular vision is a bearing-only sensor,

an inverse depth algorithm is applied in the undelayed feature initialization. The main contribution of this paper is the derivation of the hierarchical SLAM system with bearing-only observations for small UAVs. The performance of this system will be tested by both simulations and experiments. The preliminary work can be found in [21].

2 Feature detection and match

In image processing, a scale invariant feature transform (SIFT) algorithm[22] is used for feature detection. The features are described with normalized descriptors, which are invariant to scale, rotation and translation. A fuzzy-based threshold adjustment method[23] is proposed to stabilize the number of the features in one image regardless of the change of the scene. This method can reduce the computational complexity without any decrease of the accuracy. The feature match is implemented by the calculation of the Euclidean distance between descriptors. A fast nearest-neighbor algorithm is applied to the match. A matched pair is accepted if and only if the distance of the two descriptors is the shortest, less than a threshold and shorter than 0.8 times the distance of the second nearest neighbor. The bi-directional match approach is also adopted to improve the robustness of the match.

The features are used to calculate the motion of the vehicle for attitude and velocity estimations. A contrast threshold is applied to the SIFT algorithm to eliminate feature candidates with low contrast, because they are easy to be disturbed by the image noise. The remaining features are used in the homography calculation. On this basis, parts of the features whose contrasts are higher than another threshold (usually at least twice than the contrast threshold in our experiments) are considered as landmarks in the SLAM system and used in the position estimation and mapping. Fig. 1 illustrates the SIFT features of an aerial image. It can be seen that the features for the position estimation and mapping compose a part of the whole feature set for the homography calculation and motion estimation. Similarly, the feature match is also used in two items in this system: The homography calculation (attitude and velocity estimation) and the data association (position estimation and mapping). The main difference is that the homography calculation requires feature match in consecutive frames, while the data association is the match between the features in the map and observations. Besides, all the features in our algorithm are supposed to be stationary during the whole experiment. Moving features will be eliminated by random sample consensus (RANSAC) algorithm in the homography calculation and data association in the SLAM.
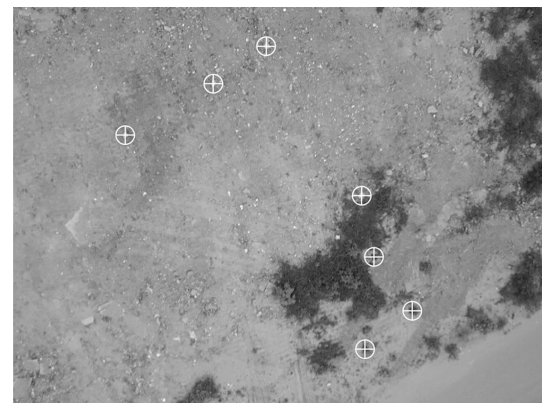
3 Attitude and velocity estimations

The attitude and velocity estimations of the SLAM are described in this section. A homography based method is used to calculate the motion of the vehicle in 6 degrees of freedom by the corresponding features in consecutive frames. Then, the visual measurement is fused with the output of the inertial sensors by an indirect EKF. State update is propagated by the output of the IMU at 50 Hz. The measurement update starts each time when the visual computer finishes the motion estimation. The flowchart of the whole process is shown in Fig. 2.
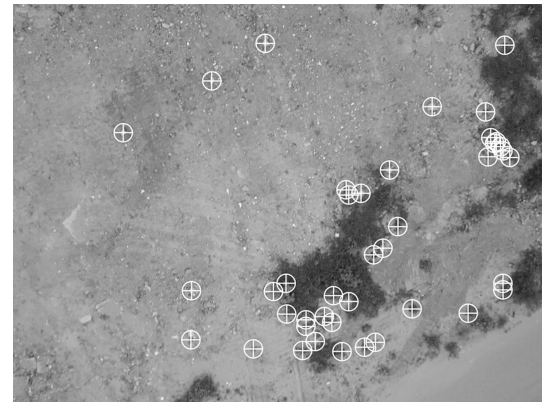
(a) For the motion estimation

(b) Landmark candidates in the SLAM
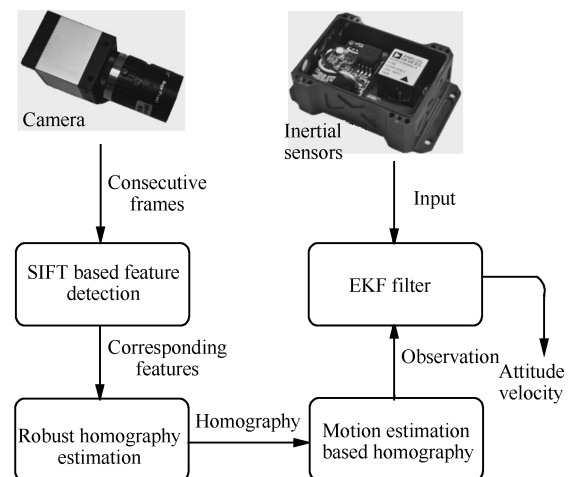
Fig. 1   SIFT features

Fig. 2   The flowchart of the motion estimation

3.1 State description and propagation

The system state for the attitude and velocity estimations is defined as

$$X^a = (v^n, q, b_a, b_\omega)^{\mathrm{T}} \qquad (1)$$

where v^n is the velocity of the vehicle in the world coordinates, q is the attitude of the vehicle in quaternions, b_a and b_ω are the bias of the three-axis accelerometers and the three-axis gyroscopes, respectively. The bias in the state can provide continuous estimation of the current drift of the IMU to compensate the time-varying bias effect. In this paper, the navigation coordinates are the same as the world coordinates. Suppose that the installation of IMU and GPS antenna has been compensated, the inertial process model in the continuous state-space form is expressed as

$$\dot{X}^a(t) = f^a(X^a(t), U(t), t) + G(X^a(t), t)w^a(t) \quad (2)$$

where

$$G = \begin{bmatrix} -C_b^n(q) & O_{3\times3} & O_{3\times3} & O_{3\times3} \\ O_{4\times3} & -\frac{1}{2}Z(q) & O_{4\times3} & O_{4\times3} \\ O_{3\times3} & O_{3\times3} & I_{3\times3} & O_{3\times3} \\ O_{3\times3} & O_{3\times3} & O_{3\times3} & I_{3\times3} \end{bmatrix}$$

$$f^a = \begin{bmatrix} C_b^n(q)\bar{a}^b + g^n \\ \frac{1}{2}\Omega(\bar{\omega}^b)q \\ O_{3\times1} \\ O_{3\times1} \end{bmatrix}$$

$$w^a = \begin{bmatrix} n_a \\ n_\omega \\ n_{ba} \\ n_{b\omega} \end{bmatrix}$$

$$\bar{a}^b = a_m^b - b_a$$

$$\bar{\omega}^b = \omega_m^b - b_\omega.$$

C_b^n is the direction cosine matrix (DCM) from the body coordinates to the world coordinates, g^n is the gravity vector in the world coordinates, as the system control vector, $U(t) = (a_m^b, \omega_m^b)$ are the outputs of the accelerometers and gyroscopes, and $w^a(t) \sim \mathcal{N}(0, W)$ represents the process noise of the system. The attitude equality in the form of quaternion is written as

$$\dot{q} = \frac{1}{2}\Omega(\omega)q = \frac{1}{2}Z(q)\omega \quad (3)$$

where

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

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

$$\Omega = (\omega_1, \omega_2, \omega_3)^{\mathrm{T}}$$

$$q = (q_0, q_1, q_2, q_3)^{\mathrm{T}}.$$

The error state vector is applied in this filter. The state model is linearized and discretized as

$$\delta X_k^a = X_k^a - \hat{X}_k^a = \Phi_{k,k-1}\delta X_{k-1}^a + W_{k-1}^a. \quad (4)$$

In (4), $\hat{X}_k^a$ is the estimation of the state, $\Phi_{k,k-1} \simeq I + \nabla f_x \Delta t$, where ∇f_x is the Jacobian of the state transition function with respect to the state, and Δt is the sampling time interval of the IMU. The covariance matrix of the equivalent white noise sequence W_k^a is derived as

$$Q_k^a \simeq \frac{(\bar{Q} + \Phi\bar{Q}\Phi^{\mathrm{T}})\Delta t}{2} \quad (5)$$

where $\bar{Q} = GWG^{\mathrm{T}}$. Then, the state propagation is written as

$$\hat{X}_{k,k-1}^a = \hat{X}_{k-1}^a + f^a(\hat{X}_{k-1}^a)\Delta t \quad (6)$$

$$P_{k,k-1}^a = \Phi_{k,k-1}P_{k-1}^a\Phi_{k,k-1}^{\mathrm{T}} + Q_{k-1}^a. \quad (7)$$

3.2 Homography based motion estimation

Homography is used to indicate the transformation between two images, including scale, rotation and translation. It is defined as

$$\lambda\tilde{m}_2 = H\tilde{m}_1 \quad (8)$$

where $\tilde{m}_1$ and $\tilde{m}_2$ are homogeneous positions for the corresponding features of two consecutive frames in the pixel coordinates, H is the homography matrix, and λ is a scale factor. The random sample consensus approach[24] is employed to eliminate the erroneous feature matches. The homography is calculated by a singular value decomposition (SVD) method with the feature pairs that pass the RANSAC test. An M-estimator algorithm is used to improve the robustness of the homography estimation.

Suppose that a small UAV with an onboard downward-looking camera is flying, and m_1 and m_2 are the two projections in the camera coordinates of a fixed point P in plane Π, as shown in Fig. 3. R_{12} and t_{12} are defined as the rotation matrix and the translation vector, respectively, which are both described in the camera coordinates of position 1 to express the motion of the vehicle. The relationship of the two projections is expressed as

$$m_2 = R_{12}(m_1 + t_{12}) = R_{12}\left(I + \frac{t_{12}n^{\mathrm{T}}}{d}\right)m_1 \quad (9)$$

where d is the Euclidean distance between position 1 and the plane Π, and n is the unit normal vector of the plane. The calibrated homography is defined as

$$H_c = A^{-1}HA = R_{12}\left(I + \frac{t_{12}n^{\mathrm{T}}}{d}\right) \quad (10)$$

where A is the camera intrinsic matrix which can be calibrated accurately. Equality (10) indicates the relationship between the homography and the motion of the camera. R_{12} and t_{12} can be obtained by the SVD of the calibrated homography H_c[25]. Since the transformation between the camera coordinates and the body coordinates is known accurately before the flight, the motion of the camera estimated by homography calculation can be converted to the motion of the UAV.
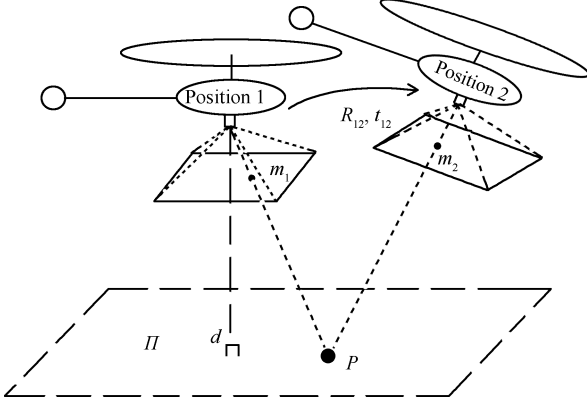
Fig. 3   Two views of the same fixed point in a plane

3.3   Measurement update

The DCM from the body coordinates at position 2 in Fig. 3 to the navigation coordinates is derived as

$$C_{b2}^n = C_{b1}^n C_c^b R_{12}^{\mathrm{T}} C_b^c \tag{11}$$

where C_{b1}^n is the DCM from the body coordinates at position 1 to the navigation coordinates, which can be calculated from the state of this filter at that moment, C_c^b and C_b^c are transformation matrixes between the camera coordinates and the body coordinates. The attitude of the vehicle is calculated from C_{b2}^n and used as a measurement of the filter. The measurement model is written as

$$Z_{1,k} = H_1 X_k^a + V_{1,k} \tag{12}$$

where

$$H_1 = [O_{4\times3} \quad I_{4\times4} \quad O_{4\times3} \quad O_{4\times3}]$$

$$Z_1 = q$$
$$V_1 \sim \mathcal{N}(0, R_1).$$

Then, the attitude measurement update is derived as

$$K_{1,k} = P_{k,k-1}^a H_1^{\mathrm{T}} (H_1 P_{k,k-1}^a H_1^{\mathrm{T}} + R_1)^{-1} \tag{13}$$

$$\hat{X}_k^a = \hat{X}_{k,k-1}^a + K_{1,k}(Z_{1,k} - H_1 \hat{X}_{k,k-1}^a) \tag{14}$$

$$P_k^a = (I - K_{1,k} H_1) P_{k,k-1}^a. \tag{15}$$

Suppose that two consecutive images are taken and calculated at time $k-m$ and k, respectively. The time interval between these two calculations is about m steps of the state propagation. The average velocity of the vehicle in the navigation coordinates during the time interval is derived as

$$\bar{v}_{k-m,k}^n = \frac{C_{b1}^n C_c^b R_{12} t_{12}}{m\Delta t}. \tag{16}$$

Since the velocity in the filter state is the instantaneous velocity, the above average velocity cannot be introduced into the filter directly. An assumption is made that during a short time interval in the flight, the velocity of the vehicle does not fluctuate remarkably. Then, the average velocity during time $k - m$ and time k can be approximated as the instantaneous velocity at time $k - \frac{m}{2}$ as

$$v_{k-\frac{m}{2}}^n \approx \bar{v}_{k-m,k}^n. \tag{17}$$

An example of the relationship between the average velocity and the instantaneous velocity is shown in Fig. 4. It can be observed that the above assumption is reasonable if the velocity is smooth. We suppose that the small UAV is in a stable flight and the above assumption is always tenable.

It is noteworthy that the velocity measurement at time $k - \frac{m}{2}$ is obtained at time k under the above assumption. A delay-based measurement update approach is proposed for the velocity measurement update. The delay-based measurement model for the velocity estimation is defined as

$$Z_{2,k} = H_2 X_k^a + V_{2,k} \tag{18}$$
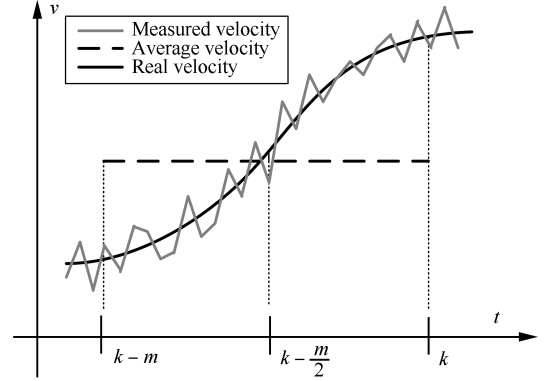

Fig. 4   Relationship between the average velocity and the instantaneous velocity

$$H_2 = [I_{3\times3} \quad O_{3\times4} \quad O_{3\times3} \quad O_{3\times3}]$$

$$Z_2 = v^n$$
$$V_2 \sim \mathcal{N}(0, R_2).$$

The velocity measurement update is derived as

$$K_{2,k} = P_{k,k-1}^a H_2^{\mathrm{T}} (H_2 P_{k,k-1}^a H_2^{\mathrm{T}} + R_2)^{-1} \tag{19}$$

$$\hat{X}_k^a = \hat{X}_{k,k-1}^a + K_{2,k}(Z_{2,k} - H_2 \hat{X}_{k-\frac{m}{2}}^a) \tag{20}$$

$$P_k^a = (I - K_{2,k} H_2) P_{k,k-1}^a. \tag{21}$$

4   Position estimation and mapping

This section presents the position estimation and mapping of the SLAM. No prior information of the scene is required in this EKF based filter, which takes the observation of the features as the measurement. The velocity estimated in Section 3 is used as the input of the process model, and the attitude estimation provides the DCMs between the navigation coordinates and the body coordinates.

4.1   Process model

The vehicle position and a map with the feature locations of the environment are estimated using relative information between the vehicle and each feature. The state vector is defined as

$$X^b = (p^n, Y_1, \cdots, Y_n)^{\mathrm{T}} \tag{22}$$

where p^n is the vehicle position in the navigation coordinates and $Y_i, i = 1, \cdots, n$ is the location of the i-th feature

in the navigation coordinates. The dynamic evolution of the position is given as

$$
\begin{aligned}
p_k^n &= f^b(X_{k-1}^a, p_{k-1}^n) + w_{k-1}^b = \\
&\quad p_{k-1}^n + v_{k-1}^n \Delta t + w_{k-1}^b
\end{aligned}
\tag{23}
$$

where v^n is the velocity estimated in Section 3 and w^b is the Gaussian noise as $w^b \sim \mathcal{N}(0, P_v)$. Feature locations are considered to be stationary and the process model of the i-th feature is given as

$$
Y_{i,k} = Y_{i,k-1}. \tag{24}
$$

It can be seen from (23) and (24) that the state transition function is linear. This is favorable for the computational complexity and stability of the filter.

4.2 Observation model

The onboard camera is able to produce relative bearing-only observations to the features in the map. As multiple features might be observed at the same time, the observation is defined as

$$
Z_3 = (z_i, \cdots, z_j)^{\mathrm{T}} \tag{25}
$$

where z_i is the observation of the i-th feature in the pixel coordinates, which is expressed as

$$
z_{i,k} = h_i(X_k^b, k) + v_{i,k} \tag{26}
$$

$$
h_i = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \dfrac{f_u x^c}{z^c} + u_0 \\ \dfrac{f_v y^c}{z^c} + v_0 \end{bmatrix} \tag{27}
$$

where $(u, v)^{\mathrm{T}}$ is the location of the feature in the pixel coordinates, v_i is the measurement noise as $v_i \sim \mathcal{N}(0, R_i)$, f_u, f_v, u_0 and v_0 are the intrinsic parameters of the camera, and $p^c = (x^c, y^c, z^c)^{\mathrm{T}}$ is the location of the feature in the camera coordinates. If the installation error of the camera to the vehicle has been compensated, then p^c can be written as

$$
p^c = C_b^c C_n^b (p_i^n - p^n) \tag{28}
$$

where C_n^b is calculated from the attitude estimation in Section 3, p^n and p_i^n are the positions of the vehicle and the i-th landmark in the map in the navigation coordinates, respectively.

4.3 Estimation process

The vehicle position is predicted using (23). The state covariance is propagated as

$$
\begin{aligned}
&P_{k,k-1}^b = \\
&\begin{bmatrix} F_{p^n}^b P_{v,k-1}(F_{p^n}^b)^{\mathrm{T}} + F_{x^a}^b P_{k-1}^a (F_{x^a}^b)^{\mathrm{T}} & P_{vf,k-1} \\ P_{fv,k-1} & P_{f,k-1} \end{bmatrix}
\end{aligned}
\tag{29}
$$

where $F_{p^n}^b$ and $F_{x^a}^b$ are the Jacobians of (23) with respect to p^n and X^a, P_v and P_f are the variances of the vehicle and the feature set, P_{vf} and P_{fv} are their covariances.

Suppose at time k, there are n features in the state and that the i-th and j-th features are observed by the camera. The measurement update is derived as

$$
K_{3,k} = P_{k,k-1}^b H_3^{\mathrm{T}} (H_3 P_{k,k-1}^b H_3^{\mathrm{T}} + R_3)^{-1} \tag{30}
$$

$$
\hat{X}_k^b = \hat{X}_{k,k-1}^b + K_{3,k}(Z_{3,k} - H_3 \hat{X}_{k,k-1}^b) \tag{31}
$$

$$
P_k^b = (I - K_{3,k} H_3) P_{k,k-1}^b \tag{32}
$$

where

$$
R_3 = \begin{bmatrix} R_i & 0 \\ 0 & R_j \end{bmatrix}
$$

$$
H_3 = \begin{bmatrix} \dfrac{\partial h_i}{\partial p^n} & \dfrac{\partial h_i}{\partial Y_1} & \cdots & \dfrac{\partial h_i}{\partial Y_n} \\ \dfrac{\partial h_j}{\partial p^n} & \dfrac{\partial h_j}{\partial Y_1} & \cdots & \dfrac{\partial h_j}{\partial Y_n} \end{bmatrix}.
$$

R_i and R_j are the measurement covariances of the i-th and j-th features in the pixel coordinates, $\frac{\partial h}{\partial p^n}$ and $\frac{\partial h}{\partial Y}$ are the Jacobians of the measurement function with respect to the vehicle position and the feature locations, respectively.

4.4 Feature initialization

In the monocular visual system, one bearing-only observation is insufficient to initialize a position of a feature in 3 dimensions for the SLAM filter with Gaussian uncertainty[11]. An inverse depth method is applied to the undelayed feature initialization[14]. This parameterization defines the position of a feature by 6 parameters as

$$
Y_i = (x_0, y_0, z_0, \theta, \phi, \rho)^{\mathrm{T}} \tag{33}
$$

where $p_0 = (x_0, y_0, z_0)^{\mathrm{T}}$ is the position of the camera optical center from which the feature is first observed, θ and ϕ are the azimuth and elevation used to define a ray going from p_0 to the feature, and ρ is the inverse distance from p_0 to the feature along the ray. This parameterization is shown in Fig. 5. $(x, y, z)^{\mathrm{T}}$ represents the navigation coordinates with the origin o. The location of this landmark in the Euclidean coordinates is derived as

$$
p_i = p_0 + \frac{m(\theta, \phi)}{\rho} \tag{34}
$$

$$
m(\theta, \phi) = \begin{bmatrix} \cos\theta \cos\phi \\ \sin\theta \cos\phi \\ \sin\phi \end{bmatrix}. \tag{35}
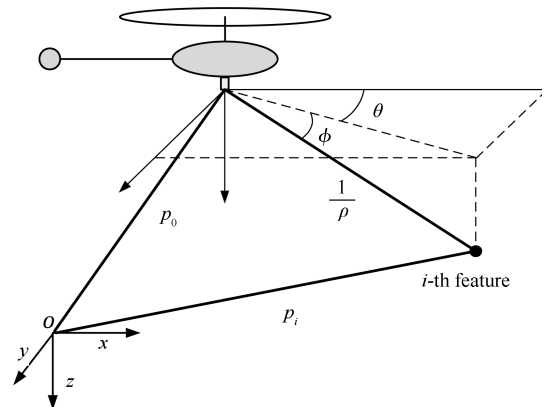$$

Fig. 5　Inverse depth parameterization

4.5 Data association

Data association is very important to any SLAM application. The association between observations and features in the map affects the stability of the system directly. Erroneous association can decrease the accuracy or even lead to system collapse. A Mahalanobis distance approach is used in the data association, which is defined as

$$r = V^{\mathrm{T}} S^{-1} V \tag{36}$$

where

$$V = Z_3 - H_3 \hat{X}^b \tag{37}$$

$$S = H_3^{\mathrm{T}} P^b H_3 + R_3. \tag{38}$$

Besides, the descriptor of the SIFT feature is also used to improve the robustness of the data association. Each SIFT feature has a unique 128-dimensional normalized descriptor, which is invariant to scale, rotation and translation. Thus, the descriptor of a new observation is compared with those of the features in the map. The fast nearest-neighbor algorithm, described in Section 2, is also used in this section.

In our SLAM framework, observation z_i is considered as the correspondence to the j-th landmark in the map if the following equation is satisfied:

$$d_{i,j} = \sqrt{r^2 + d_{\mathrm{SIFT}}^2} < \text{threshold} \tag{39}$$

where d_{SIFT} is the Euclidean distance between the SIFT descriptor vectors of the observation and the landmark. Only the observation that passes the above test is considered as a right association. Although this double check algorithm might reject correct matches that only pass one test, the system robustness is improved with erroneous associations eliminated remarkably.

5 Simulations

5.1 Simulation environment setup

An image-in-loop simulation system is set up to verify the proposed SLAM, as shown in Fig. 6. A satellite image taken from Google Earth is loaded as the terrain map of the flight. The aerial image is simulated by the pixel interpolation with a resolution of 300×300 in the pixel coordinates. The sensor data of the gyroscopes and accelerometers are simulated with white noises and bias. The intrinsic matrix of the "onboard camera" is designed according to the parameters of the real one in experiments. The features in the map are described with their uncertainties in 3 dimensions. Feature labels are also shown together with the "real-time" aerial images.

5.2 3D feature initialization

The feature initialization method is analyzed in this section. Fig. 7 shows the position errors of a feature during the first 15 observations and the expected 1-σ uncertainties. When the feature is first observed, the uncertainty is rather large due to the initial selection of the inverse depth. The uncertainty converges rapidly after other observations

are obtained. The position error lies within the 1-σ bound from the beginning to the end and converges with the increase of the observations. After 10 observations, the position estimation of the feature is stable and the inverse depth is accurate enough. Then, the feature parameters are converted from the inverse depth form to the Euclidean form. No obvious changes can be observed in the feature position estimation and the uncertainty calculation during this conversion.
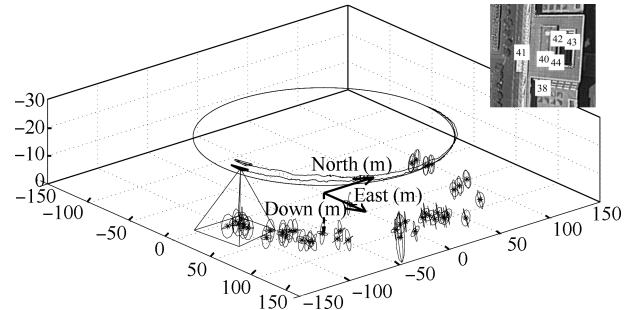
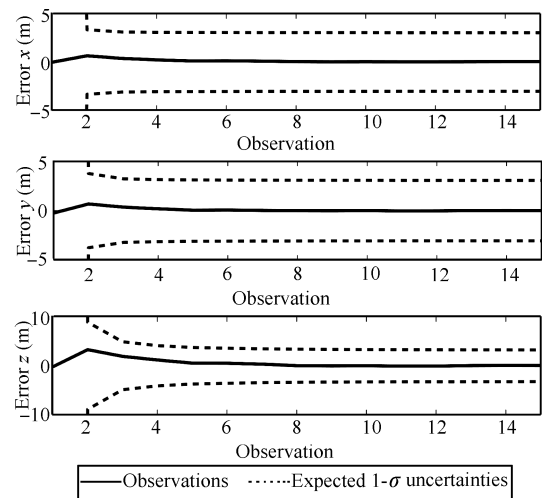Fig. 6   Image-in-loop simulation system

Fig. 7   The position errors of the feature initialization

5.3 Trajectory test with loop closure

A circle trajectory with a radius of 100 m is designed to test the SLAM system. The UAV flies two laps along the trajectory with the velocity of 10 m/s in the body coordinates, so that there is a loop closure during the flight. The processing time of this algorithm is about 120 s.

Figs. 8 and 9 are about the attitude and velocity errors with 1-σ uncertainties, respectively. Both the attitude and velocity errors are limited within the uncertainty bound during the whole experiment.

Fig. 10 illustrates the position errors with 1-σ uncertainties. The estimations in the z direction seem to be better than those in $x - y$ directions in this simulation. It could be attributed to the unchanged altitude in the designed trajectory, which makes the motion in the z direction much tiny compared to the motion in $x - y$ directions. The position errors are small during the first 20 s. Then the estimations suffer from an accumulated error which is generated by the

velocity error. The error might drift unboundedly in a few seconds without the observation of the features. But in the SLAM system, it is approximately bounded within the 1-σ uncertainties. Only the error in the x direction from 40 s to 60 s is outside the bound. The loop closure takes place at about 60 s. It can be seen that both the errors in the x and y directions decrease sharply. From this time, the UAV is in the second lap and a lot of features which are initialized in the first lap are observed again. Thus, after the loop closure, the errors keep smaller and stable within the boundaries. The uncertainty is also convergent.
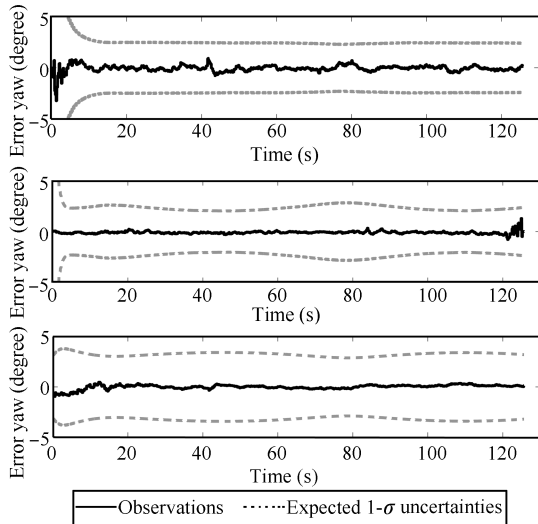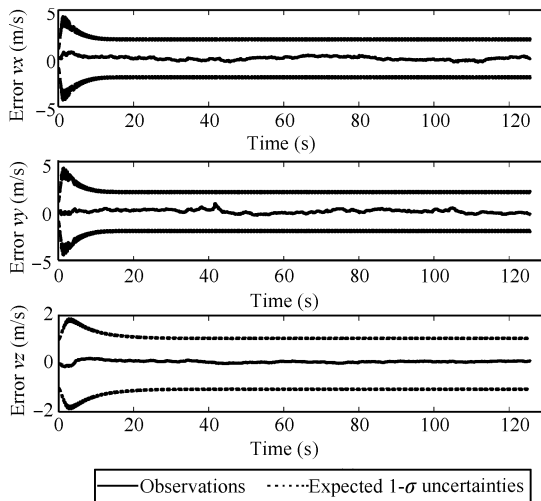

Fig. 8   The attitude errors in the simulation

Fig. 9   The velocity errors in the simulation

6   Experiments

6.1   System description

The experimental testbed is a radio-controlled model helicopter, as shown in Fig. 11. It carries a set of sensors, including an Rt 20 differential GPS receiver, an MS5540 barometric altimeter, an ADIS16355 IMU, a three-axis magnetometer composed of HMC1052 and HMC1041, and a color

industrial digital camera MV-1300UC mounted vertically downward at the bottom of the helicopter.
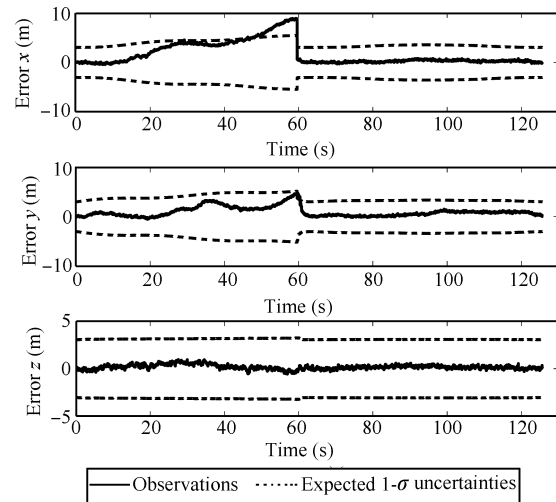

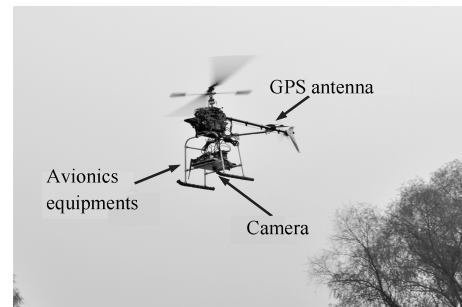Fig. 10   The position errors in the simulation

Fig. 11   The experimental helicopter testbed

There are four onboard computers in the onboard avionics. DSP TI28335 runs an extended Kalman filter to fuse the measurement of GPS and inertial sensors to build a GPS/INS (inertial navigation system) for the helicopter. ARM LPC1766 is used to guide and output the control signals to the helicopter, as well as communicate with the ground station. FPGA EP2C is the data logger and in charge of the control switching between human pilot and the autopilot. PC-104 PCM3362 is the vision computer. During the real experiment, the sensor data and the GPS/INS navigation output are logged by the FPGA and transferred to the vision computer. The vision computer receives the above data and records the aerial images simultaneously. The ground station is used to send the high-level control commands and differential GPS correction to the helicopter and receive the flight state of the helicopter, including attitude, velocity and position. The onboard avionics is shown in Fig. 12.

6.2   Real flight experiment

An experiment is carried out to test the performance of the proposed SLAM system. During the experiment, the inertial sensor data is acquired at a rate of 50 Hz and the visual calculation is about 3–4 Hz. The improvement of the visual sampling frequency is able to reduce the error introduced by the assumption in (17). But now the sampling

frequency of the visual system is limited by the computational load of the visual algorithm. The SLAM system runs in an offline mode in this paper. The performance of the proposed bearing-only inertial SLAM is evaluated by comparing with the referential GPS/INS navigation.
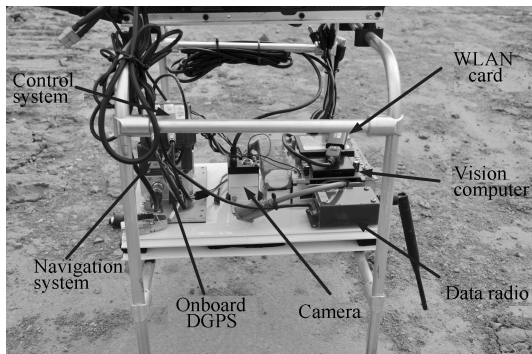

Fig. 12   The onboard avionics endcenter

The attitude comparison of the whole flight is shown in Fig. 13. It can be seen from the yaw estimations of the vehicle that the small UAV flies about 4 loops in the experiment. The visual measurement does not work during the takeoff and landing, in which the visual estimation is not accurate enough due to the rapid change of the aerial images. The visual measurement takes effect from 60 s to 240 s in this experiment. During this period, GPS signal is cutoff in our SLAM system and the output of the barometric sensor is used to provide a relative height for the visual motion estimation. Fig.13 shows that there is a strong agreement of the attitude estimations between the proposed SLAM system and the referential GPS/INS navigation. Fig. 14 is about the comparison of the velocity estimations. It shows that our system also has a high accuracy in the velocity estimation compared with the referential.
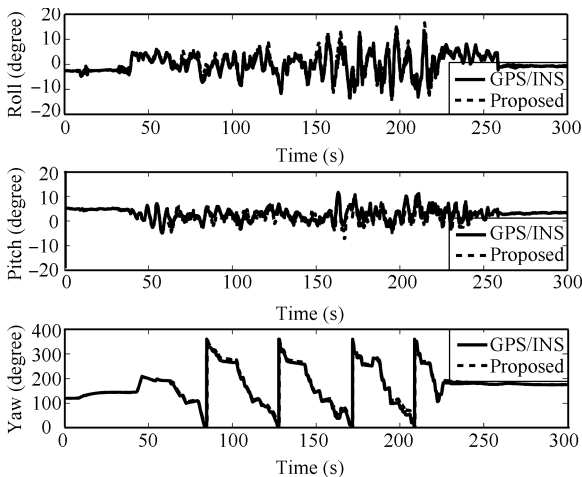

Fig. 13   The attitude comparison in the experiment

The position estimated by the direct velocity integration, which is known as visual odometry (VO), is also introduced into the position comparison as shown in Fig. 15. Despite the accurate velocity estimation shown in Fig. 14, errors

accumulate in the position estimation through direct integration. After the landing, deviations between the VO estimation and the GPS/INS system are quite obvious. The correction of the SLAM on the position estimation can be seen in the comparison. Taking the velocity estimation as the input, the proposed system is able to correct the position estimation by consecutive observations of the features in the map. The position estimated by the SLAM has a higher accuracy than the VO, and the deviations are eliminated remarkably.
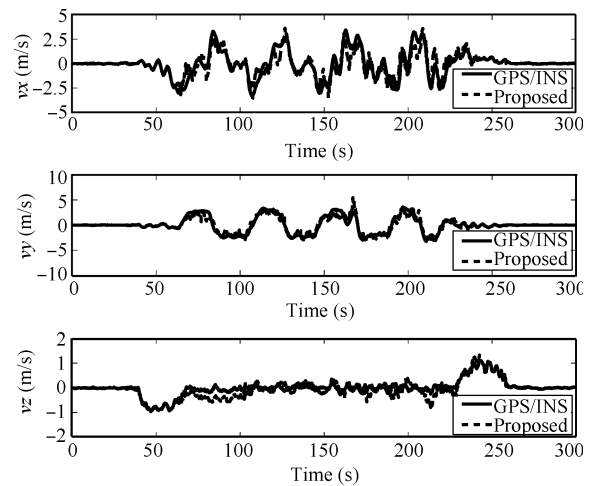

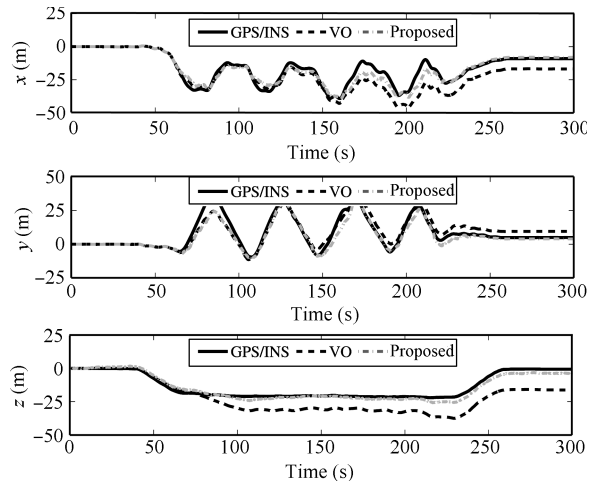Fig. 14   The velocity comparison in the experiment

Fig. 15   The position comparison in the experiment

7   Conclusions and future work

This paper describes a bearing-only visual SLAM for small UAVs in GPS-denied environments. An indirect EKF is used to estimate the attitude and velocity by the fusion of the IMU and the visual motion measurement of the vehicle, which is calculated by a homography based method. Then, these attitude and velocity estimations are input into another EKF based filter for position estimation and map building. An inverse depth method is applied to the undelayed feature initialization. A method that combines the Mahalanobis distance and the descriptor match of the SIFT

features is used to improve the robustness of the data association. Simulations and real experiments have been carried out to test the performance of the system. The SLAM system developed in this paper has a high estimation accuracy of the UAV state by comparing with the referential GPS/INS navigation. In conclusion, our bearing-only visual SLAM system can provide stable estimations of small UAVs while building a 3D feature map in GPS-denied environments.

Future work will focus on the further improvement of the SLAM on computational complexity. In the EKF based algorithm, the state is always augmented with the observations of new features in the map. The computational complexity grows quadratically with the number of the features. This is an unavoidable problem for the implementation of the real-time SLAM for small UAVs. In the future, a Rao-Blackwellized based FastSLAM will be developed based on this research for a real-time SLAM that can reduce the computation to a linear complexity.

References

[1] D. B. Kingston, A. W. Beard. Real-time attitude and position estimation for small UAVs using low-cost sensors. In *Proceedings of the AIAA 3rd Unmanned Unlimited Technical Conference, Workshop and Exhibit*, AIAA, Chicago, USA, pp. 6488–6496, 2004.

[2] B. Ludington, E. Johnson, G. Vachtsevanos. Augmenting UAV autonomy. *Robotics & Automation Magazine*, vol. 13, no. 3, pp. 63–71, 2006.

[3] M. K. Kaiser, N. R. Gans, W. E. Dixon. Vision-based estimation for guidance, navigation, and control of an aerial vehicle. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 3, pp. 1064–1077, 2010.

[4] S. Weiss, D. Scaramuzza, R. Siegwart. Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments. *Journal of Field Robotics*, vol. 28, no. 6, pp. 854–874, 2011.

[5] P. Yang, W. Y. Wu, M. Moniri, C. C. Chibelushi. A sensor-based SLAM algorithm for camera tracking in virtual studio. *International Journal of Automation and Computing*, vol. 5, no. 2, pp. 152–162, 2008.

[6] J. Artieda, J. M. Sebastian, P. Campoy, J. F. Correa, I. F. Mondragon, C. Martinez, M. Olivares. Visual 3-D SLAM from UAVs. *Journal of Intelligent & Robotic Systems*, vol. 55, no. 4–5, pp. 299–321, 2009.

[7] F. Caballero, L. Merino, J. Ferruz, A. Ollero. Vision-based odometry and SLAM for medium and high altitude flying UAVs. *Journal of Intelligent & Robotic Systems*, vol. 54, no. 1–3, pp. 137–161, 2009.

[8] J. Kelly, S. Saripalli, G. S. Sukhatme. Combined visual and inertial navigation for an unmanned aerial vehicle. In *Proceedings of the International Conference on Field and Service Robotics*, FSR, Chamonix, France, pp. 255–264, 2007.

[9] K. H. Yang, W. S. Yu, X. Q. Ji. Rotation estimation for mobile robot based on single-axis gyroscope and monocular camera. *International Journal of Automation and Computing*, vol. 9, no. 3, pp. 292–298, 2012.

[10] V. Sazdovski, P. M. G. Silson. Inertial navigation aided by vision-based simultaneous localization and mapping. *IEEE Sensors Journal*, vol. 11, no. 8, pp. 1646–1656, 2011.

[11] M. Bryson, S. Sukkarieh. Building a robust implementation of bearing-only inertial SLAM for a UAV. *Journal of Field Robotics*, vol. 24, no. 1–2, pp. 113–143, 2007.

[12] J. Kim, S. Sukkarieh. Real-time implementation of airborne inertial-SLAM. *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 62–71, 2007.

[13] A. J. Davison, I. D. Reid, N. D. Molton, O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[14] J. Civera, A. J. Davison, J. Montiel. Inverse depth parametrization for monocular SLAM. *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 932–945, 2008.

[15] G. Nützi, S. Weiss, D. Scaramuzza, R. Siegwart. Fusion of IMU and vision for absolute scale estimation in monocular SLAM. *Journal of Intelligent & Robotic Systems*, vol. 61, no. 1–4, pp. 287–299, 2011.

[16] S. D. Huang, G. Dissanayake. Convergence and consistency analysis for extended Kalman filter based SLAM. *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 1036–1049, 2007.

[17] M. Bryson, S. Sukkarieh. Observability analysis and active control for airborne SLAM. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 261–280, 2008.

[18] A. Nemra, N. Aouf. Robust airborne 3D visual simultaneous localization and mapping with observability and consistency analysis. *Journal of Intelligent & Robotic Systems*, vol. 55, no. 4–5, pp. 345–376, 2009.

[19] M. T. Bryson, S. Sukkarieh. Decentralised trajectory control for multi-UAV SLAM. In *Proceedings of the 4th International Symposium on Mechatronics and its Applications*, Sharjah, United Arab Emirates, 2007.

[20] K. Celik, C. Soon-Jo, M. Clausman, A. K. Somani. Monocular vision SLAM for indoor aerial vehicles. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, St. Louis, USA, pp. 1566–1573, 2009.

[21] C. L. Wang, T. M. Wang, J. H. Liang, Y. Chen, Y. C. Zhang, C. Wang. Monocular visual SLAM for small UAVs in GPS-denied environments. In *Proceedings of the 2012 IEEE International Conference on Robotics and Biomimetics*, IEEE, Guangzhou, China, pp. 896–901, 2012.

[22] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] C. L. Wang, T. M. Wang, J. H. Liang, Y. Chen, C. Wang. A fuzzy-based threshold method applied in SIFT for visual navigation of small UAVs. In *Proceedings of the 7th IEEE Conference on Industrial Electronics and Applications*, IEEE, Singapore, pp. 807–812, 2012.

[24] M. A. Fischler, R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[25] R. Tsai, T. Huang, W. L. Zhu. Estimating three-dimensional motion parameters of a rigid planar patch, II: Singular value decomposition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 30, no. 4, pp. 525–534, 1982.
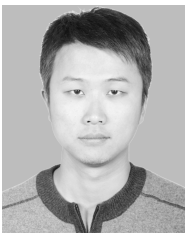
Chao-Lei Wang received the B. Sc. degree in mechanical engineering and automation from Beijing University of Aeronautics and Astronautics, China in 2008. He is now a Ph. D. candidate in mechanical engineering and automation, Beijing University of Aeronautics and Astronautics, China.

His research interests include system modeling and identification, integrated navigation and visual SLAM in small unmanned aerial vehicle.

E-mail: chaoleiwang@163.com (Corresponding author)

Tian-Miao Wang received the B. Sc. degree from Xi'an Jiaotong University, China, and the M. Sc. and Ph. D. degrees from Northwestern Polytechnic University, China in 1982 and 1990, respectively. He is currently a professor with the School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, China. He is the head of the Expertized Group in the field of advanced manufacturing technology of the National High Technology Research and Development Program of China (863 Program). He has undertaken and finished many national research projects in recent years. He has published 87 papers in local and international journals and three professional books. He is a member of China Robotics Society.

His research interests include biomimetic robotics, medical robotic technology and embedded intelligent control technology.

E-mail: wtm_itm@263.net

Jian-Hong Liang received the B. Sc. and Ph. D. degrees in mechanical engineering and automation from Beijing University of Aeronautics and Astronautics, China in 2000 and 2007, respectively. He is currently an associate professor with the School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, China. He has published 40 papers in national and international journals and two professional books.

His research interests include biomimetic underwater robotics, field robotics and small unmanned aerial vehicles.

E-mail: dommy_leung@263.net

Yi-Cheng Zhang received the B. Sc. degree in mechanical engineering and automation from Beijing University of Aeronautics and Astronautics, China in 2010. He is currently a Ph. D. candidate in mechanical engineering and automation, Beijing University of Aeronautics and As Astronautics, China.

His research interests include flight controller design and system integration of small unmanned aerial vehicle.

E-mail: zycet@126.com

Yi Zhou received the B. Sc. degree in mechanical engineering and automation from Beijing University of Aeronautics and Astronautics, China in 2012. He is currently a Ph. D. candidate in mechanical engineering and automation, Beijing University of Aeronautics and Astronautics, China.

His research interests include visual navigation and guidance of small unmanned aerial vehicle.

E-mail: cavatina@yeah.net