Adaptive Noise Identification in Vision-assisted Motion Estimation for Unmanned Aerial Vehicles

Fan Zhou[1]   Wei Zheng[1]   Zeng-Fu Wang[1, 2]

[1]Department of Automation, University of Science and Technology of China, Hefei 230027, China

[2]Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China

Abstract: Vision localization methods have been widely used in the motion estimation of unmanned aerial vehicles (UAVs). The noise of the vision location result is usually modeled as a white Gaussian noise so that this location result could be utilized as the observation vector in the Kalman filter to estimate the motion of the vehicle. Since the noise of the vision location result is affected by external environment, the variance of the noise is uncertain. However, in previous researches, the variance is usually set as a fixed empirical value, which will lower the accuracy of the motion estimation. The main contribution of this paper is that we proposed a novel adaptive noise variance identification (ANVI) method, which utilizes the special kinematic properties of the UAV for frequency analysis and then adaptively identifies the variance of the noise. The adaptively identified variance is used in the Kalman filter for more accurate motion estimation. The performance of the proposed method is assessed by simulations and field experiments on a quadrotor system. The results illustrate the effectiveness of the method.

Keywords: Adaptive noise variance identification, vision location, motion estimation, Kalman filter, unmanned aerial vehicle.

1 Introduction

The unmanned aerial vehicles (UAVs) have become more and more popular in recent years because of their wide applications in mobile missions such as surveillance, exploration and recognition in different environments. Fig. 1 shows a flying UAV system. A main problem in applications of UAVs is the estimation of the motion of the system, including 3D position and translational speed.

Fig. 1   Unmanned aerial vehicle

Many research works have been done in this field, using various kinds of location sensors including global position system (GPS)[1], laser range sensors[2, 3], doppler radars[4], ultrasonic sensors[5], etc. However, the factors of accuracy,

weight, cost, and applicable environment limit the application of these sensors on aerial vehicles. Vision sensors, with their advantage in these aspects, have become a popular choice for providing location results of the system[6].

The Kalman filter model is widely used to obtain accurate, fast updated and reliable motion estimation of the UAV system[5, 7, 8], which generally consists of two equations: the observation equation and the state equation. Results directly provided by the vision localization method are used to establish the observation equation of the Kalman filter, and measurements from inertial sensors are usually used to establish the state equation.

In the Kalman filter, the variance of the noise is needed for estimation. However, in vision localization methods, especially in the method without artificial landmarks, which is utilized in this paper, feature detecting and matching components are usually included[9]. The location accuracy is obviously affected by external environments, such as illumination, camera resolution, texture of environment, height of flight, etc. Therefore, the variance of the noise of the vision location result is changeable and hard to calibrate in the beginning. However, so far as we know, no related work has paid attention to this problem. The variance of the noise is usually simply set as an empirical parameter which will probably lower the accuracy of the motion estimation, especially when there is a large deviation between the empirical parameter and the truth variance.

In this article, some special kinematic properties of the UAV system, which were barely utilized before, are noticed and utilized for frequency analysis of the position signal (or the trajectory) of the vehicle. The derivation shows that the position signal has some characteristic in the frequency domain which helps to separate it from the noise and

therefore the variance of the noise could be identified. An simplified version of this analysis was discussed in [10].

The rest of the paper is organized as follows. Section 2 introduces the entire system where the adaptive noise variance identification (ANVI) method is applied, including the configuration of the UAV system and the principle of the motion estimation. Section 3 proposes the ANVI method detailedly. Experiment and results are shown in Section 4 to verify verifies the feasibility and performance of the proposed method. Some conclusions are presented in Section 5.

2 System introduction

2.1 UAV system configuration

The configuration of our system is shown in Fig. 2. The main sensors onboard are a downward looking monocular camera, a height sensor and an inertial measurement unit (IMU), including an accelerometer, a gyroscope and a magnetometer. A wireless link is used to share information with the ground PC, where the vision location algorithm is running. An attitude estimation algorithm is executed on the onboard micro controller. Although we do not derive the attitude estimation in this paper, some qualities of the attitude estimation algorithm are indeed utilized for analysis of the position signal of the system, which will be explained in Section 3.1.
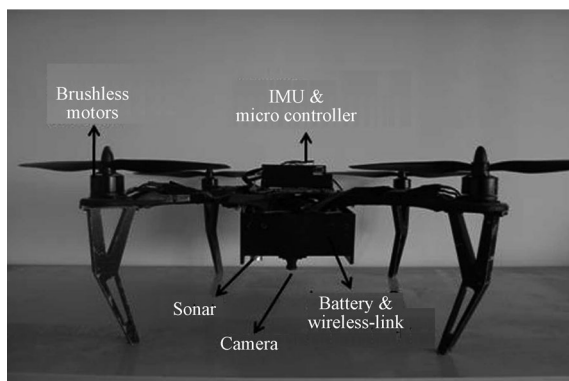

Fig. 2   The configuration of the vehicle

2.2 Principle of motion estimation

For automatic application, the states of motion need to be estimated, generally including 3D position p and 3D translational speed v, which form a 6-dimensional state vector $X = (p, v)^{\mathrm{T}}$. Fig. 3 shows the entire framework of state estimation in this paper. The unique component of ANVI will be derived in Section 3. The other three components, as general parts of the motion estimation, will be described here in Section 2.2.
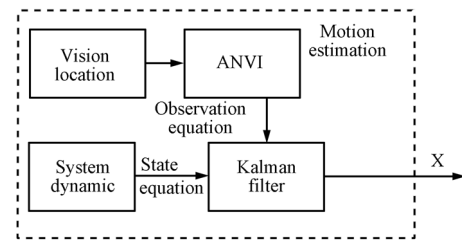

Fig. 3   The framework of motion estimation

2.2.1   Vision location

Markless vision localization methods have been widely used for navigation of UAVs in unknown environment. Most of them are accomplished in two steps. First, feature points are detected and matched between two frames of images captured in two different camera locations, as shown in Fig. 4. Secondly, relative transformation of location including rotation and translation is calculated with the corresponding points.
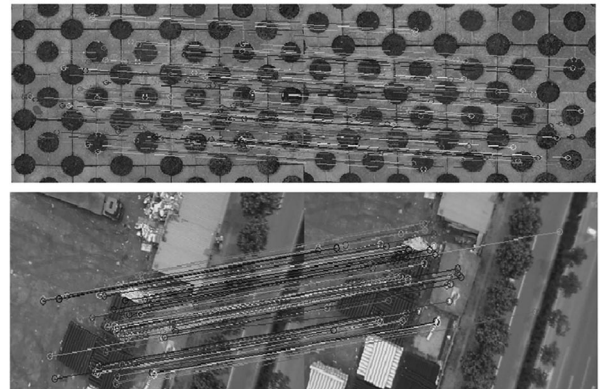

Fig. 4   Vision location of UAVs in different environments

As shown in Fig. 5, the vehicle moved from location 1 to location 2. After feature detecting and matching, a set of corresponding points was founded. Let $\widetilde{m}_1 = (u_1, v_1, 1)$ and $\widetilde{m}_2 = (u_2, v_2, 1)$ be homogeneous coordinates of one pair of corresponding points. According to the theory of structure from motion[11], the relationship between $\widetilde{m}_1$ and $\widetilde{m}_2$ could be presented as

$$d_2\widetilde{m}_2 = R_{12}d_1\widetilde{m}_1 + t_{12} \qquad (1)$$

where d_1 and d_2 are scale factors, R_{12} and t_{12} denote the relative rotation and translation, respectively. Because monocular camera could not give an absolute scale, a height sensor is used to help to recover the scale, as explained in [12]. Then, the singular value decomposition (SVD) technique is used to find the best possible R_{12} and t_{12} which relates the set of pairs of corresponding points by (1). Since the transformation between the camera frame and the body frame is known accurately when the system is installed, and the rotation between the body frame and the world frame could be obtained by the IMU[13], R_{12} and t_{12} can provide the relative location of the vehicle in the world frame.
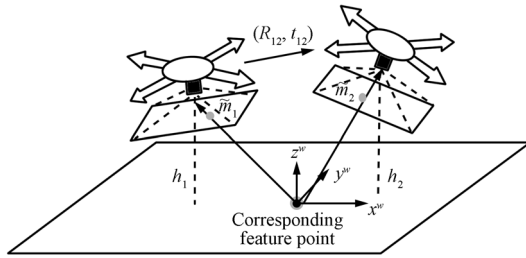
Fig. 5   Vision location

Notice that two frames of images could be matched only when the vehicle moves in a limited range of area. To obtain location results in a large area, the relative location results calculated above will be utilized in a so called "SLAM" algorithm that consists of "location" and "mapping". The main idea is that the small displacements obtained from the localization methods above are accumulated to obtain a large displacement, so the UAV could locate itself in a large area. Besides, key frames of images are stored to regulate the accumulated error. With the entire "location" and "mapping" solution, the UAV could accomplish automatic missions in a large range of space. For more details of the vision slam algorithm, one should refer to [6, 9, 14]. What we are concerned about more in this paper is the uncertain noise of the location result.

The location result provided here by the vision localization method will be used as the observation vector Z in the Kalman filter. Note that this observation contains noise, which is denoted as ξ and usually modeled as the white Gaussian noise. The variance of ξ is needed in the Kalman filter for estimation.

2.2.2   Dynamic of the system

Let w and b denote the world frame and the body frame, respectively. Axes x^w, y^w, z^w in the world frame correspond to the east, the north, and the upright directions, respectively. Axes x^b, y^b, z^b in the body frame are attached to the vehicle body and the center of the body frame is the center of mass of the system.

Therefore, the dynamic equation of the system motion is given by

$$\dot{p}^w = v^w$$
$$\dot{v}^w = a^w \qquad (2)$$

where $p^w = (x^w, y^w, z^w)$ denotes the position in the world frame. v^w, a^w denotes the velocity and acceleration in the world frame, respectively. R_b^w is defined as the rotation matrix from the body frame to the world frame. Then, the acceleration in the world frame could be obtained from

$$a^w = R_b^w a^b$$
$$a^b = a^m - n_a - \overrightarrow{g} \qquad (3)$$

where a^b and a^m denote the acceleration in the body frame and the acceleration measured by the accelerometer sensor attached to the body frame, respectively. $\overrightarrow{g}$ denotes the gravity vector. n_a denotes the Gaussian noise of the accelerometer.

Substituting (3) into (2), the dynamic model of the system is obtained as

$$\dot{p}^w = v^w$$
$$\dot{v}^w = R_b^w(a^m - n_a - \overrightarrow{g}) \qquad (4)$$

which could be transformed into the discrete form

$$\begin{bmatrix} p^w \\ v^w \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p^w \\ v^w \end{bmatrix}_k +$$
$$\begin{bmatrix} 0 \\ (R_b^w a_k^m - \overrightarrow{g})\Delta t \end{bmatrix} + \begin{bmatrix} 0 \\ -R_b^w n_a \Delta t \end{bmatrix} \qquad (5)$$

where Δt denotes the update cycle of the accelerometer.

2.2.3   Kalman filter model of the system

A popular model to fuse information from multiple sensors is the Kalman filter model, which consists of a state equation and an observation equation. The vision location results from Section 2.2.1 help to establish the observation equation, and the dynamic model of (5) helps to establish the state equation. A classic Kalman filter model is established to combine these two equations

$$\begin{cases} \begin{bmatrix} p^w \\ v^w \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p^w \\ v^w \end{bmatrix}_k + \\ \begin{bmatrix} 0 \\ (R_b^w a_k^m - \overrightarrow{g})\Delta t \end{bmatrix} + \begin{bmatrix} 0 \\ -R_b^w n_a \Delta t \end{bmatrix} \\ \\ Z_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} p^w \\ v^w \end{bmatrix}_k + \xi_k \end{cases} \qquad (6)$$

where $X = (p^w, v^w)^{\mathrm{T}}$ denotes the state vector to be estimated. The observation vector Z_k is the vision location result obtained in Section 2.2.1. ξ is the white Gaussian noise of the observation. Q_ξ is defined as the variance of ξ. In the Kalman filter, Q_ξ is needed for processing. However, as mentioned before, since the accuracy of the vision location algorithm strongly depends on the external environment, Q_ξ is uncertain, which is a common problem in many kinds of vision-assisted navigation systems. Therefore, adaptively identifying Q_ξ will certainly improve the accuracy of motion estimation.

3   Adaptive noise variance identification

The discrete observation signal $Z(k)$ is provided by the vision location algorithm at every 150 ms in our system. It consists of the real position signal $p(k)$ and the white Gaussian noise $\xi(k)$.

$$Z(k) = p(k) + \xi(k). \qquad (7)$$

The purpose of the ANVI is to identify Q_ξ from the observation signal $Z(k)$. First, some characteristic of $p(k)$ in the frequency domain will be derived. Then, it will be used to separate $p(k)$ and $\xi(k)$ so that Q_ξ could be identified.

Note that $Z(k)$, $p(k)$ and $\xi(k)$ are all actually 3-dimensional vectors in the x axis, y axis and z axis, but we only analyze one dimension of them because the quality in the other two dimensions are symmetry and quite the same.

3.1 Characteristic of position signal in frequency domain

To simplify the derivation, we first analyze $p(t)$ instead of $p(k)$, which is the continuous form of the position.

Before the derivation of the characteristic of $p(t)$ in the frequency domain, some kinematic properties of the vehicle need to be explained, i.e., the magnitude of the kinematic acceleration of the vehicle is upper-limited.

Firstly, in our experience, acceleration with a certain upper limit is enough for the vehicle to accomplish most of the automatic missions. During aerobatics, the aerial vehicle needs to fly rapidly, but for most of the common applications, such as hovering, following a certain trajectory, the vehicle flies gently.

Secondly, the limit of acceleration is actually a necessary condition for accurate attitude estimation. There are several kinds of methods for obtaining the measurement of attitude with onboard sensors. The sky line captured by the onboard camera was used to calculate the attitude[6], but the condition that the sky line needs to be unblocked and legible is strict. Yang et al.[15] used relative rotation calculated from image frames to update the attitude, but there will be accumulated error over time when the first reference image is out of view. In our system, we choose the traditional strap-down inertial attitude estimation method[16], which is popularly used in arial vehicles.

In strap-down inertial attitude estimation, the relationship between the gravity vector and the measurement of the 3-dimensional accelerometer provides an equation to calculate the direct cosine matrix of the attitude (the measurement of the magnetometer provides the other equation), as shown in (8).

$$A \times \overrightarrow{g} = \overrightarrow{a} \qquad (8)$$

where A denotes the direct cosine matrix, $\overrightarrow{g}$ is the gravity vector, and $\overrightarrow{a}$ is the measurement vector of the accelerometer.

However, (8) is correct only when the kinematic acceleration of the system could be ignored, because the accelerometer could measure both the gravity and the kinematic acceleration. Therefore, the kinematic acceleration will disturb the onboard algorithm which uses (8) for attitude estimation. The larger the acceleration of the vehicle is, the less accurate the attitude estimation result will be, which was explained in detail in [16].

Generally, by setting upper limits of the throttle, the roll angle and the pitch angle of the vehicle, the acceleration of the vehicle could be easily limited to

$$|\overrightarrow{a}| < 0.2|\overrightarrow{g}|. \qquad (9)$$

Now the position signal $p(t)$ will be analyzed by the short-time Fourier transform (STFT). The speed signal $v(t)$ and the acceleration signal $a(t)$ will be also processed for the following derivation.

First, in the STFT, signals need to be intercepted with a window. As shown in Fig. 6, $\widetilde{a}(t)$, $\widetilde{v}(t)$ and $\widetilde{p}(t)$ denote the signals intercepted from $a(t)$, $v(t)$ and $p(t)$ from t_1 to t_2, respectively.
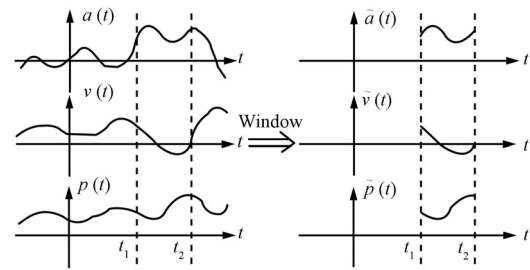

Fig. 6   Interception of signals

According to kinematic laws, the relationship between $\widetilde{a}(t)$, $\widetilde{v}(t)$ and $\widetilde{p}(t)$ is described as

$$\widetilde{p}(t) = \int_{-\infty}^{t} \widetilde{v}(t)\mathrm{d}t + p_1(u(t_1) - u(t_2))$$

$$\widetilde{v}(t) = \int_{-\infty}^{t} \widetilde{a}(t)\mathrm{d}t + v_1(u(t_1) - u(t_2)) \qquad (10)$$

where v_1 and p_1 denote the initial speed and position at time t_1, and $u(t)$ denotes the unit step function.

Then, using qualities of the Fourier transform (FT), one obtains

$$\widetilde{P}(\mathrm{j}\omega) = \frac{\widetilde{V}(\mathrm{j}\omega)}{\mathrm{j}\omega} + p_1 \times G(\mathrm{j}\omega) =$$

$$-\frac{\widetilde{A}(\mathrm{j}\omega)}{\omega^2} + \left(p_1 + \frac{v_1}{\mathrm{j}\omega}\right) \times G(\mathrm{j}\omega)$$

$$G(\mathrm{j}\omega) = \sqrt{\frac{\pi}{2}}(\mathrm{e}^{-\mathrm{j}\omega t_1} - \mathrm{e}^{-\mathrm{j}\omega t_2})\left(\frac{1}{\mathrm{j}\omega\pi} + \delta(\omega)\right). \qquad (11)$$

where $\widetilde{P}(\mathrm{j}\omega)$, $\widetilde{V}(\mathrm{j}\omega)$ and $\widetilde{A}(\mathrm{j}\omega)$ denote the FT of $\widetilde{p}(t)$, $\widetilde{v}(t)$ and $\widetilde{a}(t)$, or the STFT of $p(t)$, $v(t)$ and $a(t)$, respectively. $\delta(\omega)$ denotes the unit impulse function.

According to (9),

$$\left|\widetilde{A}(\mathrm{j}\omega)\right| = \left|\int_{t_1}^{t_2} \widetilde{a}(t)\mathrm{e}^{-\mathrm{j}wt}\mathrm{d}t\right| < 0.2g \times (t_2 - t_1). \qquad (12)$$

Therefore, according to (11) and (12),

$$\left|\widetilde{P}(\mathrm{j}\omega)\right| < \left|\frac{\widetilde{A}(\mathrm{j}\omega)}{\omega^2}\right| + \left|\left(p_1 + \frac{v_1}{\mathrm{j}\omega}\right) \times G(\mathrm{j}\omega)\right| <$$

$$\left|\frac{0.2g \times (t_2 - t_1)}{\omega^2}\right| +$$

$$\left|\left(p_1 + \frac{v_1}{\mathrm{j}\omega}\right)\right| \times \left|\sqrt{2\pi}\left(\frac{1}{\mathrm{j}\omega\pi} + \delta(\omega)\right)\right|. \qquad (13)$$

Before practical STFT processing of the position signal, the signal could be shifted so that the initial value $p_1 = 0$.

This will not affect the reconstruction of the signal in the time domain. Therefore, we obtain

$$\left| \widetilde{P}(\mathrm{j}\omega) \right| < \left| \frac{0.2g \times (t_1 - t_2)}{\omega^2} \right| + $$
$$\left| \frac{v_1}{\omega} \right| \times \left| \sqrt{2\pi} \left(\frac{1}{\mathrm{j}\omega\pi} + \delta(\omega) \right) \right|. \qquad (14)$$

In (13) and (14), the derivation of the inequalities is quite loose, which means the left-hand side part of the inequality is much smaller than the right-hand side part. And it is clear that when $|\omega|$ gets larger, the right-hand side part of inequality (14) will get smaller rapidly. In other words, the energy of the position signal in the frequency domain mostly distributes at the low frequency part.

The conclusion above is only qualitative. To make it quantitative, we collected sufficient position data during daily flight for frequency analysis experiment, and the results confirmed the conclusion and indicated that the energy of the position signal mostly distributes below 2 Hz. Details about the experiment will be described in Section 4.

In our system, the processing cycle of the vision algorithm is 150 ms. In other words, the sampling frequency from $p(t)$ to $p(k)$ is 6.66 Hz. According to the Nyquist-Shannon sampling theorem, the energy of the discrete signal $p(k)$ distributes mostly below 2 Hz and above 4.66 Hz, which is shown in Fig. 7.

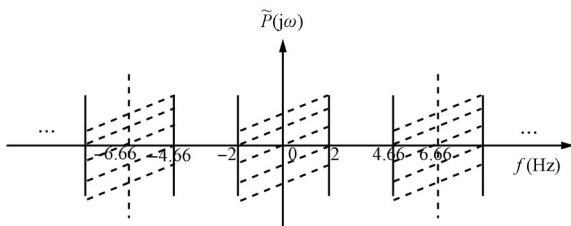Fig. 7   Affected frequency spectrum by sampling

3.2   Identification of variance

According to the analysis in Section 3.1, we can select a finite impulse response (FIR) bandpass digital filter $H(j\omega)$, whose pass band is between 2 Hz and 4.66 Hz (or included by this block). Let $h(k)$ denote the unit impulse response of the FIR filter $H(\mathrm{j}\omega)$, whose length is n. When we let the observation signal $Z(k)$ pass through this filter, the position signal $p(k)$ will be filtered out, the result signal is affected only by the noise signal $\xi(k)$.

$$\sum_{k=0}^{n-1} (h(k) \times Z(k + k_0)) = $$
$$\sum_{k=0}^{n-1} (h(k) \times (p(k + k_0) + \xi(k + k_0))) = $$
$$\sum_{k=0}^{n-1} (h(k) \times \xi(k + k_0)) \qquad (15)$$

where k_0 denotes the start of the signal sequence. As explained above, $\xi(k)$ is the white Gaussian noise, so

$$\mathrm{E}\{\xi(k_1) \times \xi(k_2)\} = \begin{cases} Q_\xi, & k_1 = k_2 \\ 0, & k_1 \neq k_2 \end{cases} \qquad (16)$$

where $\mathrm{E}\{\cdot\}$ denotes the expected value of the signal.

Then,

$$\mathrm{E}\left\{ \left(\sum_{k=0}^{n-1} (h(k) \times Z(k + k_0)) \right)^2 \right\} = $$
$$\mathrm{E}\left\{ \left(\sum_{k=0}^{n-1} (h(k) \times \xi(k + k_0)) \right)^2 \right\} = $$
$$\mathrm{E}\left\{ \sum_{k=0}^{n-1} (h^2(k) \times \xi^2(k + k_0)) \right\} = $$
$$Q_\xi \times \sum_{k=0}^{n-1} h^2(k). \qquad (17)$$

Therefore, we could obtain Q_ξ from

$$Q_\xi = \frac{\mathrm{E}\left\{ \left(\sum\limits_{k=0}^{n-1} (h(k) \times Z(k + k_0)) \right)^2 \right\}}{\sum\limits_{k=0}^{n-1} h^2(k)} \qquad (18)$$

where $\sum_{k=0}^{n-1} h^2(k)$ is a known value when we select a known digital filter. $(\sum_{k=0}^{n-1} (h(k) \times Z(k+k_0)))^2$ is the square of the the filtered result, which could be calculated. The expected value of the filtered result $\mathrm{E}\{(\sum_{k=0}^{n-1} (h(k) \times Z(k + k_0)))^2\}$ could be estimated by averaging a length of sample data. The length of the sample data is empirically selected. In practical, when the length is selected longer, the estimation of the expected value will be more accurate, but also cause a heavier computation burden.

4   Experiments and results

4.1   Frequency analysis of position signal

To prove the conclusion derived in Section 3.1 that the energy of the position signal mostly distributes in the low frequency part, and to make it quantitative, we collected sufficient position signals for frequency analysis.

An independent assistant visual system consisting of several fixed cameras is installed to collect pure position signal without noise (or nearly without noise). The position signal provided by this assistant system is highly accurate because heavy high definition cameras could be used here and the cameras don't suffer the dynamic of the vehicle. Motion information provided by this assistant system is regarded as the ground truth data for comparison.

The analysis of collected signals supported the conclusion and indicated that the energy of the position signal

distributes mostly below 2 Hz. A result of the analysis is shown in Fig. 8.
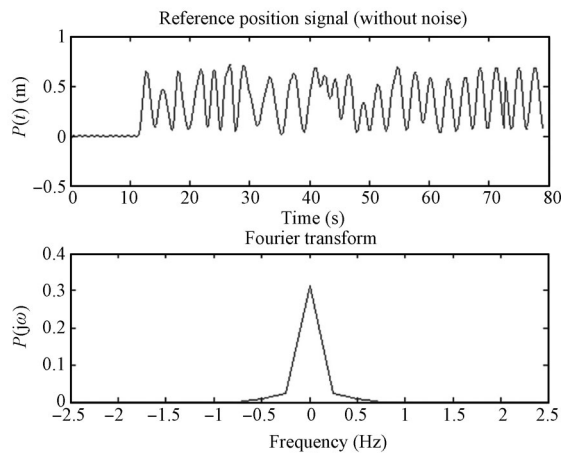

Fig. 8   Frequency spectrum of the position signal

4.2   Simulation of adaptive noise variance identification

To verify the effectiveness of the ANVI method, we add the position signal provided by the assistant visual system with white Gaussian noise whose variance is known. The position signal in Fig. 8 was mixed with different kinds of white Gaussian noises whose variance were 5 cm, 10 cm, 15 cm and 20 cm, respectively. Fig. 9 shows the estimation results of the variance with the proposed method, which is reliable and this result is practical for further motion estimation.
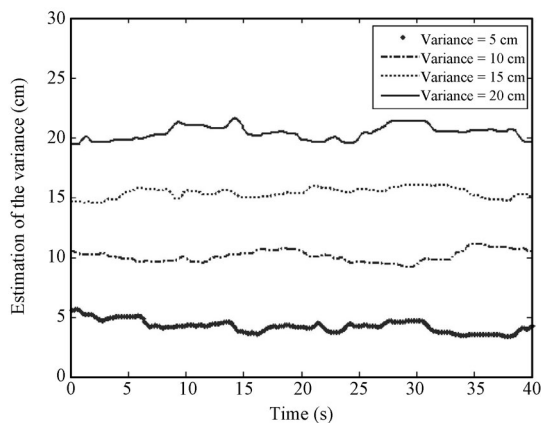

Fig. 9   Simulation: identification of the variance

In practical applications, the variance of the noise does not generally change very fast. So, to lower the computational burden of the system, the estimation of the variance is not executed in each data sample cycle as shown in Fig. 9, but every 5–10 s.

4.3   Improvement of motion estimation

The identified variance Q_ξ of the observation noise is then used in the motion estimation based on the Kalman filter

model in (6). As shown in Fig. 10, the adaptively identified variance is displayed in the top picture. Additionally, the motion estimation results with this adaptively identified variance are displayed in the next two pictures.
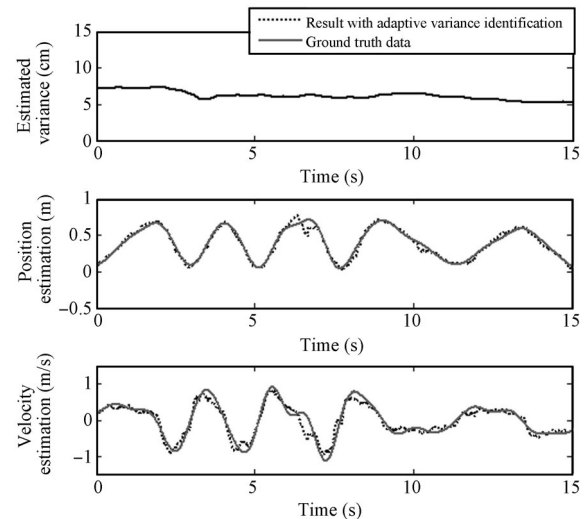

Fig. 10   Motion estimation results with identified variance

Two groups of motion estimation experiments with fixed variance were done for comparison. The first group of comparative experiments was done with fixed variance that is smaller than the true variance (variance = 0.01 m and variance = 0.03 m), with the results shown in Fig. 11. The second group of comparative experiments was done with fixed variance that is larger than the true variance (variance = 0.15 m and variance = 0.2 m), with the results shown in Fig. 12. The root-mean-square errors of all the estimated results compared with the ground truth data are shown in Table 1.
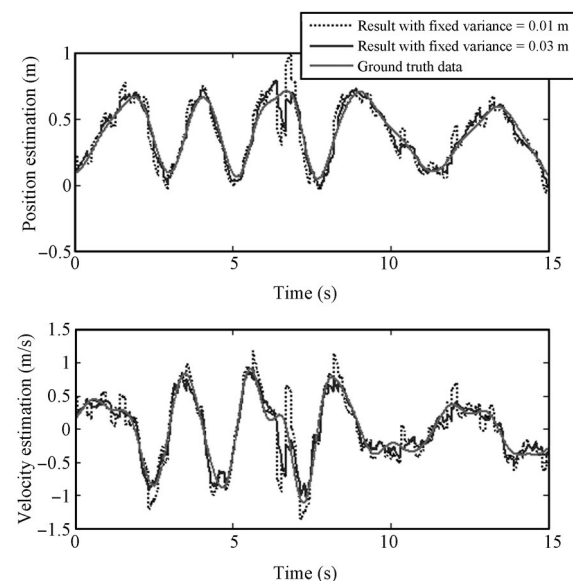

Fig. 11   Motion estimation results with fixed variance smaller than true variance
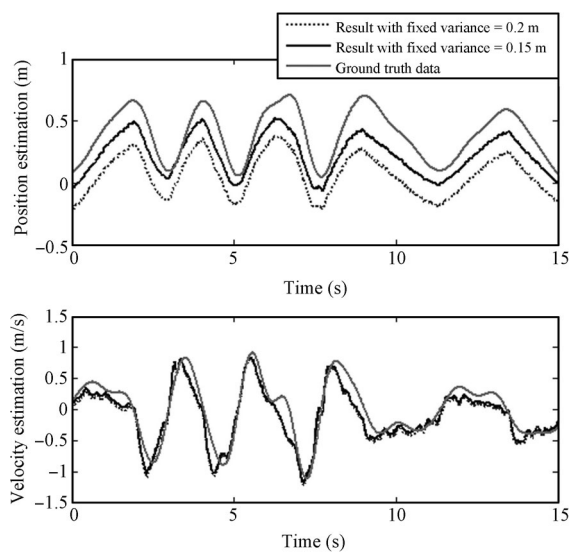
Fig. 12　Motion estimation results with fixed variance larger than true variance

As shown in Table 1, the motion estimation results with ANVI obtained the highest accuracy. During the experiment, the adaptively identified variance was about $0.05\,\mathrm{m}$ to $0.1\,\mathrm{m}$. A larger deviation between the empiric variances and the truth variance caused a larger error in the results, no matter the empiric variances are larger (refer to the truth that results with $Q_\xi = 0.2\,\mathrm{m}$ suffered bigger error than results with $Q_\xi = 0.15\,\mathrm{m}$) or smaller (refer to the truth that results with $Q_\xi = 0.01\,\mathrm{m}$ suffered bigger error than results with $Q_\xi = 0.03\,\mathrm{m}$) than the truth variance. Besides, in general, when an empiric variance gets much larger than the real variance, the estimation result will be mainly affected by the state equation, therefore the accumulated error could not be bounded effectively, which will cause big error in the position estimation (refer to the results with $Q_\xi = 0.2\,\mathrm{m}$). On the contrary, when the empiric variance gets much smaller than the real one, the estimation result will be mainly affected by the observation equation, which caused unsmooth of the results and lowered the accuracy of the velocity (refer to the results with $Q_\xi = 0.01\,\mathrm{m}$).

Table 1　Root-mean-square (RMS) errors of the estimation results

RMS error	Position (m)	Velocity (m/s)
$Q_\xi = 0.2\,\mathrm{m}$	0.273	0.176
$Q_\xi = 0.15\,\mathrm{m}$	0.152	0.158
ANVI	0.034	0.117
$Q_\xi = 0.03\,\mathrm{m}$	0.046	0.119
$Q_\xi = 0.01\,\mathrm{m}$	0.081	0.207

5　Conclusions

A novel adaptive variance identification method is proposed in this paper. Experiment shows that with this method, the variance of the noise could be identified reliably. With the ANVI method, the results of the motion es-

timation will basically be optimal. On the contrary, results will become worse when there is a large deviation between the empirical variance value and the truth value. In some application environments, the empirical variance value is really hard to choose, and the ANVI therefore becomes a necessary component for motion estimation.

The method is especially suitable for the vision-assisted motion estimation of UAVs. Because firstly, the noise of vision location results is changeable and needs adaptive identification. Secondly, the validity of the ANVI method is based on the special kinematic property of UAVs, as explained in the derivation. However, in our experience, many kinds of robots working on ground, in water or underwater may share the same kinematic properties that the kinematic acceleration has a upper limit. Therefore, the application of the proposed method could be expanded.

References

[1] C. S. Yoo, I. K. Ahn. Low cost GPS/INS sensor fusion system for UAV navigation. In *Proceedings of the 22nd Digital Avionics Systems Conference*, IEEE, Indianapolis, USA, 2003.

[2] F. Aghili, M. Kuryllo, G. Okouneva, C. English. Fault-tolerant position/attitude estimation of free-floating space objects using a laser range sensor. *IEEE Sensors Journal*, vol. 11, no. 1, pp. 176–185, 2011.

[3] J. F. Vasconcelos, C. Silvestre, P. Oliveira, B. Guerreiro. Embedded UAV model and LASER aiding techniques for inertial navigation systems. *Control Engineering Practice*, vol. 18, no. 3, pp. 262–278, 2010.

[4] L. Whitcomb, D. Yoerger, H. Singh. Advances in doppler-based navigation of underwater robotic vehicles. In *Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE, Detroit, MI, USA, pp. 399–406, 1999.

[5] H. Zhao, Z. Y. Wang. Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended Kalman filter for data fusion. *IEEE Sensors Journal*, vol. 12, no. 5, pp. 943–953, 2012.

[6] I. Mondragón, M. Olivares-Méndez, P. Campoy, C. Martínez, L. Mejias. Unmanned aerial vehicles UAVs attitude, height, motion estimation and control using visual systems. *Autonomous Robots*, vol. 29, no. 1, pp. 17–34, 2010.

[7] B. Herisse, F. X. Russotto, T. Hamel, R. Mahony. Hovering flight and vertical landing control of a VTOL unmanned aerial vehicle using optical flow. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Nice, France, pp. 801–806, 2008.

[8] M. Bošnak, D. Matko, S. Blažič. Quadrocopter hovering using position-estimation information from inertial sensors and a high-delay video system. *Journal of Intelligent & Robotic Systems*, vol. 67, pp. 43–60, 2012.

[9] C. L. Wang, T. M. Wang, J. H. Liang, Y. C. Zhang, Y. Zhou. Bearing-only visual SLAM for small unmanned aerial

vehicles in GPS-denied environments. *International Journal of Automation and Computing*, vol. 10, no. 5, pp. 387–396, 2013.

[10] F. Zhou, W. Zheng, Z. F. Wang. Adaptive noise variance identification in vision-aided motion estimation. In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods*, Angers, France, pp. 1–7, 2014.

[11] R. Szeliski. *Computer Vision: Algorithms and Applications*, New York, USA: Springer, pp. 347–348, 2010.

[12] M. Achtelik, M. Achtelik, S. Weiss, R. Siegwart. Onboard IMU and monocular vision based control for MAVs in unknown in- and outdoor environments. In *Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE, Shanghai, China, pp. 3056–3063, 2011.

[13] E. Edwan, J. Y. Zhang, J. C. Zhou, O. Loffeld. Reduced DCM based attitude estimation using low-cost IMU and magnetometer triad. In *Proceedings of the 8th Positioning Navigation and Communication*, IEEE, Dresden, Germany, pp. 1–6, 2011.

[14] J. Artieda, J. M. Sebastian, P. Campoy, J. F. Correa, I. F. Mondragn, C. Martnez, M. Olivares. Visual 3-d SLAM from UAVs. *Journal of Intelligent and Robotic Systems*, vol. 55, no. 4–5, pp. 299–321, 2009.

[15] 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.

[16] H. G. de Marina, F. J. Pereda, J. M. Giron-Sierra, F. Espinosa. UAV attitude estimation using unscented Kalman filter and TRIAD. *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4465–4474, 2012.
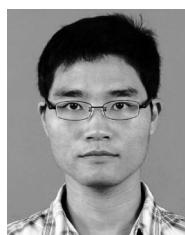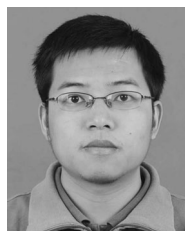
Fan Zhou   received the B. Eng. degree from University of Science and Technology of China, China in 2009. He is now a Ph. D. candidate in pattern recognition and intelligent system, University of Science and Technology of China, China.

His research interests include unmanned aerial robot, integrated navigation, adaptive signal processing and visual simultaneous localization and mapping.

E-mail: zhoufan@mail.ustc.edu.cn

ORCID iD: 0000-0003-1562-2470

Wei Zheng   received the B. Eng. degree from University of Science and Technology of China, China in 2009. He is now a Ph. D. candidate in pattern recognition and intelligent system, University of Science and Technology of China, China.

His research interests include unmanned aerial vehicle, robot localization and navigation, visual simultaneous localization and mapping and multi-sensor fusion.

E-mail: zhengwei@mail.ustc.edu.cn

Zeng-Fu Wang   received the B. Sc. degree in electronic engineering from University of Science and Technology of China, China in 1982, and the Ph. D. degree in control engineering from Osaka University, Japan in 1992. He is currently a professor of both Institute of Intelligent Machines, Chinese Academy of Sciences and University of Science and Technology of China. He has published more than 180 journal articles and conference papers.

His research interests include computer vision, human computer interaction and intelligent robots.

E-mail: zfwang@ustc.edu.cn (Corresponding author)

ORCID iD: 0000-0003-1859-900X