Adaptive Noise Variance Identification in Vision-aided Motion Estimation for UAVs

Fan Zhou[1], Wei Zheng[1] and Zengfu Wang[2]
[1]*Department of Automation, University of Science and Technology of China, Hefei, China*
[2]*Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China*

Keywords: Adaptive Noise Variance Identification, Vision Location, Motion Estimation, Kalman Filter.

Abstract: Vision location 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 the white gaussian noise so that this 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. In this paper, a novel adaptive noise variance identification (ANVI) method is proposed, which utilizes the special kinematic property of the UAV for frequency analysis and adaptively identify the variance of the noise. Then, the adaptively identified variance are used in the kalman filter for 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.

1 INTRODUCTION

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

Many research works have been done in this field, using various kinds of location sensors including GPS (Yoo and Ahn, 2003), laser range sensors (Farhad et al., 2011; Vasconcelos et al., 2010), doppler radars (Whitcomb et al., 1999), ultrasonic sensors (Zhao and Wang, 2012), etc. However, the factors of accuracy, weight, cost, and applicable environment limit the application of these sensors on aerial vehicles. Vision sensors, with advantage in these aspects, therefore have become a popular choice for providing location results of the system (Mondragón et al., 2010).

The kalman filter model is widely used to obtain accurate,fast updated and reliable motion estimation of the UAV system (Zhao and Wang, 2012; Chatila et al., 2008; Bosnak et al., 2012), which generally consists of two equations: the observation equation and the state equation. Results directly provided by the vision location 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. The variance of the noise of inertial sensors is basically statical and could be obtained from the hardware data sheet. However, in vision location methods, feature detecting and matching component is usually included, whose accuracy is obviously affected by external environment, such as illumination, camera resolution, texture of environment, height of flight etc. Therefore, the variance of the noise of vision location results is changeable and simply setting an empirical parameter of it will probably lower the accuracy of the motion estimation.

In this article, some special kinematic properties of the UAV system, which were barely utilized before, are observed and utilized for frequency analysis of the position signal (or the trajectory) of the vehicle. Derivation shows that the position signal have some characteristic in the frequency domain which helps to separate it from the noise and therefore the variance of the noise could be identified.

The rest of the paper is organized as follows. Section 2 introduces the entire system where the ANVI method is applied, including the configuration of the UAV system and the principle of the motion estima-

Zhou F., Zheng W. and Wang Z..
Adaptive Noise Variance Identification in Vision-aided Motion Estimation for UAVs.
DOI: 10.5220/0004921707530758
In *Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods* (ICPRAM-2014), pages 753-758
ISBN: 978-989-758-018-5

tion. Section 3 proposed the ANVI method detailedly. Experiment and results are shown in section 4 which verified 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 Figure 1. The main sensors onboard is a downward looking monocular camera, a height sensor and an IMU unit, including an accelerometer, a gyroscope, and a magnetometer. The wireless link is used to share information with the ground PC computer, where the vision location algorithm is processing. An strap-down inertial attitude estimation algorithm is executed on the onboard micro controller. In the algorithm, the gravity is measured by the 3-dimensional accelerometer, which helps to establish the observation equation of the attitude. Additionally, the angular velocity measured by the gyroscope helps to establish the propagation equation of the attitude. The two equations constitute the kalman model for the attitude estimation. Although we don't discuss 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.

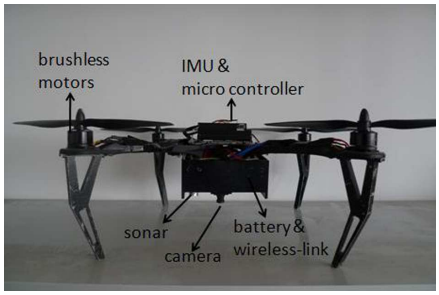

Figure 1: The configuration of the vehicle.

2.2 Principle of Motion Estimation

For automatical application, states of motion need to be estimated, generally including 3D position p and 3D translational speed v, which form a 6-dimensional state vector $X = (p, v)^T$. Figure 2 shows the entire framework of the states estimation in this paper. The unique component of ANVI will be derived in Section 3. The other components, as general parts of the motion estimation, will be described in Section 2.

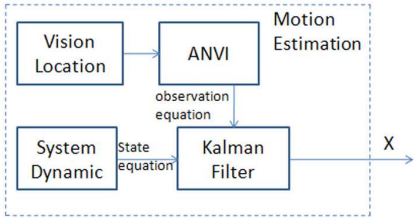

Figure 2: The framework of motion estimation.

2.2.1 Markless Vision Location

Figure 3 shows the environment of visual observation. The first frame captured by the camera is set as the reference frame. Then, every frame captured after will be compared with the reference frame. After feature detecting and matching, which had been widely studied in computer vision (R.Szeliski, 2010), a location method could be presented with the pairs of corresponded points.

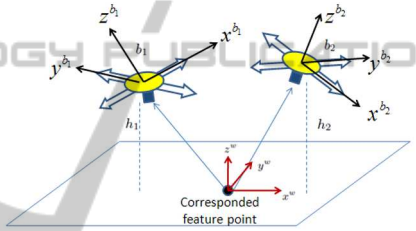

Figure 3: Vision location.

As shown in Figure 3, the vehicle moved from location 1 to location 2 (the attitude changed too), and the related body frame are b_1 and b_2, respectively. b_1 is where the reference video frame is captured. A world coordinate frame is established with the center point coincides with one of the corresponded feature points $p_0^w = (x_0^w, y_0^w, z_0^w)^T$.

With the knowledge of image formation process, the coordinate of p_0^w in the world frame could be transformed into the coordinate in the image:

$$s_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = M \begin{bmatrix} R_1 & t_1 \end{bmatrix} \begin{bmatrix} x_0^w \\ y_0^w \\ z_0^w \\ 1 \end{bmatrix} \quad (1)$$

Where $(u_1, v_1)^T$ denotes the coordinate in the image. M denote the intrinsics matrix of the camera, which could be obtained through calibration. R_1 and t_1 denote the rotation and translation from the world frame to the camera frame, respectively. In the vehicle system, the rotation matrix R_1 could be obtained by onboard strap-down inertial navigation system (Savage, 1998; Edwan et al., 2011; de Marina et al., 2012) so it could be treated as a known matrix.

s_1 is a scale factor. $(x_0^w, y_0^w, z_0^w)^T$ denote the coordinate of the feature point in the world frame. Since the world frame is established with p_0^w as the center point, $(x_0^w, y_0^w, z_0^w)^T = (0,0,0)^T$. Substituting it in (2), one obtains

$$t_1 = s_1 M^{-1} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \tag{2}$$

Another equation could be established with the measurement of the height sensor:

$$\left[R_1^{-1} t_1\right]_3 = h_1 \tag{3}$$

Where the subscript 3 denotes the third element of the vector. R_1^{-1} is the inverse of R_1 and denotes the rotation from the body frame to the world frame. Obviously $R_1^{-1} t_1$ is the location of the camera in the world frame. h_1 is the measurement of the height sensor.

Using (2) and (3), we have four equations with four undetermined parameters (scale factor s_1 and translation vector t_1). The equations could be solved and then the location of the camera in the world frame $p_{c_1}^w$ could be represent as:

$$p_{c_1}^w = R_1^{-1} t_1 \tag{4}$$

When the vehicle moved to location 2, as shown in Figure 3, the location $p_{c_2}^w$ could be obtained with the same equations presented above. Then, the relative 3D position between location 1 and 2 could be easily calculated:

$$p = p_{c_2}^w - p_{c_1}^w \tag{5}$$

The position result provided here by the vision location 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

The dynamic equation of the system motion is given by

$$\begin{aligned} \dot{\boldsymbol{p}}^w &= \boldsymbol{v}^w \\ \dot{\boldsymbol{v}}^w &= \boldsymbol{a}^w \end{aligned} \tag{6}$$

Where $\boldsymbol{p}^w = (x^w, y^w, z^w)$ denotes the position in the world frame. $\boldsymbol{v}^w$, $\boldsymbol{a}^w$ denotes the velocity and acceleration in the world frame, respectively. Define R_b^w as the rotation matrix from the body frame to the world frame. Then the acceleration in the world frame could be obtained from

$$\begin{aligned} \boldsymbol{a}^w &= R_b^w \boldsymbol{a}^b \\ \boldsymbol{a}^b &= \boldsymbol{a}^m - \boldsymbol{n}_a - \overrightarrow{\boldsymbol{g}} \end{aligned} \tag{7}$$

Substituting (7) into (6), one obtains the dynamic model of the system:

$$\begin{aligned} \dot{\boldsymbol{p}}^w &= \boldsymbol{v}^w \\ \dot{\boldsymbol{v}}^w &= R_b^w (\boldsymbol{a}^m - \boldsymbol{n}_a - \overrightarrow{\boldsymbol{g}}) \end{aligned} \tag{8}$$

Which could be transformed into discrete form:

$$\begin{bmatrix} p^w \\ v^w \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p^w \\ v^w \end{bmatrix}_k + \begin{bmatrix} 0 \\ (R_b^w a_k^m - \overrightarrow{g})\Delta t \end{bmatrix} + \begin{bmatrix} 0 \\ -R_b^w n_a \Delta t \end{bmatrix} \tag{9}$$

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 helps to establish the observation equation and the dynamic model helps to establish the state equation. A classic kalman filter model is established combining these two equations:

$$\begin{cases} \begin{bmatrix} p^w \\ v^w \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p^w \\ v^w \end{bmatrix}_k + \begin{bmatrix} 0 \\ (R_b^w a_k^m - \overrightarrow{g})\Delta t \end{bmatrix} + \begin{bmatrix} 0 \\ -R_b^w n_a \Delta t \end{bmatrix} \\ Z_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{p}^w \\ \boldsymbol{v}^w \end{bmatrix}_k + \xi_k \end{cases} \tag{10}$$

Where $X = (p^w, v^w)^T$ 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. In the kalman filter, the variance of the noise is needed for processing. Define Q_ξ as the variance of ξ. Since the accuracy of the vision location algorithm strongly depends on external environment, Q_ξ needs to be adaptively identified to improve the accuracy of estimation.

3 ADAPTIVE NOISE VARIANCE IDENTIFICATION

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

$$Z(k) = p(k) + \xi(k) \tag{11}$$

3.1 Characteristic of the Position Signal in the 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. That is, the magnitude of the kinematic acceleration of the vehicle is upper limited.

On one hand, acceleration with a certain upper limit is enough for the vehicle to accomplish most of the automatical missions. On the other hand, limit of acceleration is actually a necessary condition for accurate attitude estimation. It is easy to understander that the kinematic acceleration will disturb the attitude algorithm which uses the gravity for attitude estimation. The larger the acceleration of the vehicle is, the less accurate the attitude estimation result will be, which was detailedly explained in (de Marina et al., 2012). Generally, by setting a upper limit of the throttle, the roll angle and the pitch angle of the vehicle, the acceleration is limited to:

$$|a| < 0.2g \tag{12}$$

Now the position signal $p(t)$ will be transformed from the time domain to the frequency domain by the Short-time Fourier transform (STFT). The speed signal $v(t)$ and the acceleration signal $a(t)$ will be used in the following derivation, so they are also processed here.

First, in the STFT, signals need to be intercepted with a window. As shown in figure 4, $\widetilde{a}(t)$, $\widetilde{v}(t)$ and $\widetilde{p}(t)$ denote the signal intercepted from $a(t)$, $v(t)$ and $p(t)$ from t_1 to t_2, respectively.

According to kinematic laws, the relation ship between $\widetilde{a}(t)$, $\widetilde{v}(t)$ and $\widetilde{p}(t)$ is described in equation (13):

$$\begin{aligned} \widetilde{p}(t) &= \int_{-\infty}^{t} \widetilde{v}(t)dt + p_1(u(t_1 - u(t_2))) \\ \widetilde{v}(t) &= \int_{-\infty}^{t} \widetilde{a}(t)dt + v_1(u(t_1 - u(t_2))) \end{aligned} \tag{13}$$

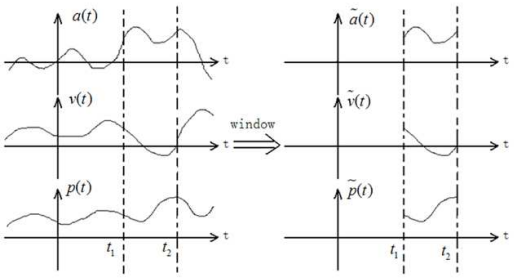

Figure 4: Interception of signals.

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

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

$$\begin{aligned} \widetilde{P}(j\omega) &= \frac{\widetilde{V}(j\omega)}{j\omega} + p_1 \cdot G(j\omega) \\ &= -\frac{\widetilde{A}(j\omega)}{\omega^2} + (p_1 + \frac{v_1}{j\omega}) \cdot G(j\omega) \end{aligned} \tag{14}$$

where

$$G(j\omega) = \sqrt{\frac{\pi}{2}}(e^{-j\omega t_1} - e^{-j\omega t_2})(\frac{1}{j\omega\pi} + \delta(\omega))$$

Where $\widetilde{P}(j\omega)$, $\widetilde{V}(j\omega)$ and $\widetilde{A}(j\omega)$ 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 (12),

$$\left|\widetilde{A}(j\omega)\right| = \left|\int_{t_1}^{t_2} \widetilde{a}(t)e^{-jwt}dt\right| < 0.2g \cdot (t_2 - t_1) \tag{15}$$

Therefore, according to (14) and (15)

$$\begin{aligned} \left|\widetilde{P}(j\omega)\right| &< \left|\frac{\widetilde{A}(j\omega)}{\omega^2}\right| + \left|(p_1 + \frac{v_1}{j\omega}) \cdot G(j\omega)\right| \\ &< \left|\frac{0.2g \cdot (t_2 - t_1)}{\omega^2}\right| + \\ &\quad \left|(p_1 + \frac{v_1}{j\omega})\right| \cdot \left|\sqrt{2\pi}(\frac{1}{j\omega\pi} + \delta(\omega))\right| \end{aligned} \tag{16}$$

Before practical STFT processing of the position signal, the signal could be shifted so that the initial value $p_1 = 0$, this won't affecting the reconstruction of the signal in the time domain. Therefore we obtain

$$\begin{aligned} \left|\widetilde{P}(j\omega)\right| &< \left|\frac{0.2g \cdot (t_1 - t_2)}{\omega^2}\right| + \\ &\quad \left|\frac{v_1}{\omega}\right| \cdot \left|\sqrt{2\pi}(\frac{1}{j\omega\pi} + \delta(\omega))\right| \end{aligned} \tag{17}$$

Equation (16) indicates the energy of the position signal in the frequency domain mostly distribute at the low frequency part. 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 distribute below 2Hz.

3.2 Identification of the Variance

According to the analysis in Section 3.1, we can Select a FIR bandpass digital filter $H(j\omega)$, whose pass band is above 2Hz. Let $h(k)$ denotes the unit impulse response of the FIR filter $H(j\omega)$, whose length is n. When we let the observation signal $Z(k)$ through this filter, the position signal $p(k)$ will be filtered out, the result signal is affected only by the noise signal $\xi(k)$:

$$\begin{aligned}&\sum_{k=0}^{n-1}(h(k)\cdot Z(k+k_0))\\ =&\sum_{k=0}^{n-1}(h(k)\cdot(p(k+k_0)+\xi(k+k_0)))\\ =&\sum_{k=0}^{n-1}(h(k)\cdot\xi(k+k_0))\end{aligned} \tag{18}$$

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

$$E\{\xi(k_1)\cdot\xi(k_2)\}=\begin{cases}Q_\xi, k_1=k_2\\ 0, k_1\neq k_2\end{cases} \tag{19}$$

Where $E\{\cdot\}$ denote the expected value of the signal.

Therefore

$$\begin{aligned}&E\{(\sum_{k=0}^{n-1}(h(k)\cdot Z(k+k_0)))^2\}\\ =&E\{(\sum_{k=0}^{n-1}(h(k)\cdot\xi(k+k_0)))^2\}\\ =&E\{\sum_{k=0}^{n-1}(h^2(k)\cdot\xi^2(k+k_0))\}\\ =&Q_\xi\cdot\sum_{k=0}^{n-1}h^2(k)\end{aligned} \tag{20}$$

Therefore, we could obtain Q_ξ from:

$$Q_\xi=\frac{E\{(\sum_{k=0}^{n-1}(h(k)\cdot Z(k+k_0)))^2\}}{\sum_{k=0}^{n-1}h^2(k)} \tag{21}$$

Where $\sum_{k=0}^{n-1}h^2(k)$ is a known value when we selected a known digital filter. $(\sum_{k=0}^{n-1}(h(k)\cdot Z(k+k_0)))^2$ is the Square of the the filtered result, which could be calculated. The expected value of the filtered result $E\{(\sum_{k=0}^{n-1}(h(k)\cdot Z(k+k_0)))^2\}$ could be estimated by average 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 causes heavier computation burden.

4 EXPERIMENTS AND RESULTS

4.1 Simulation of Adaptive Noise Variance Identification

To verify the effectiveness of the ANVI method, we plus the position signal provided by the assistant visual system with white gaussian noise whose variance is known. The position signalwas mixed with different kinds of white gaussian noise whose variance were 5cm, 10cm, 15cm and 20cm, respectively. Figure 5 shows the estimation result of the variance with the proposed method, which is reliable and this result is practical for further motion estimation.

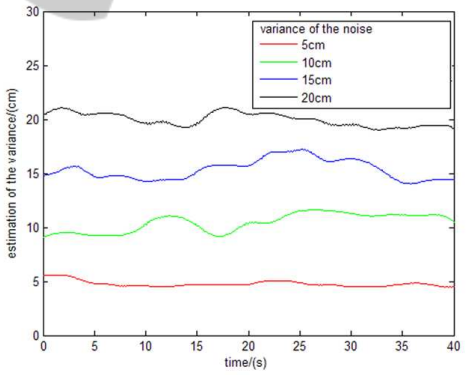

Figure 5: Simulation: identification of the variance.

In practical application, the variance of the noise generally changes not very fast. So to lower the computational burden of the system, the estimation of the variance is not executed every data sample cycle as shown in Figure 5, but every 5-10 seconds.

4.2 Improvement of the 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 equation (10). The results of

the motion estimation with a fixed empiric variance value and with the ANVI method are recorded for comparison.

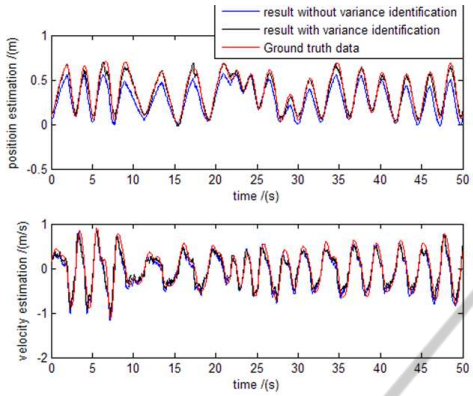

Figure 6: Result of the motion estimation.

As shown in Figure 6, motion estimation includes position estimation and velocity estimation, and the accuracy of both improved with the ANVI method. The root-mean-square error of the results compared with the groudtruth data is shown in Table 1.

Table 1: Root-mean-square error of the results in Figure 7.

RMS error	position(m)	velocity(m/s)
without ANVI	0.110	0.151
with ANVI	0.037	0.105

5 CONCLUSIONS

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

The method is especially suitable in the vision-aided 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 properties of UAVs, as explained in the derivation. However, since most kinds of robot share the same kinematic properties that the kinematic acceleration has a upper limit. Therefore, with certain adjustments of the parameters, the proposed method could be used in wide applications.

ACKNOWLEDGEMENTS

The work in this paper was supported by the National Science and Technology Major Projects of the Ministry of Science and Technology of China: ITER (No.2012GB102007).

REFERENCES

Bosnak, M., Matko, D., and Blazic, S. (2012). Quadrocopter hovering using position-estimation information from inertial sensors and a high-delay video system. *Journal of Intelligent & Robotic Systems*, 67(1):43–60.

Chatila, R., Kelly, A., and Merlet, J, P. (2008). Hovering flight and vertical landing control of a vtol unmanned aerial vehicle using optical flow. In *IEEE/RSJ IROS*, pages 801–806. IEEE.

de Marina, H, G., Pereda, F, J., Giron-Sierra, J, M., and Espinosa, F. (2012). Uav attitude estimation using unscented kalman filter and triad. *IEEE Transactions on Industial Electronics*, 59(11):4465–4474.

Edwan, E., Zhang, J, Y., and Zhou, J, C. (2011). Reduced dcm based attitude estimation using low-cost imu and magnetometer triad. In *Proceedings of the 8th WPNC*, pages 1–6. IEEE.

Farhad, A., Marcin, K., and Galina, O. (2011). Fault-tolerant position/attitude estimation of free-floating space objects using a laser range sensor. *IEEE Sensors Journal*, 11(1):176–185.

Mondragón, I., Olivares-Méndez, M., Campoy, P., Martínez, C., and Mejias, L. (2010). Unmanned aerial vehicles uavs attitude, height, motion estimation and control using visual systems. *Autonomous Robots*, 29:17–34.

R.Szeliski (2010). *Computer Vision: Algorithms and Applications*. Springer, USA, 1nd edition.

Savage, P, G. (1998). Strapdown inertial navigation integration algorithm design part 1: attitude algorithms. *Journal of Guidance, Control, and Dynamics*, 21(1):19–28.

Vasconcelos, J, F., Silvestre, C., and Oliveira, P. (2010). Embedded uav model and laser aiding techniques for inertial navigation systems. *Control Engineering Practice*, 18(3):262–278.

Whitcomb, L., Yoerger, D., and Singh, H. (1999). Advances in doppler-based navigation of underwater robotic vehicles. In *Proceedings of ICRA*, volume 1, pages 399–406. IEEE.

Yoo, C, S. and Ahn, I, K. (2003). Low cost gps/ins sensor fusion system for uav navigation. In *Proceedings of the 22nd DASC*, volume 2, pages 8.A.1–8.1–9. IEEE.

Zhao, H. and Wang, Z, Y. (2012). Motion measurement using inertial sensors, ultrasonic sensors, and magnetometers with extended kalman filter for data fusion. *IEEE Sensors Journal*, 12(5):943–953.