Homography Based State Estimation for Aerial Robots

Jakob Schwendner

DFKI Robotics Lab
Robert-Hooke-Str. 5, D-28359, Bremen, Germany

Abstract. Low-cost Unmanned Aerial Vehicles have large potential for applications in the civil sector. Cheap inertial sensors alone can not provide the degree of accuracy required for control and navigation of the UAVs. Additional sensors, and sensor fusion techniques are needed to reduce the state estimation error. In the context of this work, the possibility of using visual information to augment the inertial data is investigated. The projected image of the ground plane is processed by a homography constrained optical flow description and used as a measurement update for an EKF. A simulation environment is used, which models the UAV dynamics, the rate-gyro and accelerometer errors as well as a low resolution vision system.

1 Introduction

Over the recent years Unmanned Aerial Vehicles (UAV) have received an increased academic, commercial and military interest. A 2004 report by NASA [4] projects one third of US Airforce aircraft to be unmanned by 2010. UAVs current main application is in the military sector, and most of the current development effort is geared towards it. Decreasing costs in Sensor and Embedded System technology however make UAV technology an interesting prospect for civil use.

In order to be applicable to civil applications, the technology at use must be safe, reliable and also cost effective. An increased level of autonomy can help with this. One of the main challenges of operating an autonomous aerial robot is the sensing and situation awareness part. It combines multi-sensory input into an internal representation of the current system state and puts it into relation to known world facts and mission objectives.

Low-cost inertial sensors can provide the basis for a state estimation system. Due to the nature of the integration process however they are prone to extensive drift. Often GPS systems provide absolute references to compensate the drift. A different approach which is also followed by this work is to use visual means to aid the state estimation process. A method is described which constraints the optical flow using a homography transformation of the perceived ground plane. Further, an Extended Kalman Filter is developed where the state transition is based on a generic rigid body model using the inertial sensors as system inputs, and the homography matrix is used to perform a measurement update.

The validity of the approach is tested using a simulation environment. A scenario is presented where the homography based update is compared to inertial state estimation alone.

A. Dengel et al. (Eds.): KI 2008, LNAI 5243, pp. 332–339, 2008.

2 Related Work

A number of ways have been investigated to augment inertial navigation with visual sensors to minimise localisation drift. An important concept in this context is the *motion field*, which is the projection of 3-D points and their velocity vectors onto a 2-D plane [9]. Egomotion, the motion of the observer in space, can be partly recovered due to the nature of the motion field [2]. If the speed of the observer is known, the motion field can also be used to calculate distance information, which is very helpful for obstacle avoidance [10]. *Optical flow* methods can under certain conditions be used to recover the motion field information. The optical flow is specified as the vector field that maps the brightness of points in one image to the brightness values in the next image in the sequence. The field is under-specified and suffers from what is known as the aperture problem. Various algorithms can be used to calculate an approximation of the motion field based on an image sequence. Because of the under-specified nature of the problem, these methods are recursive in nature and computationally expensive. A good overview of the different algorithms and their performances is given by Galvin et al. [5].

A number of different approaches have been used to apply optical flow measurements to UAV navigation and control. Barrows and Neely [1] have proposed a hardware centred approach in which an array of photoreceptors is used to measure the optical flow, thus making it possible to be used in micro aerial vehicles (MAV) with a weight requirement of less than a gram per sensor. A similar approach is used by Stocker [14], who has successfully demonstrated a solution for a 2D optical flow sensor in silicon.

A simulated environment is used by Muratet et al. [11] to test the application of optical flow in navigating urban canyons. The simulation assumes the presence of a sensor array for inertial, GPS and altimeter data. This data is combined with optical flow information in a traditional control scheme using PID controllers in order to avoid obstacles.

The previously shown examples for using optical flow information for controlling UAVs were mostly based for collision detection. However, when assuming a camera motion relative to a plane, it is possible to express the optical flow as a combination of orientation, linear velocity, rotational velocity and distance to the plane [12]. The optical flow field can thus be expressed as a 3x3 homography matrix, relating the projective coordinates from one frame to the next.

A homography based approach for landing is used by Bosse et al. [2], who propose an algorithm to fuse the information with existing state information with the help of a Kalman filter. The algorithm uses image sequences for which it calculates the time derivative of the projection matrix, which is equal to the optical flow. The global smoothness constraint of [8] is replaced by the quadratic function of image location, relating the optical flow at a certain image point to the derivative of the projection matrix. The system state values are extracted using eigenvector decomposition. The system's current estimate of altitude is used to solve the scale ambiguities.

3 Method

The system model is defined by its state transition function $f(x)$ and the measurement function $h(x)$. The environment is modeled as an inertial right-handed reference frame $\mathcal{I} = \{E_1, E_2, E_3\}$, with the positive base $e_2 \in E_2$ pointing upwards. A ground plane is assumed to be at position $0 \in \mathcal{I}$ with a normal of $(0, 1, 0) \in \mathcal{I}$. The gravity vector $-ge_2$ is assumed constant with $g = 9.81m/s^2$, friction and air resistance are not part of the model.

Rigid Body Dynamics. The position of the centre of mass of the flyer is given by $\xi \in \mathcal{I}$. A second frame of reference, the body frame, is given by $\mathcal{B} = \{E_1^a, E_2^a, E_3^a\}$ its origin is the flyers centre of mass and fixed to the flyers orientation. The orientation of the body frame $\mathcal{B}$ is identified by the unity quaternion q. The systems dynamics are described using Newton's equations of motion. [7]

$$\dot{\xi} = v \tag{1}$$

$$m\dot{v} = F - mge_2 \tag{2}$$

$$\dot{q} = \frac{1}{2}\begin{bmatrix} 0 \\ \omega \end{bmatrix} q \tag{3}$$

$$I\dot{\omega} = -\omega \times I\omega + \tau \tag{4}$$

where $v \in \mathcal{I}$ is the linear velocity and $F \in \mathcal{I}$ the external forces acting on the flyer not including gravity. Furthermore $\omega \in \mathcal{B}$ describes the angular velocity and $\tau \in \mathcal{B}$ the torque in the body reference frame. I is the inertia tensor in the body fixed frame $\mathcal{B}$ and m the overall mass of the flyer. We consider F and τ to be unknown noise inputs for now.

Image Sensor. A perspective projection with known image plane distance is assumed for the optical system of the sensor. The image sensors output is then $\tilde{G} = G_{\xi,q} + n_G$, with $G_{\xi,q}, \tilde{G} \in \mathbb{R}_{0:1}^{w \times h}$ being a arrays of image brightness values in the range $\mathbb{R}_{0:1} = \{x \in \mathbb{R} | 0 \leq x \leq 1\}$, noise term n_G and (w, h) the image resolution. With the assumption of a constant ground plane texture, the noise free image $G_{\xi,q}$ is uniquely identified by the sensor position ξ and the orientation q.

To be able to relate consecutive frames, the image brightness constraint states that for any given point $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$ and an image brightness $I(x, t)$ at time t,

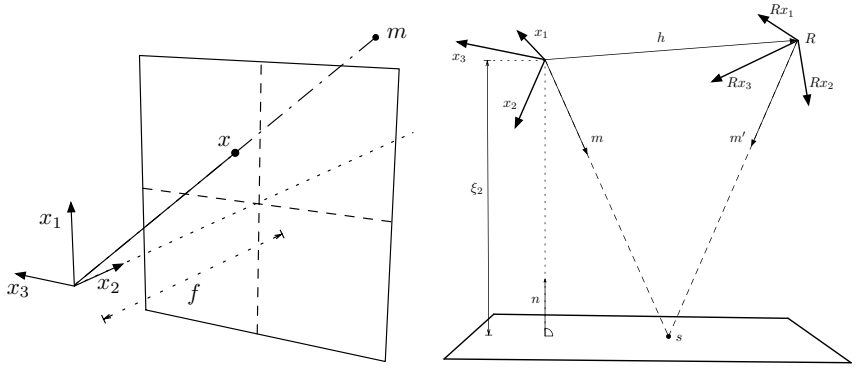
$$I(x, t) = I(x + \delta x, t + \delta t). \tag{5}$$

With the ground plane as the only scene object assumed, Bosse et al. [2] give a method to relate image sequences to the camera state. A perspective image coordinate m is defined as,

$$x = \psi(m) = \frac{f}{m_2}\begin{pmatrix} m_1 \\ m_3 \end{pmatrix}. \tag{6}$$

where the focal length f can be assumed unity. We can also define the inverse of ψ by setting the second component of a three vector to unity.

$$m = \psi^{-1}(x) = \begin{pmatrix} x_1 \\ 1 \\ x_2 \end{pmatrix} \tag{7}$$



(a) Perspective coordinates (b) Projective transformation

Fig. 1. Point on plane related to different camera states

If we now assume a camera with position ξ and an orientation q which is exerted to a translation of h and a rotation of R the projective transform of a perspective coordinate can be given as

$$\mathcal{A} = \frac{1}{k}(I + ph^T)R, \quad k = \sqrt[3]{1 + p^T h} \tag{8}$$

where p is the ground plane surface normal expressed in the reference frame given by R_q, divided by the distance of the camera to the plane ξ_2.

$$p = \frac{R_q^T e_2}{\xi_2} \tag{9}$$

Because perspective coordinates can be scaled arbitrarily, the k factor makes sure that $\det(\mathcal{A}) = 1$ is normalised.

The transform $\mathcal{A}$ allows now to relate the image point x to the projected point x' by using (6) and (7). The concept is illustrated in Figure 1.

$$x' = \psi(\mathcal{A}^T \psi^{-1}(x)) \tag{10}$$

The problem is to find the transformation matrix $\mathcal{A}$ which fits best for the brightness constraint criterion (5). Given a set of image points X we can formulate such a constraint easily using (10) and (5).

$$\min_{\mathcal{A} \in \mathbb{R}^{3\times 3}} \sum_{x \in X} \left(I(\psi(\mathcal{A}^T \psi^{-1}(x)), t + \delta t) - I(x, t) \right)^2 \tag{11}$$

Equation (11) has the geometric interpretation of projecting the image at $t + \delta t$ using the transformation and aligning it with the image at t. Because the problem is an optimisation of a least square formulation, the requirements for applying the Levenberg-Marquardt method are fulfilled.

The image sensor provides an array of brightness values. However, the optimisation criterion (11) requires a spatially continuous function $I(x)$. This is achieved by using a bicubic interpolation method [13].

Sensor Fusion. The state estimation approach is partly based on [6] and [15]. An Extended Kalman Filter with a directly formulated state is used, where the state vector x consists of

$$x = (\xi, v, q, \omega)^T \tag{12}$$

The state vector carries the angular velocity ω, which is necessary to form the measurement estimate of the perspective transform for the vision system (8).

$$\dot{\xi} = v \tag{13}$$
$$\dot{v} = R_q a + g e_2 \tag{14}$$
$$\dot{q} = -\frac{1}{2}\Omega(\omega) q \tag{15}$$

The derivative of q which for the system description (3) is given as a quaternion multiplication is replaced with the matrix equivalent $\Omega(\omega)$ defined for the body reference frame [3] as

$$\Omega(\omega) = \begin{pmatrix} 0 & \omega^T \\ -\omega & -[\omega]_\times \end{pmatrix} = \begin{pmatrix} 0 & \omega_1 & \omega_2 & \omega_3 \\ -\omega_1 & 0 & \omega_3 & -\omega_2 \\ -\omega_2 & -\omega_3 & 0 & \omega_1 \\ -\omega_3 & \omega_2 & -\omega_1 & 0 \end{pmatrix} \tag{16}$$

This system description is discretised using a first-order Euler update.

$$p_{k+1} = p_k + v\Delta t \tag{17}$$
$$v_{k+1} = v_k + (R_q a + g e_2)\Delta t \tag{18}$$
$$q_{k+1} = \left(I(\cos(s) + \eta\Delta t\gamma) - \frac{\sin(s)}{2s}\Omega(\omega)\Delta t \right) q_k, \quad \gamma = 1 - |q_k|^2 \tag{19}$$

Where γ a Lagrange multiplier term [6], which ensures that q always converges to unity, with $\eta\Delta t < 1$ determining the speed of convergence.

The descriptions for the discrete dynamics (18) both still contain the measurement noise vectors n_a and n_ω, which can easily be extracted.

$$v_{k+1} = v_k + (R_q \tilde{a} + g e_2)\Delta t + p_a \tag{20}$$
$$q_{k+1} = \left(I(\cos(s) + \eta\Delta t\gamma) - \frac{\sin(s)}{2s}\Omega(\tilde{\omega})\Delta t \right) q_k + p_\omega \tag{21}$$

with

$$p_a = R_q n_a \Delta t, \quad p_\omega = \frac{-\sin(s)}{2s}\Omega(n_\omega \Delta t) q_k \tag{22}$$

The noise vector n_a and n_ω are given in the continuous time domain. However, for the discrete filter a discretised form of the noise vectors is required. We specify $\bar{n}_a$ and $\bar{n}_\omega$ to be the discrete versions of n_a and n_ω. The covariances of $\bar{n}_a$ and $\bar{n}_\omega$ can actually be measured and are given by

$$\bar{n}_a \sim N(0, P_a), \quad \bar{n}_\omega \sim N(0, P_\omega) \tag{23}$$

For the filtering process however, we need the covariance of the direct noise terms p_a and p_ω.

$$Q_v = \mathrm{cov}(p_a) = R_q P_a R_q^T \Delta t^2 \tag{24}$$

The case is not as straightforward for p_ω. However, it is possible to find a solution for the equation $p_k = \Gamma(q)\omega$ by defining $\Gamma(q)$ as

$$\Gamma(q) = \begin{pmatrix} q_2 & q_3 & q_4 \\ -q_1 & q_4 & -q_3 \\ -q_4 & -q_1 & q_2 \\ q_3 & -q_2 & -q_1 \end{pmatrix} \tag{25}$$

so that we can give the covariance of p_ω as

$$Q_q = \mathrm{cov}(p_\omega) = \Gamma(q_{sk}) P_\omega \Gamma(q_{sk})^T, \quad q_{sk} = \frac{-\sin(s)\Delta t}{2s} q_k \tag{26}$$

The covariance of the angular velocity is of course equal to that measurement covariance of the gyro sensors, $Q_\omega = P_\omega$. Further, the cross covariance of Q_q and Q_ω is given by

$$Q_{q,\omega} = \Gamma(q_{sk}) P_\omega \tag{27}$$

so that we can give the complete process noise covariance matrix Q as

$$Q_k = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & Q_v & 0 & 0 \\ 0 & 0 & Q_q & Q_{q,\omega} \\ 0 & 0 & Q_{q,\omega}^T & Q_\omega \end{pmatrix} \tag{28}$$

Recalling that the measurement for the filter is the projection matrix given in (8), we need to formulate $h(x)$ to map to the projection matrix $\mathcal{A}$ based on the current state vector x.

$$\mathcal{A}_s = \left(I + \frac{R_{q_k}^T e_2 (v_k \Delta t)^T}{\xi_2} \right) e^{[\omega_k]_\times \Delta t} \tag{29}$$

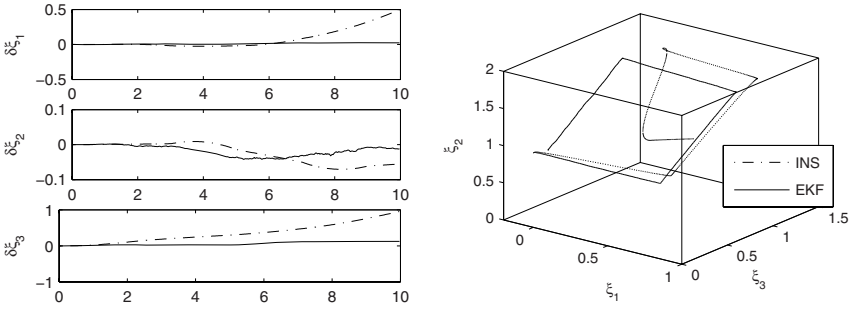
The matrix exponential can be expressed in the closed form as

$$e^{[\omega_k]_\times \Delta t} = I + \frac{\sin(s)}{s} [\omega \Delta t]_\times + \frac{1 - \cos(s)}{s^2} [\omega \Delta t]_\times^2, \quad s = |\omega \Delta t| \tag{30}$$

the measurement function is then given by the matrix in (29) normalised to unity.

$$h(x) = \frac{1}{\sqrt[3]{\det(\mathcal{A}_s)}} \mathcal{A}_s \tag{31}$$

The covariance of the measurement noise is specified by R_k. This matrix is directly calculated from measurements with known ground truths and not analytically derived.



(a) Position error (b) Spatial plot

Fig. 2. Evaluation scenario position comparison

4 Experiments

To verify the results of the EKF implementation a simulated environment based on the rigid body and image sensor description given in the previous section is used. The accelerometer and gyro noise has been modelled to be similar in properties to that of low-cost sensor components. A raytracer is used to simulate the image sensor output, and an additional pixel noise term is added. The ground texture wich is used is a grayscale random pattern.

Figure 2 shows a plot of the position errors for an evaluated example scenario, where the simulated UAV was made to perform a rectangular movement on a tilted plane. The start and end position are the same.

The overall error is significantly smaller for both plots in $\delta\xi_2$ than for the other two components. This is related to the fact that the e_2 component of the accelerometer has a lower noise part to reflect the same situation in an already implemented real system. For all three components of the position vector, the aided EKF offers a better performance than the inertial sensors alone. Figure 2 (b) gives a direct comparison of the position and allows a spatial interpretation of the results. For a total distance of about $5m$ the INS (Inertial Navigation System) solution has an error of more than $1m$, where the aided estimation is around $5cm$. A similar trend is visible in other evaluated scenarios.

5 Conclusion

The results of the simulation are very promising, as they show that the combination of the visual information and the inertial sensors can for particular situations perform better by a factor of 20 when compared to simple integration of low-cost inertial sensors. Previous works using similar methods did not use the projection matrix directly as a measurement. They extracted the state information from the matrix and relied on third sensors to solve the ambiguity in the velocity and height data. The proposed approach is able to handle this situation without the need for additional reference information and is stable even if the state estimate is inaccurate. The proposed method uses the constant brightness constraint method for relating consecutive images. Another possible solution is

to perform feature extraction and relate the features of consecutive images. For this approach to work, the image requires enough visual information so that features can accurately be identified. A combination of the two approaches could potentially solve a wider class of problems. The sensor fusion method applied in the context of this work uses the Extended Kalman Filter, which has known problems with highly non-linear systems. An alternative approach which might have advantages in this area is the class of Unscented Kalman Filters.

References

1. Barrows, G., Neely, C.: Mixed-mode vlsi optic flow sensors for in-flight control of a micro air vehicle. In: SPIE 45th Annual Meeting (2000)
2. Bosse, M., Karl, W., Castanon, D., DeBitetto, P.: A vision augmented navigation system. In: Intelligent Transportation System, Boston, MA, USA, pp. 1028–1033 (1997)
3. Coutsiasy, E.A., Romeroz, L.: The quaternions with an application to rigid body dynamics
4. Cox, T.H., Nagy, C.J., Skoog, M.A., Somers, I.A.: Civil UAV capability assessment. Technical report, NASA (December 2004)
5. Galvin, B., McCane, B., Novins, K., Mason, D., Mills, S.: Recovering motion fields: An evaluation of eight optical flow algorithms. In: Proceedings of the British Machine Vision Conference 1998 (1998)
6. Gavrilets, V.: Autonomous aerobatic maneuvering of miniature helicopters. PhD thesis, Massachusetts Institute of Technology (2003)
7. Hamel, T., Mahony, R., Lozano, R., Ostrowski, J.-N.: Dynamic modelling and configuration stabilization for an x4-flyer. In: 15th IFAC World Congress 2002 (2002)
8. Horn, B.K., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17(1-3), 185–203 (1981)
9. Jähne, B., Haußecker, H.: Computer Vision and Applications. Academic Press, London (2000)
10. Low, T., Wyeth, G.: Obstacle detection using optical flow. In: Proceedings of the 2005 Australasian Conference on Robotics & Automation (2005)
11. Muratet, L., Doncieux, S., Briere, Y., Meyer, J.-A.: A contribution to vision-based autonomous helicopter flight in urban environments. Robotics and Autonomous Systems 50(4), 195–209 (2005)
12. Negahdaripour, S., Horn, B.K.P.: Direct passive navigation: Analytical solution for planes. IEEE Robotics and Automation. Proceedings. 3, 1157–1163 (1986)
13. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, New York (1992)
14. Stocker, A.A.: Analog integrated 2-d optical flow sensor. Analog Integr. Circuits Signal Process 46(2), 121–138 (2006)
15. van der Merwe, R., Wan, E.A., Julier, S.: Sigma-point kalman filters nonlinear estimation and sensor fusion - applications in integrated navigation. In: AIAA Guidance Navigation and Controls Conference (March 2004)