State Estimation using Optical Flow from Parallax-Weighted Feature Tracking

Joseph J. Kehoe[*] Adam S. Watkins[†] Ryan S. Causey[‡] and Rick Lind[§]
Department of Mechanical and Aerospace Engineering
University of Florida

Computer vision presents an attractive sensor option for micro aerial vehicle (MAV) applications due to the payload and performance restrictions imposed by typical mission scenarios. Optical flow can be measured by tracking the perceived motion of feature points between successive image frames. This perceived feature-point motion yields information regarding vehicle motion as described through geometric relationships. This paper presents an optimization-based approach to estimate aircraft angular rates and wind-axis angles using monocular vision. A bias in the optical-flow equations is leveraged to decouple components resulting from angular and translational motion, respectively. Attempts to resolve the ambiguity introduced by the loss of depth information are avoided through this decoupling. Additionally, estimator performance is shown to rely on proper selection of feature points used for the estimation process. Parallax measurements are used to identify features that are most likely to yield accurate state estimates. The technique is then demonstrated through simulation.

I. INTRODUCTION

Unmanned aerial vehicles (UAVs) have been considered for a variety of applications. Recent progress in the development of small-scale, fixed wing UAVs has facilitated mission scenarios that were previously not possible. A micro aerial vehicle (MAV) demonstrates the stealth and agility to maneuver undetected throughout complex and cluttered environments; however, the control challenges posed by such a task are diverse and many. While a variety of miniature inertial-measurement packages have been emerging for use in MAV applications, these units add weight and often provide inaccurate information during aggressive maneuvers. Additionally, many of these units rely on position measurements from GPS to generate their measurements. Access to GPS signals during urban flight operations is likely to be intermittent at best. As such, navigation capabilities can be severely impaired for large segments of the mission.

Alternatively, small video cameras present a lightweight, low-power, and information-rich sensor. Computer vision has been demonstrated as an enabling technology for MAV control tasks. Vision-based techniques are especially effective when maneuvering relative to nearby objects. These situations provide a diverse set of information concerning camera motion. Thus, computer vision is well-suited to aid navigation during GPS-denied mission segments.

Vision processing techniques allow the extraction of information concerning both the environment as well as camera motion. In particular, techniques that can identify and track points of significance in successive images enable a variety of problems to be addressed. These "feature points," as depicted in Figure 1, are commonly detected using the intensity gradient of the image, and are then tracked using techniques such as template registration, point correlation, dynamic filtering, or some combination thereof.[1–3] Recent efforts have moved these tracking algorithms towards real-time implementation for images with large motion between frames for flight control applications.[2,4]

Figure 1. Tracking feature points (green) yields perceived relative motion

Tracking features between consecutive frames gives an indication of the perceived motion in the image. This apparent velocity is denoted optical flow

[*]Graduate Student, *jjk@ufl.edu*
[†]Graduate Student, *awatts@ufl.edu*
[‡]Graduate Student, *aerocauz@ufl.edu*
[§]Assistant Professor, *ricklind@ufl.edu*

and can be characterized by the gradient of the image intensity function.[5] A variety of techniques exist for computing optical flow, including differential methods, frequency-based methods, and correlation-based methods.[6] For the case of a fixed environment, optical flow results solely from camera motion and can therefore provide information related to the vehicle state.

Researchers have exploited this observation previously for state estimation. Soatto et al. derived a form of the Kalman filter that uses the relationship between vision-based measurements and the motion of the camera.[7,8] The resulting implicit extended Kalman filter (IEKF) can be used to recover the camera motion states. Gurfil and Rotstein recast Soatto's work in terms of an aircraft state-estimation problem by incorporating aircraft dynamics into the IEKF framework.[9] The resulting formulation partially estimated the aircraft states but exhibited relatively slow convergence. Improvements have been demonstrated by Webb et al. who also used an aircraft model.[10,11] Unfortunately, accurate MAV models are often not available within an aggressive flight regime where the aerodynamics are notoriously difficult to characterize.

Other techniques, such as the eight-point algorithm,[12] do not require a vehicle model to obtain state estimates. The eight-point algorithm computes the relative translation and rotation of the camera between two vantage points using a geometric relationship. Epipolar geometry requires that a single point in space viewed from two vantage points will be coplanar with each vantage point. This constraint is used to compute the essential matrix and recover camera motion.

Several techniques have utilized the kinematic relationship between camera motion and the resulting optical flow to directly solve for unknown motion parameters using constrained optimization.[13–15] These techniques depend on at least partial knowledge of the translational velocity for use in the optimization. This knowledge often depends on GPS measurements and therefore is not available for the mission segments of interest here.

This paper addresses the problem of estimating aircraft states during a GPS-denied mission segment. An iterative optimization approach is adopted to determine the angular rates and the wind-axis angles. No knowledge of vehicle velocity is required. The coupled aircraft-camera kinematics are used to solve for aircraft states in similar fashion to previous efforts; however, velocity dependencies are removed through decoupling the optical flow resulting from angular and translational motion, respectively. Angular rate estimates are obtained initially and used to set up a simple linear least-squares problem for the aerodynamic angles. Performance of the least-squares problem is further improved through the application of a weighting scheme derived from parallax measurements.

The remainder of the paper is organized as follows. Section II develops the coupled equations of motion describing the dynamics of the aircraft-camera system, thus forming a mathematical basis for the remainder of the paper. Section III discusses some properties of the relationship between camera motion and optical flow. Section IV details the vision-based approach to state estimation. Finally, Section V discusses a demonstrative example performed in simulation using a nonlinear aircraft model.

II. COUPLED AIRCRAFT-CAMERA MOTION

Video data from a camera moving relative to features in the environment can provide a great deal of information concerning the motion of the camera. The relationship between the image data and the motion of the vehicle carrying the camera must be characterized in order to incorporate this information into estimation and control techniques. The general framework developed here allows for the formulation and solution of a variety of vision-based control problems.[16]

II.A. Problem Geometry

The geometry for the general case of an aircraft carrying a single camera is described by the vector diagram of Figure 2 and Eqs. 1a-1d. The aircraft body basis, **B**, is fixed at the aircraft center of gravity and is located relative to the inertial frame, **E**, by the vector $\vec{R}$ as in Eq. 1a. The camera reference frame, **C**, is fixed to the aircraft at an arbitrary position and is located relative to **B** by the vector $\vec{\Delta}$ as in Eq. 1b.

$$\vec{R} = R_1\hat{e}_1 + R_2\hat{e}_2 + R_3\hat{e}_3 \tag{1a}$$

$$\vec{\Delta} = \Delta_1\hat{b}_1 + \Delta_2\hat{b}_2 + \Delta_3\hat{b}_3 \tag{1b}$$

$$\vec{\xi}_i = \xi_{1_i}\hat{e}_1 + \xi_{2_i}\hat{e}_2 + \xi_{3_i}\hat{e}_3 \tag{1c}$$

$$\vec{\eta}_i = \eta_{1_i}\hat{c}_1 + \eta_{2_i}\hat{c}_2 + \eta_{3_i}\hat{c}_3 \tag{1d}$$

The i^{th} point-feature of the environment is denoted m_i and is located relative to **E** by $\vec{\xi}_i$ as in Eq. 1c. The vector

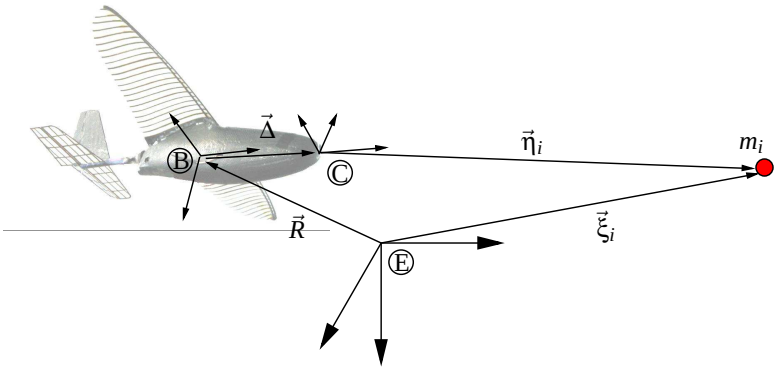


Figure 2. Problem Geometry

$\vec{\eta}_i$ then relates the relative position of m_i with respect to **C**, as described by Eq 1d. This relative position vector can be expressed in terms of the aircraft position and orientation via the coordinate transformations T_{EB} and T_{BC}, which represent simple axis rotations from **E** to **B** and **B** to **C**, respectively. The usual definitions for pitch, roll, and yaw Euler angles are used to define T_{EB} as in Eq. 2. The transformation T_{BC} is a specified quantity that is dependent upon the physical mounting of the camera to the airframe.

$$\begin{aligned} T_{EB} &= T_\phi \cdot T_\theta \cdot T_\psi \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{2}$$

The relative position of m_i can be expressed as a vector sum of the quantities in Equations 1a-1c, as shown in Eq. 3. Each vector in Eq. 3 is transformed to **C** in order to maintain consistency in bases with $\vec{\eta}$. Note that the dependence of $\vec{\eta}$ on the position and orientation of the aircraft is introduced in Eq. 3 through the inertial position, $\vec{R}$, and the transformation from **E** to **B**, T_{EB}.

$$\vec{\eta}_i = T_{BC} T_{EB} (\vec{\xi}_i - \vec{R}) - T_{BC} \vec{\Delta} \tag{3}$$

II.B. Camera Model

The pinhole camera model maps the projection of a point in space, m_i, onto a plane that is normal to the $\hat{c}_3$ axis. This plane is denoted the image plane and is displaced along the $\hat{c}_3$ axis by the focal length, f, which is an intrinsic camera parameter. The mapping results from a transformation that is dependent upon both the relative position, $\vec{\eta}$, and f. The resulting projection of m_i is located on the image plane at the coordinates μ_i and ν_i which are determined as in Eq. 4. The geometry of this mapping is depicted in Figure 3.

$$\begin{bmatrix} \mu_i \\ \nu_i \end{bmatrix} = \frac{f}{\eta_{3_i}} \begin{bmatrix} \eta_{1_i} \\ \eta_{2_i} \end{bmatrix} \tag{4}$$

The mapping to an image is restricted by the field of view (FOV) such that $\mu \in [\underline{\mu}, \overline{\mu}]$ and $\nu \in [\underline{\nu}, \overline{\nu}]$. This constraint on the two-dimensional image position translates to an angular constraint in three-dimensions. This angular description of the constraint results from the property that the pinhole camera actually relates the line-of-sight (LOS) to features within the camera FOV. The projection of a point in three-dimensional space to the two-dimensional image maps back to a ray in three-dimensional space along the LOS as seen in Figure 3. Inspection of Eq. 4 and Figure 3 reveals that the values of the image coordinates, (μ, ν), are proportional to the tangents of the angles that orient this LOS in the vertical and horizontal directions, respectively.

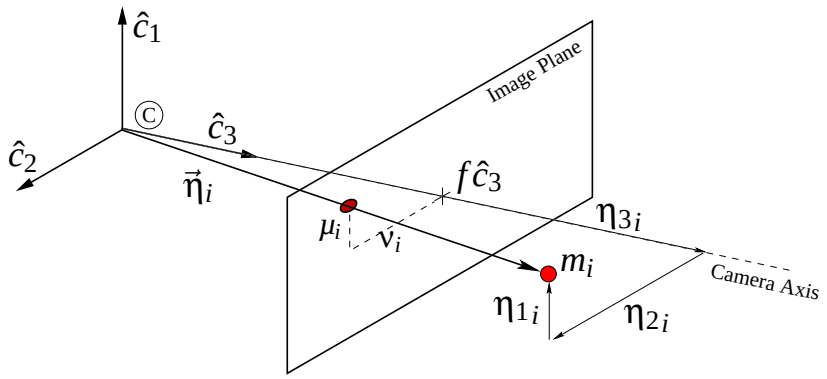


Figure 3. Mapping m_i to the Image Plane

II.C. Coupled Dynamics

The relative position of the i^{th} feature point with respect to the **C** basis, $\vec{\eta}_i$, is expressed in terms of aircraft states in Eq. 3 as discussed previously. This vector can be differentiated to yield a set of state equations to augment the aircraft equations of motion. The new coupled system can then be used with the pinhole camera model to provide a continuous dynamic description of image-based motion of the projection of m_i as the vehicle state evolves.

The derivative is taken prior to any coordinate transformations with respect to the **E** reference frame as shown in Eq. 5. Evaluation of the derivatives requires consideration for the rotating bases of the respective vectors as defined in Eqs. 1a-1d. Therefore the angular velocities of the **B** and **C** frames with respect to the **E** frame are introduced as in Eq. 6 and are represented by ${}^E\vec{\omega}^B$ and ${}^E\vec{\omega}^C$, respectively. These angular velocities are expressed in their respective rotating reference frames. The relationship is then manipulated to express $\dot{\vec{\eta}}_i$ in terms of the remaining variables in Eq. 7.

$$\frac{{}^E d}{dt}(\vec{\eta}_i) = \frac{{}^E d}{dt}\left(\vec{\xi}_i\right) - \frac{{}^E d}{dt}\left(\vec{R}\right) - \frac{{}^E d}{dt}\left(\vec{\Delta}\right) \tag{5}$$

$$\frac{{}^C d}{dt}(\vec{\eta}_i) + \left({}^E\vec{\omega}^C \times \vec{\eta}_i\right) = \dot{\vec{\xi}}_i - \dot{\vec{R}} - \left(\dot{\vec{\Delta}} + {}^E\vec{\omega}^B \times \vec{\Delta}\right) \tag{6}$$

$$\dot{\vec{\eta}}_i = \dot{\vec{\xi}}_i - \dot{\vec{R}} - \left(\dot{\vec{\Delta}} + {}^E\vec{\omega}^B \times \vec{\Delta}\right) - \left({}^E\vec{\omega}^B + {}^B\vec{\omega}^C\right) \times \vec{\eta}_i \tag{7}$$

Equation 7 can be simplified through the application of several assumptions. The camera is assumed to have a fixed position and orientation with respect to **B**. This assumption implies that $\dot{\vec{\Delta}} = 0$ and ${}^B\vec{\omega}^C = 0$. Further, the environment is assumed to be static such that the inertial velocity of m_i, given by $\dot{\vec{\xi}}_i$, evaluates to zero as well. Finally, the camera location is assumed to coincide with the center of gravity such that $\vec{\Delta} = 0$. The remaining quantities are transformed from their respective bases to the **C** frame for consistency, as shown in Eq. 8.

$$\dot{\vec{\eta}}_i = -T_{BC}T_{EB}\dot{\vec{R}} - \left(T_{BC}\,{}^E\vec{\omega}^B \times \vec{\eta}_i\right) \tag{8}$$

Equation 8 can be further simplified by recognizing the definitions of the body-axis translational and angular velocities, as shown in Equations 9 and 10. Finally, the transformation T_{BC} is specified as in Eq. 11. This form has the $\hat{c}_3$ axis aligned with the $\hat{b}_1$ axis such that the camera is pointing directly out the nose of the airplane. The $\hat{c}_1$ axis is located in the $\hat{b}_1$-$\hat{b}_3$ plane pointing upwards while the $\hat{c}_2$ axis points along the $\hat{b}_2$ axis out the right wing.

$$T_{EB}\dot{\vec{R}} = \vec{V}_B = \begin{bmatrix} u & v & w \end{bmatrix}^T \tag{9}$$

$$^E\vec{\omega}^B = \begin{bmatrix} p & q & r \end{bmatrix}^T \tag{10}$$

$$T_{BC} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \tag{11}$$

The resulting differential equation is shown as Eq. 12. This equation expresses the time rate of change of relative feature position with respect to aircraft motion. Equation 12 can be used to augment the well-known aircraft equations of motion for each tracked feature to form a fully-coupled system.

$$\dot{\vec{\eta}}_i = -\begin{bmatrix} -w - p\eta_{2_i} + q\eta_{3_i} \\ v + p\eta_{1_i} + r\eta_{3_i} \\ u - q\eta_{1_i} - r\eta_{2_i} \end{bmatrix} \tag{12}$$

III. OPTICAL FLOW

Optical flow refers to perceived motion in an image resulting from relative motion between the camera and the environment.[17] This motion can be generated by translation of the camera, rotation of the camera, or motion of features within the environment. The environment is considered static for this work and thus optical flow is assumed to be generated by camera motion alone for the remainder of discussion. Further, this paper interprets the motion of specific point features of the environment as the optical flow. This motion is described through the camera model of Section II.B. Hence, the optical flow of a particular point, m_i, can be considered as a velocity vector,$\vec{V}_{flow,i}$, defined on the image plane as in Eq. 13. The components of $\vec{V}_{flow,i}$ are resolved in the image coordinate directions and therefore consist of the time-derivatives of m_i's image coordinates, (μ_i, v_i), as shown in Figure 4(a). These quantities can be measured directly from video data for a set of n points given the assumption that each point can be tracked between two image frames. A variety of techniques discussed in the vision processing literature have been shown to successfully detect and track feature points through a series of images.[1–3]

$$\vec{V}_{flow,i} = \begin{bmatrix} \dot{\mu}_i & \dot{v}_i \end{bmatrix}^T \tag{13}$$

The expression for the velocity of m_i's projection in the image is found through differentiation of Eq. 4. The developments of Section II allow this derivative to be rewritten in terms of aircraft states. The general expression for the focal plane velocities of a feature point is given by Eq. 14.

$$\begin{bmatrix} \dot{\mu}_i \\ \dot{v}_i \end{bmatrix} = \frac{f}{{\eta_{3_i}}^2} \begin{bmatrix} \eta_{3_i} & 0 & -\eta_{1_i} \\ 0 & \eta_{3_i} & -\eta_{2_i} \end{bmatrix} \begin{bmatrix} \dot{\eta}_{1_i} \\ \dot{\eta}_{2_i} \\ \dot{\eta}_{3_i} \end{bmatrix} \tag{14}$$

Upon substitution of the components of Eq. 12 into Eq. 14, several terms cancel while others can be grouped into the form of the focal-plane coordinates given previously by Eq. 4. Camera calibration is assumed known such that the focal length can be normalized to $f = 1$. The equations can now be expressed as in Eqs. 15a and 15b, where the depth component of $\vec{\eta}_i$ has been replaced with d_i to simplify notation.

$$\dot{\mu}_i = \mu_i \frac{u}{d_i} + \frac{w}{d_i} + v_i p - (1 + {\mu_i}^2)q - \mu_i v_i r \tag{15a}$$

$$\dot{v}_i = v_i \frac{u}{d_i} - \frac{v}{d_i} - \mu_i p - \mu_i v_i q - (1 + {v_i}^2)r \tag{15b}$$

III.A. Motion-Decoupled Feature Tracking

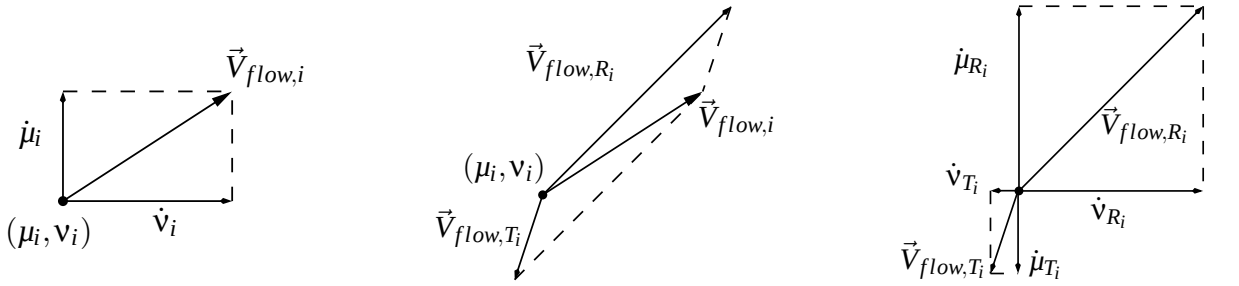
While it is natural to consider components of $\vec{V}_{flow}$ with respect to the orthogonal image basis and the corresponding coordinates, (μ, v), the optical flow can also be decomposed into contributions from motion parameters.[18] The decomposition shown by Eq. 16 distinguishes between the flow of m_i induced by translational motion, $\vec{V}_{flow,T_i}$, and the flow

induced by rotational motion, $\vec{V}_{flow,R_i}$. This decomposition is non-orthogonal in general as depicted in Figure 4(b). The non-orthogonal motion components are then resolved in terms of the image basis, as shown by Figure 4(c) and Eqs. 17 and 18.

$$\vec{V}_{flow,i} = \vec{V}_{flow,T_i} + \vec{V}_{flow,R_i} \tag{16}$$

$$\vec{V}_{flow,T} = \begin{bmatrix} \dot{\mu}_{T_i} & \dot{\nu}_{T_i} \end{bmatrix}^T \tag{17}$$

$$\vec{V}_{flow,R} = \begin{bmatrix} \dot{\mu}_{R_i} & \dot{\nu}_{R_i} \end{bmatrix}^T \tag{18}$$



(a) Orthogonal Decomposition (b) Motion Decomposition (c) Orthogonal Motion Decomposition

Figure 4. Projecting an Optic Flow vector on orthogonal and non-orthogonal bases

The expressions relating the components of $\vec{V}_{flow,T_i}$ and $\vec{V}_{flow,R_i}$ are easily recognized from the total flow expressions of Eqs. 15a and 15b. The translational components simply consist of the terms of Eqs. 15a and 15b that contain the velocity states, (u,v,w), as shown by Eqs. 19a and 19b. Similarly, the rotational components consist of the terms of Eqs. 15a and 15b that contain the angular rate states, (p,q,r), as shown by Eqs. 20a and 20b. Equations 21a and 21b follow directly from this simple decoupling method.

$$\dot{\mu}_{T_i} = \frac{\mu_i u + w}{d_i} \tag{19a}$$

$$\dot{\nu}_{T_i} = \frac{\nu_i u - v}{d_i} \tag{19b}$$

$$\dot{\mu}_{R_i} = \nu_i p - (1+\mu_i^2)q - \mu_i \nu_i r \tag{20a}$$

$$\dot{\nu}_{R_i} = -\mu_i p - \mu_i \nu_i q - (1+\nu_i^2)r \tag{20b}$$

$$\dot{\mu}_i = \dot{\mu}_{T,i} + \dot{\mu}_{R,i} \tag{21a}$$

$$\dot{\nu}_i = \dot{\nu}_{T,i} + \dot{\nu}_{R,i} \tag{21b}$$

III.B. Parallax and the Focus of Expansion

The translational optical flow component can be considered as a measure of parallax. Parallax refers to the apparent relative motion of points at different depths resulting from camera displacement. Essentially, points exhibiting large parallax are likely to be nearby while points exhibiting small parallax are likely distant. These effects are seen to result from the inverse depth relationship present in Eqs. 19a and 19b. Hence, parallax is considered here as a measure of relative depth of feature points with respect to each other.

A degenerate case of this observed relationship occurs for translation directly toward a point in space. No apparent motion is observed for this case. While this lack of motion degrades the potential information given by parallax measurements, it does yield useful information regarding the direction of velocity. Recall from Section II.B that each point in the image plane maps to a ray in three-dimensional space. The point that maps to the direction of the velocity vector corresponds to this degenerate case and is denoted the focus of expansion (FOE).[19] The FOE exhibits zero translational flow and is given by Eqs. 22a and 22b. Further, all other translational vectors, $\vec{V}_{flow,T_i}$, radiate from this point.[19] As such, each of the $\vec{V}_{flow,T_i}$ can be related to the FOE coordinates through Eq. 23. This equation simply states that the lines resulting from an infinite extension of each $\vec{V}_{flow,T_i}$ intersect at the coordinates, (μ_F, ν_F). The concept is depicted in Figure 5.

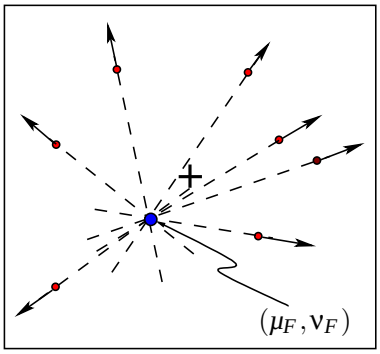


Figure 5. Translational optical flow, or parallax, radiates from the focus of expansion.

$$\mu_F = -w/u \tag{22a}$$

$$\nu_F = v/u \tag{22b}$$

$$\frac{(\mu_i - \mu_F)}{(\nu_i - \nu_F)} = \frac{\dot{\mu}_{T,i}}{\dot{\nu}_{T,i}} \tag{23}$$

Equations 22a and 22b show that the FOE location is related directly to aircraft velocity information. The FOE coordinates, (μ_F, ν_F), actually relate angular information regarding the orientation of the velocity vector. This relationship is described in terms of image information through a tangent relationship as discussed in Section II.B. This information is analogous to the angles describing the transformation from aircraft body axis to wind axis, α and β. Using some simple trigonometric relations, the FOE coordinates are related to α and β through Eqs. 24a and 24b. The angles are not related directly by the tangents of the FOE coordinates due to the wind-axis transformation sequence. The angles related by μ_F and ν_F are each measured relative to the **C** frame. The sideslip angle, β, is defined as a rotation from the intermediate axis that results from rotating the **B** axis through α.

$$\mu_F = -f \cdot \tan\alpha \tag{24a}$$

$$\nu_F = \left(\sqrt{1+\mu_F^2}\right) \cdot \tan\beta \tag{24b}$$

IV. STATE ESTIMATION USING OPTICAL FLOW

An optimization-based approach is adopted to determine the aircraft angular rates and aerodynamic angles given only measurements derived from a sequence of images. The relationships derived in Section III between aircraft motion and image-based parameters are exploited to set up a least-squares minimization problem. Values of the aircraft motion parameters are effectively fitted to optical flow measurements using variations of Eqs. 15a and 15b. Two methods that have been presented previously in the literature are discussed followed by a new method presented here.

IV.A. Method 1

Inspection of the optical flow expressions given by Eqs. 15a and 15b reveals two equations in seven unknowns for a single tracked feature point. This conclusion assumes that the image coordinates, (μ_i, ν_i), and the optical flow components, $(\dot{\mu}_i, \dot{\nu}_i)$, are measurable quantities. The seven unknowns consist of the three aircraft linear velocity components, $(u,\ v,\ w)$, the three aircraft angular velocity components, $(p,\ q,\ r)$, and the feature depth, d_i. Tracking each additional feature point results in the gain of two additional equations at the expense of only a single added unknown variable: the respective feature depths, d_i, of the additional tracked points. As such, a minimum of six points must be tracked to determine the system. A feasible optimization problem can be posed given this over-determined system. The minimization utilizes the cost function given as Eq. 26, which consists of a sum of squared terms of the form shown in Eq. 25. The decision variables consist of the angular velocity, ${}^E\vec{\omega}^B$, the translational velocity, $\vec{V}_B$, and the vector of feature depths for each point, $\vec{d}$.

$$J_{1,i}\left({}^{E}\vec{\omega}^{B},\vec{V}_B,d_i\right)=\left(\frac{\mu_i u+w}{d_i}+\nu_i p-(1+{\mu_i}^2)q-\mu_i\nu_i r-\dot{\mu}_i\right)^2+\left(\frac{\nu_i u-v}{d_i}-\mu_i p-\mu_i\nu_i q-(1+{\nu_i}^2)r-\dot{\nu}_i\right)^2 \tag{25}$$

$$J_1\left({}^{E}\vec{\omega}^{B},\vec{V}_B,\vec{d}\right)=\sum_i J_{1,i} \tag{26}$$

This method has been previously shown to yield accurate estimates for the angular rate variables; however, errors are encountered in the computation of the body-axis velocities.[15] To account for this deficiency, the motion decoupling concept described in Section III.A is leveraged to determine an estimate of the FOE. This value is then used to compute estimates of the wind-axis angles, α and β. Specifically, the rate estimates that result from the minimization of Eq. 26 can be used in conjunction with Eqs. 20a and 20b to compute the rotational component of the optical flow over the entire image. This information allows access to the translational component, or parallax, which can be used to estimate the FOE. Estimates of the angle of attack, α, and the angle of sideslip, β, follow from a straightforward computation. Hence, only directional velocity information can be recovered.

The optical flow component due to translational motion cannot be found explicitly from Eqs. 19a and 19b due to the poor estimates of the body-axis velocities, and the unknown depths to each individual feature. Alternatively, flow measurements and the estimates of angular velocity can be used to recover the translational flow field as in Eqs. 27 and 28. This result is achieved using Eqs. 20a and 20b with the angular velocity estimates to compute $\dot{\mu}_{R_i}$ and $\dot{\nu}_{R_i}$ for all feature points.

$$\dot{\mu}_{T_i}=\dot{\mu}_i-\dot{\mu}_{R_i} \tag{27}$$

$$\dot{\nu}_{T_i}=\dot{\nu}_i-\dot{\nu}_{R_i} \tag{28}$$

Recall from Section III.A that the translational optical flow radiates from the FOE and that the location of this point, (μ_F,ν_F), can be used to obtain α and β through Eqs. 24a and 24b. Hence, an estimate of the FOE position leads to an estimate of the wind-axis angles. An approximation for the FOE location is found by extending the translational optical-flow vectors and seeking the common point of intersection. Using n feature points results in a system of n equations in 2 unknowns. A linear least-squares optimization is used as in Eq. 29 to estimate the FOE coordinate values.

$$\begin{bmatrix}\mu_F\\ \nu_F\end{bmatrix}=\arg\min_{\substack{\mu\in[\underline{\mu},\overline{\mu}]\\ \nu\in[\underline{\nu},\overline{\nu}]}}\frac{1}{2}\parallel C\begin{bmatrix}\mu\\ \nu\end{bmatrix}-\gamma\parallel^2 \tag{29}$$

where

$$C=\begin{bmatrix}1 & 1 & \cdots & 1\\ -\frac{\dot{\mu}_{t_1}}{\dot{\nu}_{t_1}} & -\frac{\dot{\mu}_{t_2}}{\dot{\nu}_{t_2}} & \cdots & -\frac{\dot{\mu}_{t_n}}{\dot{\nu}_{t_n}}\end{bmatrix}^T \tag{30a}$$

$$\gamma=\begin{bmatrix}\left(\mu_1-\frac{\dot{\mu}_{t_1}}{\dot{\nu}_{t_1}}\nu_1\right) & \left(\mu_2-\frac{\dot{\mu}_{t_2}}{\dot{\nu}_{t_2}}\nu_2\right) & \cdots & \left(\mu_n-\frac{\dot{\mu}_{t_n}}{\dot{\nu}_{t_n}}\nu_n\right)\end{bmatrix}^T \tag{30b}$$

Once an approximation is found for the FOE, estimates of α and β are straightforward to compute using Eqs. 24a and 24b. The actual expressions are given as Eqs. 31a and 31b.

$$\hat{\alpha}=-\tan^{-1}\left(\frac{1}{f}\mu_F\right) \tag{31a}$$

$$\hat{\beta}=\tan^{-1}\left(\nu_F/\left(\sqrt{1+\mu_F^2}\right)\right) \tag{31b}$$

This method has been previously shown to exhibit reasonable estimation performance; however, the requirement to compute feature depths for each tracked point drives complexity up and results in poor conditioning of the optimization.[15] Additionally, approximations of the total velocity and average feature point depths are required to initialize

the optimization at each step. Average feature points depths require knowledge about the scaling of the environment which may not always be accurate or available. Further, the most likely source of the required velocity information is a GPS sensor; however, the vision system is intended to provide state estimates during mission segments for which GPS data is not available. Therefore the assumption of a constant "trim velocity" would be the next closest approximation; where this value is taken as the either the last measured velocity prior to loss of GPS signal or some known average cruise speed.

IV.B. Method 2

As an alternative to using Eqs 15a and 15b directly as in Eqs. 25 and 26, the optical flow expressions can be combined to a single equation through the elimination of the depth variable, d_i. This single equation is then used in a similar fashion to set up an optimization problem. After some algebraic manipulation, the resulting cost function for minimization is seen as $J_2\left({}^E\vec{\omega}^B, \vec{V}_B\right)$ in Eq. 33 which consists of terms of the form shown in Eq. 32. This form presents an attractive alternative in comparison to Eq. 25 in that the number of variables has been reduced to only those of interest to the state-estimation problem. A large number of measured feature points can be used in the optimization to ensure robustness without an unnecessary increase in problem complexity.

$$J_{2,i}\left({}^E\vec{\omega}^B, \vec{V}_B\right) = \begin{bmatrix} u & v & w \end{bmatrix} \begin{bmatrix} \left(\mu_i^2 + \nu_i^2\right) p - \nu_i q + \mu_i r + \mu_i \dot{\nu}_i - \dot{\mu}_i \nu_i \\ -\nu_i p + \left(1 + \mu_i^2\right) q + \mu_i \nu_i r + \dot{\mu}_i \\ \mu_i p + \mu_i \nu_i q + \left(1 + \nu_i^2\right) r + \dot{\nu}_i \end{bmatrix} \tag{32}$$

$$J_2\left({}^E\vec{\omega}^B, \vec{V}_B\right) = \sum_i \left(J_{2,i}\right)^2 \tag{33}$$

The constrained minimization problem featuring the cost function of Eq. 33 has been shown to admit a unique solution under certain conditions.[14] This solution results in accurate estimates such that the linear least squares step to recover the FOE is unnecessary. The constraints under which the existence proof is performed are shown as Eqs. 34 and 35, respectively, where V is the known total speed of the vehicle. The constraints are required in order to resolve a velocity scale-ambiguity. Specifically, Eqs. 15a,15b, and 32 are satisfied for any scalar multiple of the vehicle velocity.

$$u > 0 \tag{34}$$

$$\|\vec{V}_B\| = V > 0 \tag{35}$$

Inclusion of the constraints shown in Eqs. 34 and 35 is problematic for the current application. As with the previous method, GPS measurements are required to obtain a value for the total velocity. These measurements are assumed unavailable for missions of interest to this paper. Therefore, an approximation such as the "trim velocity" assumption described previously is required to express the constraint shown in Eq. 35.

IV.C. Method 3

In contrast, the approach taken for the method presented here is to eliminate explicit velocity dependencies from the optimization altogether. The optical flow is decoupled according to motion contributions as described in Section III.A prior to posing the optimization problem. Consequently, the angular motion and translational motion can be treated as separate problems. Solution of the angular motion allows the conversion of optical flow measurements to their translational components. These components then yield directional velocity information in the form of the wind-axis angles, α and β. Further, the decoupled flow permits access to measurements of the parallax induced on each tracked feature point. This measure of relative depth can be leveraged to aid in feature selection and weighting for proper conditioning of the optimization.

Estimates of the aircraft angular rates are achieved through the solution of an optimization problem that does not explicitly include vehicle velocity. Specifically, the motion-decoupled optical flow is used to form a quadratic cost function that is minimized for the correct values of the aircraft states in question.

Section III.A describes the decomposition of the optical flow of a feature point, m_i, into contributions from translational motion and angular motion. Further, the translational component, or parallax, is shown to expand radially from the FOE. This condition is re-expressed here as Eq. 36.

$$\dot{\nu}_{T,i}\left(\mu_i - \mu_F\right) - \dot{\mu}_{T,i}\left(\nu_i - \nu_F\right) = 0 \tag{36}$$

Recall that the image velocity components of m_i, $(\dot{\mu}_i, \dot{\nu}_i)$, can be expressed as a sum of the components of $\vec{V}_{flow,T_i}$ and $\vec{V}_{flow,R_i}$ as in Eqs. 21a and 21b. These vectors represent the optical flow of m_i resulting from translation and rotation, respectively. The decomposed flow can be used to express the translational flow in terms of the total flow and the rotational component, as shown in Eqs. 37a and 37b.

$$\dot{\mu}_{T,i} = \dot{\mu}_i - \dot{\mu}_{R,i} \tag{37a}$$

$$\dot{\nu}_{T,i} = \dot{\nu}_i - \dot{\nu}_{R,i} \tag{37b}$$

After substitution of Eqs. 37a and 37b into Eq. 36, the FOE condition is expressed as Eq. 38. This expression captures the dependence of the measurable optical flow components, $(\dot{\mu}_i, \dot{\nu}_i)$, on the vehicle motion without explicit references to translational velocity terms. The rotational flow terms, $(\dot{\mu}_{R,i}, \dot{\nu}_{R,i})$, are defined by Eqs. 20a and 20b and are dependent only upon the measurable feature point image positions and the unknown vehicle angular velocities. The influence of the translational motion is contained entirely in the unknown FOE position, (μ_F, ν_F).

$$(\dot{\mu}_i - \dot{\mu}_{R,i})(\nu_i - \nu_F) - (\dot{\nu}_i - \dot{\nu}_{R,i})(\mu_i - \mu_F) = 0 \tag{38}$$

Eq. 38 is comprised of the three unknown angular velocities and the unknown FOE position along with measurable image quantities. No additional unknowns are added through tracking additional feature points, therefore the system is completely determined when at least five features are tracked. An optimization problem to solve for the unknown quantities can be formulated using terms taking the form of Eq. 38. After substitution of the expressions for the rotational flow components, the cost term associated with each tracked feature point is shown as Eq. 39 The cost function is then given as the sum of the squares of terms of the form shown in Eq 39 for all visible feature points, as shown by Eq. 40.

$$J_{3,i}\left({}^{E}\vec{\omega}^{B}, \mu_F, \nu_F\right) = \begin{bmatrix} p & q & r \end{bmatrix} \begin{bmatrix} \left(\nu_i^2 - \nu_i\nu_F\right) + \left(\mu_i^2 - \mu_i\mu_F\right) \\ \mu_i\nu_i(\mu_i - \mu_F) - \left(1+\mu_i^2\right)(\nu_i - \nu_F) \\ \left(1+\nu_i^2\right)(\mu_i - \mu_F) - \mu_i\nu_i(\nu_i - \nu_F) \end{bmatrix} + \begin{bmatrix} -(\nu_i - \nu_F) & (\mu_i - \mu_F) \end{bmatrix} \begin{bmatrix} \dot{\mu}_i \\ \dot{\nu}_i \end{bmatrix} \tag{39}$$

$$J_3\left({}^{E}\vec{\omega}^{B}, \nu_F, \mu_F\right) = \sum_i (J_{3,i})^2 \tag{40}$$

Finally, the minimization problem can be posed as Eq. 41. The solution of Eq. 41 yields estimates for the aircraft angular velocity, ${}^{E}\hat{\vec{\omega}}^{B}$, and for the FOE coordinates, $(\hat{\mu}_F, \hat{\nu}_F)$. While $(\hat{\mu}_F, \hat{\nu}_F)$ could be used to yield estimates of α and β via Eqs. 24a and 24b, improved accuracy can be achieved through parallax-dependent feature selection and a second optimization step similar to that discussed in Section IV.A. Hence, the currently obtained FOE coordinates are disregarded and only the estimates for angular velocity are kept.

$$\left({}^{E}\hat{\vec{\omega}}^{B}, \hat{\mu}_F, \hat{\nu}_F\right) = \arg \min_{\left({}^{E}\vec{\omega}^{B}, \mu_F, \nu_F\right)} J_3\left({}^{E}\vec{\omega}^{B}, \mu_F, \nu_F\right) \tag{41}$$

The angular rate estimates obtained through the optimization problem posed in Eq. 41 are used to decouple the optical flow and set up the linear least-squares problem shown previously as Eq. 29. Further, the most accurate information for the aerodynamic angle estimation step is achieved through the use of flow that is dominated by relative velocity. Flow vectors that exhibit a large amount of parallax give a strong indication of the FOE location. Vectors exhibiting less parallax are subject to inaccuracies resulting from measurement and estimation error. Potential improvements in the estimation accuracy of (μ_F, ν_F) can be made through proper feature selection based on parallax measurements.

In addition to selecting the largest translational flow vectors, the least-squares optimization can be further improved by weighting the vectors that give a better indication of the actual FOE. The weight for each translational flow vector is computed as a function of parallax such that larger translational flow vectors are given greater influence in the optimization.

V. EXAMPLE

V.A. Simulation

Aircraft states are estimated using data generated by a flight simulation. Specifically, a high-fidelity model of the nonlinear equations of motion for an F-16 is simulated to fly through a region of obstacles.[20] The model includes an autopilot which enables the vehicle to track a set of maneuvers and progress between waypoints using inertial sensors.

Images are generated by simulating a camera mounted on the aircraft. This camera is mounted at the center of gravity with the optical axis aligned along the nose of the vehicle. The camera model uses a 70 *deg* field of view and a capture rate of 30 *images/second*.

The environment is constructed to approximate an urban scene scaled to the dynamics of the F-16. This environment is shown in Figure 6. The building surfaces of the environment are populated with feature points in a grid pattern with a spacing of 500 *ft*. Points are scattered randomly on the ground plane according to a uniform distribution. Several points are also scattered at a large distance in all directions to simulate feature points corresponding to clouds and background clutter.

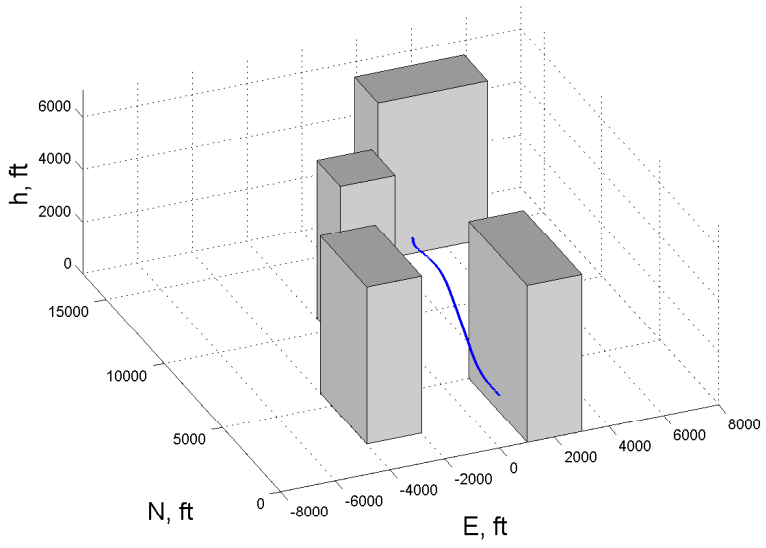


Figure 6. Virtual Environment and Simulated Trajectory

The aircraft maneuvers through the environment by following a prescribed trajectory. This trajectory, as shown in Figure 6, involves several maneuvers. The vehicle is initially flying at trim for straight and level flight at sea level. A maneuver is executed which results in an aggressive climb followed by a negative roll to an inverted configuration. This inverted flight is characterized by a positive pitch rate throughout. The aircraft then rolls back to a wings-level condition while maintaining a nose-high attitude. The position data in Figure 7 and attitude data in Figure 8 quantify these maneuvers.

A critical aspect of this trajectory is the variation in rates and velocities associated with these maneuvers. These rates are shown in Figure 9 while the velocities are represented in Figure 10 by the total airspeed along with the angle of attack and angle of sideslip. The large variations of these parameters are indicative of maneuvers, such as traverssing an urban environment by a UAV, for which vision-based feedback might be especially valuable.

V.B. State Estimation without Noise

A sequence of images are measured that do not contain any random variations due to noise. These images are effectively perfect measurements of the environment. As such, the analysis of perfect measurements should demonstrate an upper bound on the theoretical accuracy of the state estimates generated by minimizing each of the cost functions.

The estimated rates are shown in Figure 11 along with the actual rates. The minimization of each cost function is able to generate highly accurate estimates of the rates. The esimated values are visually indistinguishable from the actual values.

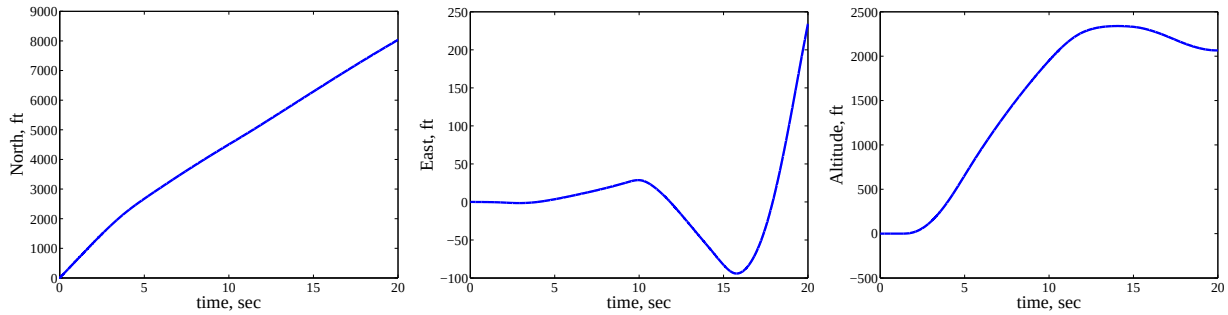


Figure 7. Vehicle position through 20 *sec* duration of maneuver

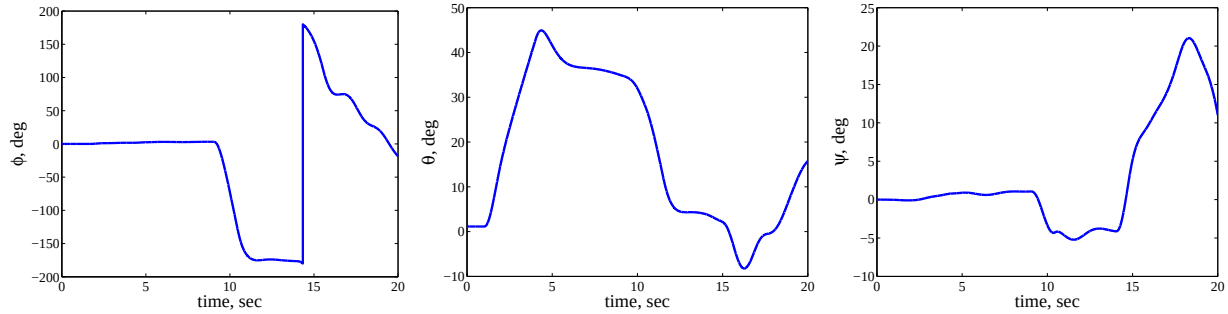


Figure 8. Vehicle attitude through 20 *sec* duration of maneuver

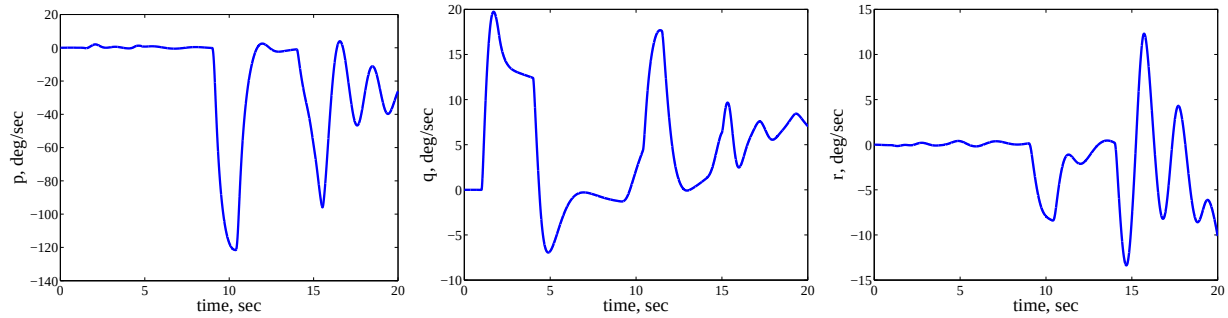


Figure 9. Vehicle rates through 20 *sec* duration of maneuver

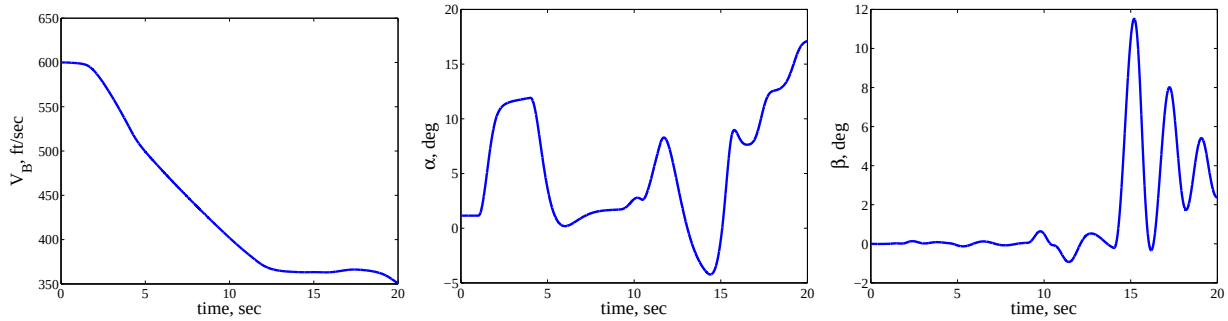


Figure 10. Vehicle velocity through 20 *sec* duration of maneuver

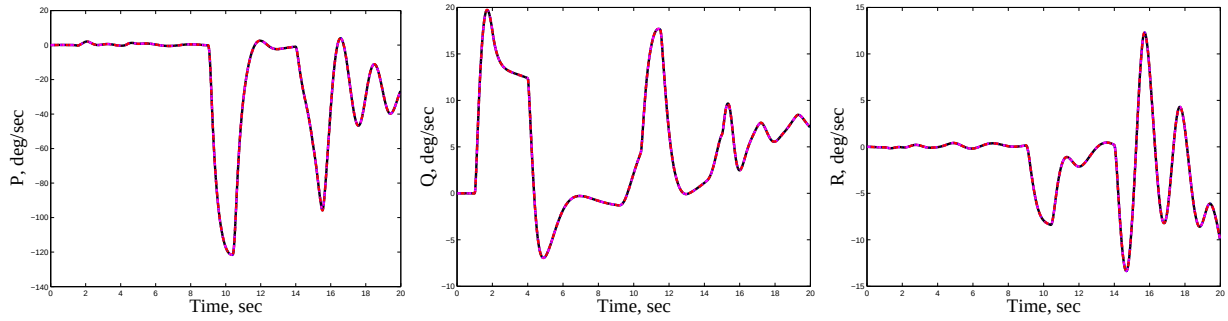


Figure 11. Angular rate estimates for the noise-free case. Results are shown for J_1 (dashed magenta), J_2 (dotted blue), and J_3 (dash-dot red) overlayed on truth value (solid black)

The estimates of aerodynamic angles are shown in Figure 12 along with actual angles. In thise case, the estimated values are quite close to the true values; however, a small deviation is noted in the data associated with the J_2 cost from Equation 32. The formulation of J_2 assumed a constant approximated "trim velocity" as described in Section IV. This condition is clearly violated by the airspeed, as shown in Figure 10, nears its minimal value at the time of the error in the angles.

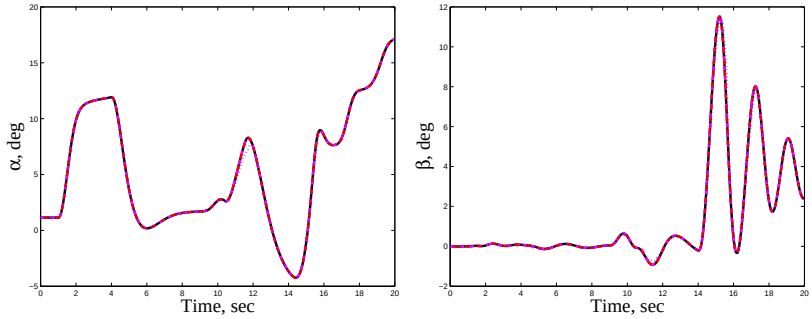


Figure 12. Aerodynamic angle estimates for the noise-free case. Results are shown for J_1 (dashed magenta), J_2 (dotted blue), and J_3 (dash-dot red) overlayed on truth value (solid black)

The accuracy of the state estimates can be approximated by the 2-norm metric of each error. These metrics, given in Table 1, demonstrate that minimizing J_3 generates the best estimates. These estimates are always improved over minimization of J_1 by a factor of at least 2 and over minimization of J_3 by an order of magnitude.

cost function	p	q	r	α	β
J_1	0.0245	0.0219	0.0449	0.2538	0.3503
J_2	0.2113	0.3002	0.4146	5.2542	5.7387
J_3	0.0059	0.0089	0.0056	0.1036	0.0552

Table 1. 2-Norm of Total error during 20 *sec* maneuver, noise-free

V.C. State Estimates with Noise

Noise is introduced to the images to reflect the performance of each cost function using imperfect measurements. A noise value up to 0.5 pixels, both horizontally and vertically, is associated with each feature point in the image. Such noise is a realistic level that could result from lens distortion, errors in calibration, and numerical errors in image processing.

The angular rates which are estimated are shown in Figure 13 to compare with the actual values. In each case, the state estimates are reasonably accurate. Slight variation is noted in each estimate; however, these variations are quite minor.

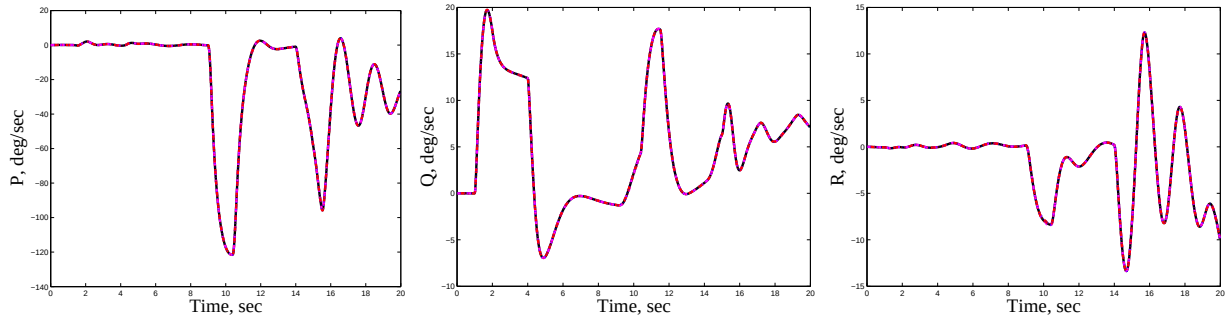


Figure 13. Angular rate estimates for the case of 0.5 pixel-width noise that is uniformly distributed about measured image position. Results are shown for J_1 (dashed magenta), J_2 (dotted blue), and J_3 (dash-dot red) overlayed on truth value (solid black)

The estimates of the aerodynamics angles, as shown in Figure 14, are somewhat more sensitive to the noise. The estimates resulting from minimization of each cost function show significant error around 10 seconds into the simulation. Additional error is present as the simulation ends.

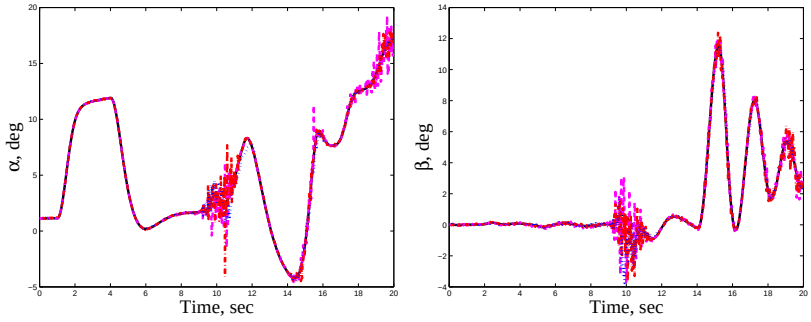


Figure 14. Aerodynamic angle estimates for the case of 0.5 pixel-width noise that is uniformly distributed about measured image position. Results are shown for J_1 (dashed magenta), J_2 (dotted blue), and J_3 (dash-dot red) overlayed on truth value (solid black)

The effect of noise is evident in the 2-norm metrics of each error signal. These values, given in Table 2, are noticeably higher than the corresponding metrics associated with noise-free images in Table 1. The cost function whose minimization resulted in the smallest error actually varies with the state estimate. The error in rates was smallest for J_1 and largest for J_2; conversely, the error in angle of attack was smallest for J_2 and largest for J_1.

cost function	p	q	r	α	β
J_1	0.2789	0.4200	0.3963	14.6860	12.5938
J_2	0.4323	0.6296	0.7359	10.8991	13.2984
J_3	0.3495	0.5204	0.5996	13.2852	12.3931

Table 2. 2-Norm of Total error during 20 *sec* maneuver, with pixel position noise

V.D. Computational Cost

The computational cost of state estimation is another metric with which to evaluate the cost functions. The application of such approaches is presumable for autonomous operation during flight; consequently, the speed at which information is available for control becomes critical. The neccessity for speed is compounded as the flight environment becomes more dense and agile maneuvering is required.

The CPU required for each estimation is given in Table 3 as determined by MATLAB. Clearly, the minimization of J_3 is achieved with a reduced workload as compared to J_1 and J_2.

cost function	without noise	with noise
J_1	9383.7	9492.5
J_2	1678.3	1995.1
J_3	680.3	791.4

Table 3. Total CPU time in seconds

VI. CONCLUSION

This paper has presented a technique for partial aircraft state estimation using vision as a sensor. Optical flow is used to compute accurate estimates of the aircraft body axis angular rates as well as the aerodynamic angles that determine the orientation of the velocity vector. Estimation accuracy is aided by selectively tracking those feature points more likely to exhibit motion that is beneficial to the estimation process. Development of this purely visual technique is a step towards using vision for flight control in complex and cluttered environments.

VII. ACKNOWLEDGMENTS

This work was supported jointly by the Air Force Research Laboratory and the Air Force Office of Scientific Research under F49620-03-1-0381 with Johnny Evers, Neal Glassman, Sharon Heise, and Robert Sierakowski as project monitors.

References

[1]Lucas, B., and Kanade, T., "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. of the DARPA Image Understanding Workshop*, 1981, pp.121-130.

[2]Kanade, T., Amidi, O., and Ke, Q., "Real-Time and 3D Vision for Autonomous Small and Micro Air Vehicles," *43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, December 2004.

[3]Yao, Y.S., Chellapa, R., "Dynamic Feature Point Tracking in an Image Sequence," 12^{th} *IAPR International Conference on Pattern Recognition*, Vol. 1, October 1994, pp. 654-657.

[4]He, Z., Iyer, R.V., and Chandler, P.R., "Vision-Based UAV Flight Control and Obstacle Avoidance," *Proc. of the 2006 IEEE American Control Conference*, Minneapolis, MN, June, 2006.

[5]Horn, B., and Schunck, B., "Determining Optical Flow," *Artificial Intelligence*, Vol. 17, pp. 185-203, October 1981.

[6]Beauchemin, S., and Barron, J., "The Computation of Optical Flow," *ACM Computing Surveys*, Vol. 27, No. 3, September 1995, pp. 433-467.

[7]Soatto, S., Frezza, R., and Perona, P., "Motion Estimation via Dynamic Vision," *IEEE Transactions on Automatic Control*, Vol. 41, No. 3, 1996, pp. 393-413.

[8]Soatto, S., Perona, P., "Recursive 3-D Visual Motion Estimation Using Subspace Constraints," *International Journal of Computer Vision*, Vol. 22, No. 3, 1997, pp. 235-259.

[9]Gurfil, P., and Rotstein, H., "Partial Aircraft State Estimation from Visual Motion Using the Subspace Constraints Approach," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 5, 2001, pp. 1016-1028.

[10]Webb, T., Prazenica, R., Kurdila, A., and Lind, R., "Vision-Based State Estimation for Autonomous Micro Air Vehicles," *Proc. of the AIAA Guidance, Navigation, and Control Conference*, AIAA-2004-5349, Providence, RI, August 2004.

[11]Webb, T., Prazenica, R., Kurdila, A., and Lind, R., "Vision-Based State Estimation for Uninhabited Aerial Vehicles," *Proc. of the AIAA Guidance, Navigation, and Control Conference*, AIAA-2005-5869, San Francisco, CA, August 2005.

[12]Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S., *An Invitation to 3-D Vision: From Images to Geometric Vision*, New York, NY:Springer-Verlag, 2004.

[13]Gebert, G., Snyder, D., Lopez, J., Siddiqi, N., and Evers, J., "Optical Flow Angular Rate Determination," *Proc. of the International Conference on Image Processing*, Vol. 1, 2003, pp. 949-52.

[14]Iyer, R.V., He, Z., and Chandler, P.R., "On the Computation of the Ego-Motion and Distance to Obstacles for a Micro Air Vehicle," *Proc. of the 2006 IEEE American Control Conference*, Minneapolis, MN, June, 2006.

[15]Kehoe, J., Causey, R., Arvai, A., and Lind, R., "Partial Aircraft State Estimation from Optical Flow Using Non-Model-Based Optimization," *Proc. of the 2006 IEEE American Control Conference*, Minneapolis, MN, June, 2006.

[16]Causey, R., and Lind, R., "Aircraft-Camera Equations of Motion," *AIAA Journal of Aircraft*, In Preparation.

[17]Barrows, G.L., and Neely, C., "Mixed-Mode VLSI Optic Flow Sensors for In-Flight Control of a Micro Air Vehicle," *Proceedings of the* 45^{th} *Annual SPIE*,San Diego, CA, July, 2000.

[18]Roderick, A., Kehoe, J., and Lind, R., "Vision-Based Navigation Using Multi-Rate Feedback from Optic Flow and Scene Reconstruction," *Proc. of the AIAA Guidance, Navigation, and Control Conference*, AIAA-2005-6093, San Francisco, CA, August 2005.

[19]Branca, A., Stella, E., and Distante, A., "Passive Navigation using Focus of Expansion," *Proc. of the* 3^{rd} *IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, December 1996, pp. 64-69.

[20]Nguyen, L, Ogburn, M., Gilbert, W., Kibler, K., Brown, P., and Deal, P., “Simulator Study of Stall/Post-Stall Characteristics of a Fighter Airplane with Relaxed Longitudinal Static Stability,” Tech. Report 1538, NASA Langley Research Center, Hampton, VA, 1979.