UNMANNED SYSTEMS PAPER

Vision-based Target Geo-location using a Fixed-wing Miniature Air Vehicle

**D. Blake Barber · Joshua D. Redding ·
Timothy W. McLain · Randal W. Beard ·
Clark N. Taylor**

Abstract This paper presents a method for determining the GPS location of a ground-based object when imaged from a fixed-wing miniature air vehicle (MAV). Using the pixel location of the target in an image, measurements of MAV position and attitude, and camera pose angles, the target is localized in world coordinates. The main contribution of this paper is to present four techniques for reducing the localization error. In particular, we discuss RLS filtering, bias estimation, flight path selection, and wind estimation. The localization method has been implemented and flight tested on BYU's MAV testbed and experimental results are presented demonstrating the localization of a target to within 3 m of its known GPS location.

Key words computer vision · geo-location · localization · micro air vehicles · unmanned air vehicles

1 Introduction

Unmanned air systems are prime candidates for tasks involving risk and repetition, or what the military calls the "dull, dirty and dangerous" [13]. For tasks that involve tracking, reconnaissance, and delivery, one objective of unmanned air systems is to accurately determine the location of ground-based objects.

This paper presents a method for determining the location of objects in world/inertial coordinates using a gimballed EO/IR camera on board a fixed-wing miniature air vehicle (MAV). We focus on fixed-wing UAVs (as opposed to rotary wing aircraft or blimps) due to the unique benefits available from fixed-wing aircraft,

D. B. Barber · T. W. McLain · R. W. Beard (✉) · C. N. Taylor
Magicc Lab, Brigham Young University, Provo, UT 84602, USA
e-mail: beard@ee.byu.edu

J. D. Redding
Scientific Systems Company, Inc., Woburn, MA 01801, USA

including: adaptability to adverse weather, enhanced fuel efficiency, a shorter learning curve for the untrained operator, and extreme durability in harsh environments. Also, minimum airspeed requirements associated with fixed-wing aircraft requires images to be provided from multiple vantage points, allowing for more robust localization.

In this paper we have assumed that the target is identified in the video stream by a human end user. The target is then automatically tracked at frame rate using a combination of color segmentation and feature tracking [11]. After the object has been identified in the video stream and an initial estimate of its world coordinates has been determined, the MAV adjusts its path autonomously in order to orbit the object and collect additional information that is used to further enhance the estimate. Due to the specific nature of MAVs, there are several sources of error that affect the position estimates. In this paper, we analyze the error sources and present four steps to enhance the accuracy of the estimated target location.

While vision-based localization is well understood, previously published results focus on unmanned ground vehicles [4, 17], or stationary air vehicles such as a blimps [3] or rotorcraft [21]. However, blimps are not well suited for use in high winds or inclement weather, and the costs and complexities associated with rotorcraft are non-trivial. The objective of this paper is to explore localization methods using fixed-wing MAVs which tend to be more robust and less-expensive platforms.

Previous results on geo-locating objects from fixed-wing aircraft have several limitations not present in the system described in this paper. In [9, 10], all information collected by an aerial camera is accurately geo-located through registration with pre-existing geo-reference imagery. In contrast, our system focuses on geo-locating a specific object in the video stream and does not require pre-existing geo-referenced imagery. A method for creating geo-referenced mosaics from aerial video is presented in [18], however, this method assumes an extremely accurate IMU that is impractical for MAVs due to weight and power restrictions.

Several previous works on target tracking/localization from UAVs are focused on control of the UAV to keep the object in view, as opposed to actually geo-locating the objects [7, 16, 19, 22]. In [16] flight paths for fixed-wing UAVs are designed to maintain a constant line-of-sight with a ground-based target. Stolle and Rysdyk [19] present similar results with some useful details on camera control. Both references focus on pointing a UAV-mounted camera at a known target location and present simulation results. The accuracy of the localization results is not discussed.

The geo-location system presented in [8] is similar to our work, however, the reported errors are in excess of 20 m, while our method achieves localization errors under 5 m. Whang et al. [23] and Dobrokhodov et al. [5] describe a geo-location solution that is similar to the work presented in this paper. Range estimates in [5] are obtained using a terrain model, and a nonlinear filter is used to estimate the position and velocity of moving ground based targets. Campbell and Wheeler [2] also presents a vision based geolocation system that is similar to our solution. The estimation scheme proposed in [2] is based on a square root sigma point filter and can handle moving objects. Bounds on the localization error are explicitly derived from the filter. However, the results presented in [5] and [2] both exhibit biases in the estimate, and neither paper addresses the sensitivity of the solution to heavy wind conditions. Early versions of the results appearing in this paper are presented in [14].

The remainder of the paper is organized as follows. In Section 2, we present the basic mathematics used to obtain the raw target localization estimates from a single frame within the video. In Section 3 we discuss four techniques for improving the localization estimate of the target. We present flight results demonstrating the effectiveness of our method in Section 4, and offer some concluding remarks in Section 5.

2 The Geometry of Geo-location

In this section, we present our method for generating raw estimates of the target's location in the world frame. We assume throughout the paper that the target's pixel location in the video image is known. Experimental results are obtained by allowing a user to select the target to be imaged and using a color segmentation algorithm to track the target in the image plane.

2.1 Coordinate Frames

The coordinate frames associated with this problem include the inertial frame, the vehicle frame, the body frame, the gimbal frame, and the camera frame. Figures 1 and 2 show schematics of the different coordinate frames. The inertial frame, denoted by (X_I, Y_I, Z_I), is a fixed frame with X_I directed North, Y_I directed East, and Z_I directed toward the center of the earth. The vehicle frame, denoted by (X_v, Y_v, Z_v), is oriented identically to the inertial frame but its origin is at the vehicle center of mass. The body frame, denoted by (X_b, Y_b, Z_b) also originates at the center of mass but is fixed in the vehicle with X_b pointing out the nose, Y_b pointing out the right wing, and Z_b pointing out the belly. As shown in Figures 1 and 2, the gimbal frame, represented by (X_g, Y_g, Z_g) originates at the gimbal rotation center and is oriented so that X_g points along the optical axis, Z_g points down in the image plane, and Y_g points right in the image plane. The camera frame, denoted (X_c, Y_c, Z_c), originates at the optical center with X_c pointing up in the image, Y_c pointing right in the image plane, and Z_c directed along the optical axis.

Figure 1 A graphic showing a lateral view of the coordinate frames. The inertial and vehicle frames are aligned with the world, the body frame is aligned with the airframe, and the gimbal and camera frames are aligned with the camera.
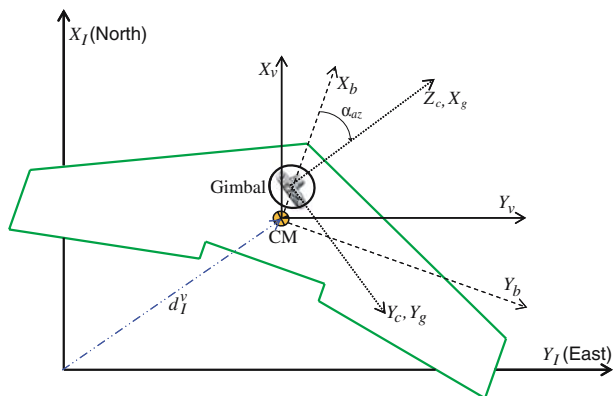
Figure 2 A graphic showing a longitudinal view of the coordinate frames.
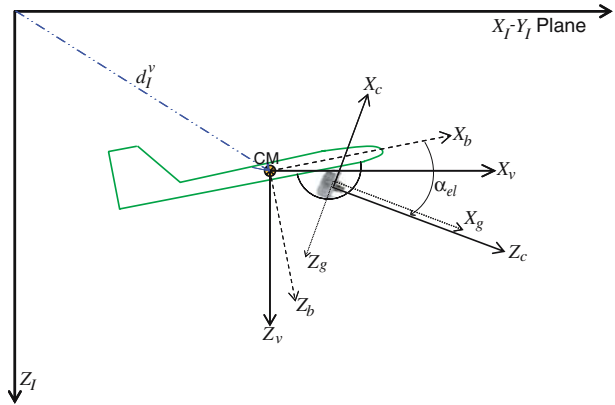
The notation v^i implies that vector v is expressed with respect to frame i. The rotation matrix and the translation vector from frame i to frame j are denoted by R_i^j and d_i^j respectively. The homogeneous transformation matrix from frame i to frame j is given by

$$T_i^j = \begin{pmatrix} R_i^j & -d_i^j \\ \mathbf{0} & 1 \end{pmatrix}, \tag{1}$$

where $\mathbf{0} \in \mathbb{R}^3$ is a row vector of zeros. Note that d_i^j is resolved in the jth coordinate frame. The inverse transformation is given by

$$T_j^i \triangleq T_i^{j-1} = \begin{pmatrix} R_i^{jT} & R_i^{jT} d_i^j \\ \mathbf{0} & 1 \end{pmatrix}.$$

The transformations used in this papers are defined in Table I. The derivation for each of the transformations will be discussed below.

Table I Homogeneous transformation matrices

Transformation	Description
T_I^v	Inertial to MAV vehicle frame
T_v^b	MAV vehicle to MAV body frame
T_b^g	MAV body to gimbal frame
T_g^c	Gimbal to camera frame

2.1.1 Transformation from the Inertial to the Vehicle Frame

The transformation from the inertial to the vehicle frame is a simple translation. Therefore T_I^v is given by

$$T_I^v = \begin{bmatrix} I & -d_I^v \\ 0 & 1 \end{bmatrix}, \qquad \text{where}$$

$$d_I^v = \begin{bmatrix} x_{\text{MAV}} \\ y_{\text{MAV}} \\ -h_{\text{MAV}} \end{bmatrix}, \tag{2}$$

and where x_{MAV} and y_{MAV} represent the North and East location of the MAV as measured by its GPS sensor, and h_{MAV} represents the MAV's altitude as measured by a calibrated, on-board barometric pressure sensor.

2.1.2 Transformation from the Vehicle to the Body Frame

The transformation from the vehicle frame to the MAV body frame, T_v^b, consists of a rotation based on measurements of Euler angles. If ϕ, θ and ψ represent the MAV's roll, pitch and heading angles in radians, then the transformation is given by

$$T_v^b = \begin{bmatrix} R_v^b & 0 \\ 0 & 1 \end{bmatrix}, \qquad \text{where}$$

$$R_v^b = \begin{bmatrix} c_\theta c_\psi & c_\theta s_\psi & -s_\theta \\ s_\phi s_\theta c_\psi - c_\phi s_\psi & s_\phi s_\theta s_\psi + c_\phi c_\psi & s_\phi c_\theta \\ c_\phi s_\theta c_\psi + s_\phi s_\psi & c_\phi s_\theta s_\psi - s_\phi c_\psi & c_\phi c_\theta \end{bmatrix} \tag{3}$$

and where $c_\varphi \triangleq \cos \varphi$ and $\sin_\varphi \triangleq \sin \varphi$. On our platform, the Euler angles are estimated by a two-stage Kalman filter as described in Eldredge [6]. The Kalman filter uses rate gyros for the propagation model, and accelerometers for the measurement update.

2.1.3 Transformation from the Body to the Gimbal Frame

The transformation from the MAV body to the gimbal frame, T_b^g, will depend on the location of the MAV's center of mass with respect to the gimbal's rotation center. This vector, denoted by d_b^g, is resolved in the gimbal frame. T_b^g will also depend on the rotation that aligns the gimbal's coordinate frame with the MAV's body frame. This rotation is denoted R_b^g and requires measurements of the camera's azimuth and elevation angles. Let α_{az} denotes the azimuth angle of rotation about Z_g, and α_{el} the

elevation angle of rotation about Y_g, after α_{az}. Both α_{az} and α_{el} can be deduced from the gimbal servo commands. The transformation is given by

$$T_b^g = \begin{bmatrix} R_b^g & -d_b^g \\ 0 & 1 \end{bmatrix}, \qquad \text{where}$$

$$R_b^g = R_{y,\alpha_{el}} R_{z,\alpha_{az}}$$

$$= \begin{bmatrix} c_{el} & 0 & s_{el} \\ 0 & 1 & 0 \\ -s_{el} & 0 & c_{el} \end{bmatrix} \begin{bmatrix} c_{az} & s_{az} & 0 \\ -s_{az} & c_{az} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} c_{el}c_{az} & c_{el}s_{az} & s_{el} \\ -s_{el} & c_{az} & 0 \\ -s_{el}c_{az} & -s_{el}s_{az} & c_{el} \end{bmatrix}. \tag{4}$$

### 2.1.4 Transformation from the Gimbal to the Camera Frame

The transformation from gimbal to camera reference frames, T_g^c, depends on the vector d_g^c, which describes the location of the gimbal's rotation center relative to the camera's optical center and is resolved in the camera's coordinate frame. T_g^c also depends on a fixed rotation R_g^c, which aligns the camera's coordinate frame with that of the gimbal since we have chosen $X_c = -Z_g$ and $Z_c = X_g$. The transformation is given by

$$T_g^c = \begin{bmatrix} R_g^c & -d_g^c \\ 0 & 1 \end{bmatrix}, \text{ where}$$

$$R_g^c = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{5}$$

2.2 Camera Model

A simple camera projection model is shown in Figure 3. The point $q = (x_{ip}, y_{ip}, 1, 1)^T$ is the homogeneous projection of the point $p_{obj}^c = (p_x, p_y, p_z, 1)^T$ onto the image

Figure 3 A graphic showing the coordinate frames associated with the camera. The coordinate frame represented by $\{X_c, Y_c, Z_c\}$ has origin at the camera *center* and its elements have units of meters. The frame $\{X_{im}, Y_{im}, Z_{im} = Z_c - f\}$ is *centered* at the image plane and has units of meters. The frame (X_{ip}, Y_{ip}) is *centered in the upper left hand corner* of the image and has units of pixels.
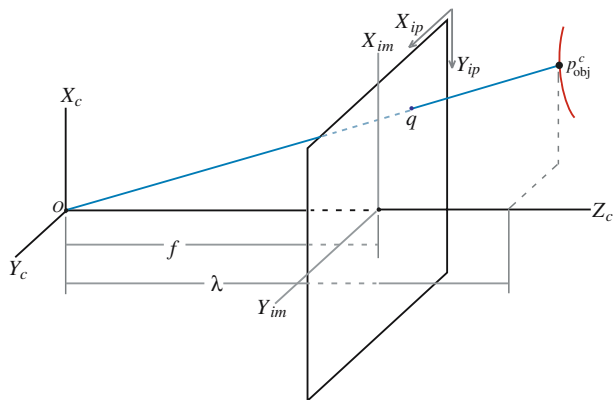
plane in pixels, where p_{obj}^c denotes the location of an object p relative to the center of the camera. Trucco and Verri [20] show that the change from pixels to meters in the image frame is accomplished by

$$
\begin{aligned}
x_{im} &= (-y_{ip} + 0_y)S_y \\
y_{im} &= (x_{ip} - 0_x)S_x \,,
\end{aligned} \tag{6}
$$

where the units of (x_{ip}, y_{ip}) are pixels and the units of (x_{im}, y_{im}) are meters. The parameters 0_x and 0_y denote the x and y offsets to the center of the image from the upper-left hand corner in pixels, and S_x and S_y denote the conversion factors from pixels to meters.

By similar triangles we get that

$$
\frac{x_{im}}{f} = \frac{p_x}{p_z}, \qquad \text{and} \qquad \frac{y_{im}}{f} = \frac{p_y}{p_z},
$$

where f is the focal length of the camera. Using Eq. 6, and defining $\lambda \triangleq p_z$ we get

$$
\Lambda q = \underbrace{\begin{bmatrix} 0 & f_x & 0_x & 0 \\ -f_y & 0 & 0_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{C} p_{\text{obj}}^c \,, \tag{7}
$$

where $f_x \triangleq \frac{f}{S_x}$, $f_y \triangleq \frac{f}{S_y}$, and $\Lambda = \begin{pmatrix} \lambda I & 0 \\ 0 & 1 \end{pmatrix}$. The matrix C is known as the calibration matrix.

Our objective is to determine p_{obj}^I, the object's position in the inertial frame. Using the homogeneous transformations derived in the previous sections we have

$$
\Lambda q = C p_{\text{obj}}^c = C T_g^c T_b^g T_v^b T_I^v p_{\text{obj}}^I.
$$

Solving for p_{obj}^I gives

$$
p_{\text{obj}}^I = \left[C T_g^c T_b^g T_v^b T_I^v \right]^{-1} \Lambda q. \tag{8}
$$

Therefore, p_{obj}^I can be determined when λ is known.

2.3 Image Depth

The image depth λ refers to the distance along the camera's optical axis to the object of interest in the image [11]. In this paper we describe a technique for estimating λ based on a flat earth assumption. A similar technique can be used if a terrain map is available.

Let p_{cc} be the location of the camera's optical center. If p_{cc} is resolved in the camera frame we have $p_{cc}^c = (0, 0, 0, 1)^T$. Therefore, resolving in the inertial frame gives

$$p_{cc}^I = \begin{pmatrix} x_{cc}^I \\ y_{cc}^I \\ z_{cc}^I \\ 1 \end{pmatrix} = \left[T_g^c T_b^g T_v^b T_I^v \right]^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{9}$$

Figure 4 shows the location $q = [x_{ip} \; y_{ip} \; 1 \; 1]^T$. Define q_{obj}^I as q resolved in the inertial frame, i.e.,

$$q_{obj}^I = \begin{pmatrix} x_{obj}^I \\ y_{obj}^I \\ z_{obj}^I \\ 1 \end{pmatrix} = \left[CT_g^c T_b^g T_v^b T_I^v \right]^{-1} q. \tag{10}$$

Note from Figure 4 that the flat earth model implies that the relationship between the z-components of q_{obj}^I and p_{cc}^I is given by

$$0 = z_{cc}^I + \lambda \left(z_{obj}^I - z_{cc}^I \right) . \tag{11}$$

If a terrain model is known, the zero on the left-hand side of Eq. 11 would be modified to reflect the altitude at the point where the optical axis intersects the terrain. Since both z_{cc}^I and z_{obj}^I are known from Eqs. 9 and 10 respectively, λ can be computed as

$$\lambda = \frac{z_{cc}^I}{z_{cc}^I - z_{obj}^I}. \tag{12}$$

Since Z_I is defined positive toward the center of the earth, z_{cc}^I will be negative for flight altitudes greater than the calibrated zero. Thus, Eq. 12 yields a positive value for λ, as expected.

Figure 4 The range to the target λ, is estimated using a *flat earth model* and knowledge of the location and orientation of the MAV and its camera system.
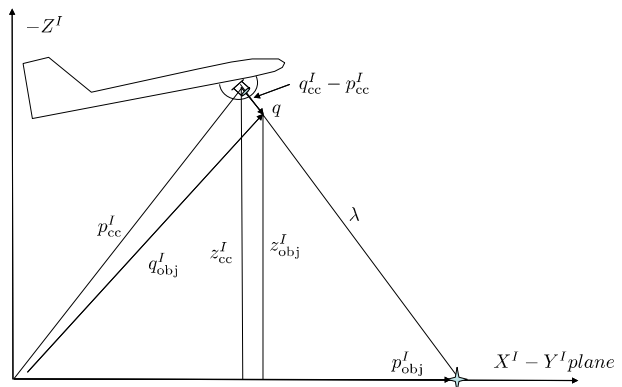
2.4 Target Location

Given λ, the inertial location of the object is given by

$$p_{\text{obj}}^I = \left[C T_g^c T_b^g T_v^b T_I^v \right]^{-1} \Lambda q$$
$$= T_v^I T_b^v T_g^b T_c^g C^{-1} \Lambda q , \tag{13}$$

or equivalently, in the more computationally efficient form

$$\bar{p}_{\text{obj}}^I = \bar{p}_{\text{cc}}^I + \lambda \left(\bar{q}_{\text{obj}}^I - \bar{p}_{\text{cc}}^I \right), \tag{14}$$

where $\bar{p}$ represents the first three elements of p.

Using these equations, we can estimate the geo-location of a target using the telemetry data from the MAV and a time-synchronized video frame containing the target. Unfortunately, every term on the right-hand side of Eq. 13 is computed using measured (i.e. noisy and biased) data. In particular, the transformation matrices (T) and Λ are computed using sensor readings, which for MAVs are typically low grade. In the next section we discuss the effects of low quality sensors on the estimation error and introduce four techniques that can be used to reduce the error.

3 Enhancing the Geo-location Accuracy

Sensor noise and uncertainty in the MAV geometry introduces error in the geo-location estimate provided by Eq.(13). Figure 5 shows the results of a flight test using the MAV system described in Section 4.1. The MAV was commanded to orbit the target location and a color segmentation algorithm was used to track the target location in the image. The error (in meters) of the raw estimates of the geo-location of the target are shown in Figure 5. The raw estimates have errors that typically range from 20 to 40 m.

Figure 5 The error, in meters, of raw geo-location estimates obtained by using Eq. 13. Sensor noise and geometric uncertainties result in typical estimation errors of 20 to 40 m.
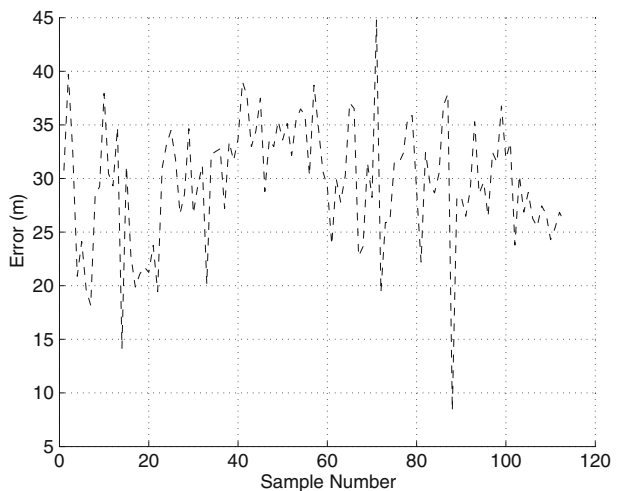
The primary contribution of this paper is to propose four techniques for enhancing the accuracy of the geo-location estimate. These techniques include: (1) recursive least squares filtering, (2) bias estimation, (3) flight path selection, and (4) wind estimation. Each technique is discussed in more detail below.

3.1 Recursive Least Squares

As shown in Figure 5, there is significant noise in the estimation error. In this paper, we assume that the target location is stationary. Therefore, a well known technique to remove the estimation error is to use a recursive least squares (RLS) filter [12]. The RLS filter minimizes the average squared error of the estimate using an algorithm that only requires a scalar division at each step and is therefore suitable for on-line implementation.

The result of using the RLS filter on the data shown in Figure 5 is shown in Figure 6. Note that the RLS filter quickly converges to an error of approximately 5 m. While the improvement in geo-location accuracy is significant, it will be shown in the following three sections that it is possible to further improve the geo-location accuracy by exploiting the structure inherent in the problem.
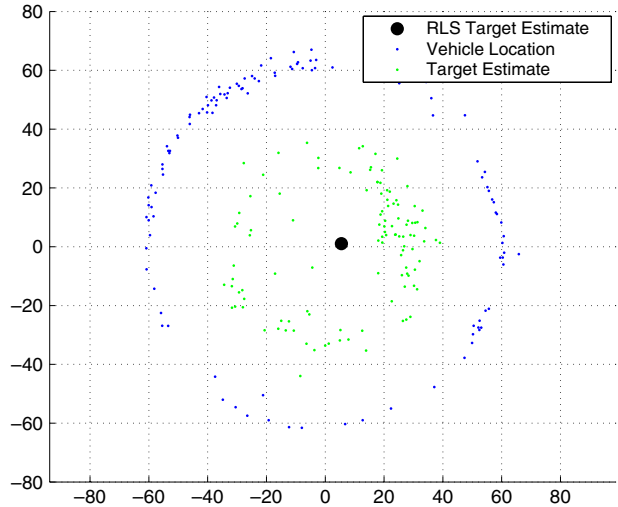
3.2 Bias Estimation

The sensor noise and the geometric uncertainties introduce both zero-mean noise and a constant bias. While the RLS algorithm is effective at removing the zero-mean noise, it is not effective at removing the bias. The geo-location error is particularly sensitive to biases in the roll and the gimbal azimuth measurement. Although bias errors can be mitigated by advanced attitude estimation schemes and precise mounting and calibration of the camera gimbal, it is impossible to totally remove these bias errors.

Fortunately, by executing a loiter pattern around a specific object, the biases and zero-mean noise can be easily distinguished. Because the bias errors are uncorrelated

Figure 6 Result of using the RLS algorithm. The error in the geo-location estimate decreases from 20 to 40 m to approximately 5 m.
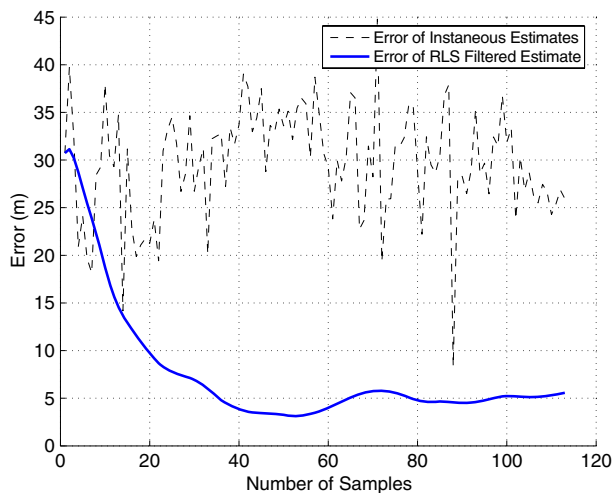
Figure 7 Localization error
before gimbal calibration. The
errors in the localization
estimates exhibit a *circular
pattern* about the target
location due to the biases
introduced by imprecisely
calibrated sensors and
geometric modeling errors.

with respect to position along the desired flight path, and the flight path is symmetric
about the target, the bias errors result in geolocation estimates that are also sym-
metric about the target. For the case of a circular flight path centered at the target,
this results in the localization estimates forming a ring around the desired target, as
shown in Figure 7.

If the biases are removed from the localization estimates, the geo-location er-
rors collapse to a 2-D Gaussian distribution centered at the object, as shown in
Figure 8. The covariance of the distribution is a function of the zero-mean noise on
raw attitude estimates and the selected radius and altitude of the loiter trajectory.

Since the biases may change from flight to flight, an on-line learning algorithm
was developed to estimate and remove them. The algorithm exploits the observation
that biases add a ring-like structure to the location estimates, effectively increasing

Figure 8 Localization error
after gimbal calibration. The
structured bias in the estimates
has been removed.
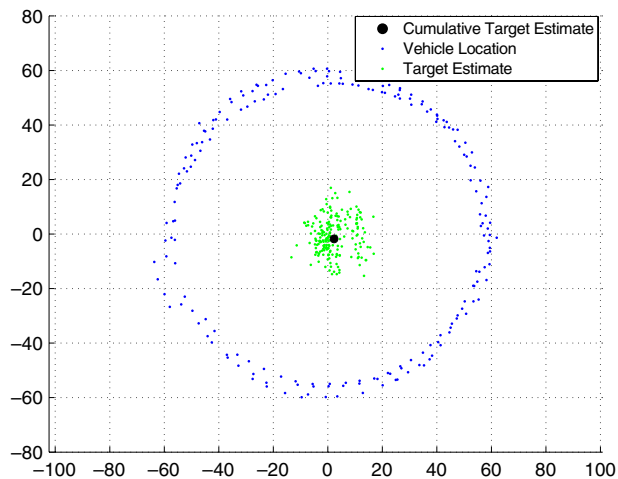
the variance of the estimates. Therefore, if the flight path is a circular orbit about the target and the bias errors are uncorrelated with position along the flight path, then the distribution of location estimates with the smallest variance will be obtained from the unbiased estimate of the target location. As a result, the bias estimation problem can be posed as the following optimization problem:

$$\min_{\bar{\alpha}_{az}, \bar{\alpha}_{el}, \bar{\phi}, \bar{\theta}, \bar{\psi}, \bar{z}} \sigma^2_{\text{localization}}(\bar{\alpha}_{az}, \bar{\alpha}_{el}, \bar{\phi}, \bar{\theta}, \bar{\psi}, \bar{z}) \tag{15}$$

where $\bar{\alpha}_{az}$, $\bar{\alpha}_{el}$, $\bar{\phi}$, $\bar{\theta}$, $\bar{\psi}$, and $\bar{z}$ are the biases associated with the measurements of gimbal azimuth, gimbal elevation, roll, pitch, yaw, and altitude, respectively.

For the fixed-wing MAVs used in this paper, the center of mass and the gimbal center are located close to each other. Therefore, as can be seen from Figure 1 the rotation axes for heading ψ, and gimbal azimuth angle α_{az} are nearly aligned, making biases introduced by these quantities, virtually indistinguishable. In an orbit pattern, the gimbal azimuth angle will be close to 90°, which implies that the airframe roll axis and the gimbal elevation axis will be nearly aligned, again making biases introduced by ϕ and α_{el} nearly indistinguishable. Even when the flight path is not an orbit, if the body pitch angle is close to zero, then biases introduced by the roll and heading measurements are indistinguishable from biases introduced by gimbal elevation and azimuth measurements, respectively. For the MAVs used in this paper, the angle of attack is approximately 5°, implying that the pitch angle is close to zero for constant altitude flight patterns. Extensive flight testing has also shown that for certain altitude–orbit radius pairs, the estimation error is not significantly affected by biases in pitch and altitude. Therefore, bias estimation can be reduced to the following optimization problem:

$$\min_{\bar{\alpha}_{az}, \bar{\alpha}_{el}} \sigma^2_{\text{localization}}(\bar{\alpha}_{az}, \bar{\alpha}_{el}). \tag{16}$$

This problem is solved on-line using a quasi-Newton based method.

Once the biases have been determined, their effects are removed by using the corrected measurements for α_{el} and α_{az} in Eq. 13 to obtain unbiased raw estimates. The effects of bias estimation and correction on the dispersion of raw target estimates can be seen in Figure 8. It is clear from Figure 8 that the ring structure characteristic of bias errors has been dramatically reduced.

3.3 Flight Path Selection

With the bias removed, we turn our attention to minimizing the variance of the resulting zero-mean estimation error. The variance is primarily due to noisy estimates of the attitude and the position of the MAV. Redding [15] presents a study of the sensitivity of Eq. 13 to errors in attitude and position. The conclusion of that study is that for circular orbits, the geo-location estimate is most sensitive to errors in roll. However, the sensitivity is a strong function of altitude and orbit radius. As shown in Figure 9, as the altitude of the MAV increases, the distance to the target also increases. Therefore, a fixed error in roll, produces a localization error that increases

Figure 9 The sensitivity of the localization error to imprecise attitude estimates, is highly dependent on altitude. At *low altitudes*, the sensitivity is due to the *obliqueness of the angle* to the target. At *high altitudes*, the sensitivity is due to distance from the target.

Localization error sensitivity increases with altitude

Localization error sensitivity increases with obliqueness
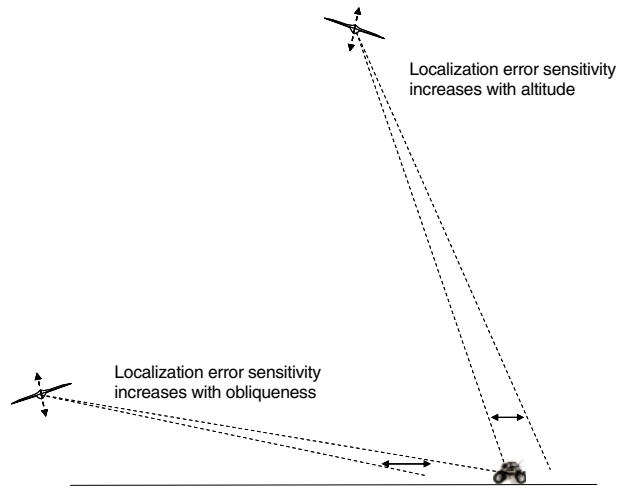
with altitude. On the other hand, Figure 9 illustrates that low altitudes also increase the error sensitivity since the angle to the target becomes more oblique as altitude decreases. For an identical error in roll, increasingly oblique angles produce a larger localization error. To explore the relationship between sensitivity to roll and the altitude and orbit radius, consider the simplified situation shown in Figure 10, where we have assumed that the camera is aligned with the right wing and is pointing directly at the target, and that the pitch angle is zero. Therefore, the nominal roll angle is $\phi_{\mathrm{nom}} = \tan^{-1} \frac{h}{R}$ where h is the altitude and R is orbit radius. If the roll angle deviates from the nominal value by $\delta\phi$, Eq. 13 will indicate a geo-location of $R - \delta R$ instead of R. For the simplified geometry shown in Figure 10 we have that

$$R - \delta R = \frac{h}{\tan(\phi + \delta\phi)}.$$

Figure 10 Simplified geometry used to derive an expression for the sensitivity of the localization error to the roll angle as a function of the orbit altitude and radius.

$$R$$

$$\phi_{\mathrm{nom}} = \tan^{-1} \frac{h}{R}$$

$$\delta\phi$$

$$h$$

$$\delta R$$
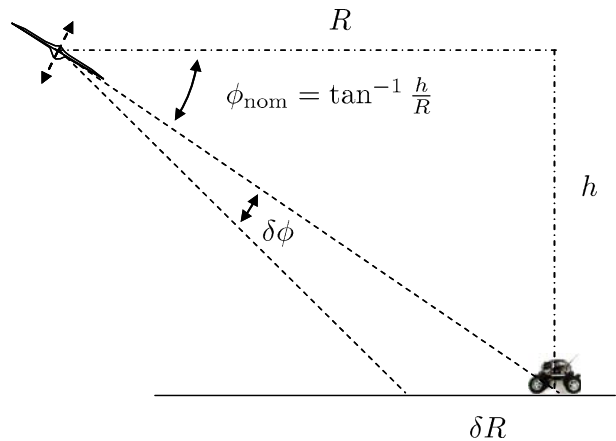
Therefore, using the relations $\tan(A + B) = \frac{\tan(A) + \tan B}{1 - \tan(A)\tan(B)}$ and $\tan\phi = \frac{h}{R}$, we obtain

$$
\begin{aligned}
\delta R &= R - \frac{h}{\tan(\phi + \delta\phi)} \\
&= R - h\frac{1 - \tan\phi\tan\delta\phi}{\tan\phi + \tan\delta\phi} \\
&= R - h\frac{R - h\tan\delta\phi}{h + R\tan\delta\phi} \\
&= \frac{(R^2 + h^2)\tan\delta\phi}{h + R\tan\delta\phi}.
\end{aligned}
\tag{17}
$$

Figure 11 shows a plot of Eq. 17 as a function of h for $\delta\phi = 5°$ and $R = 100$ m. It is clear that for a fixed radius, there is an optimal altitude that minimizes the sensitivity of the localization error to deviations in the roll attitude. The optimal altitude is found by differentiating Eq. 17 with respect to h and solving for the unique minimizer:

$$
h^* = R\left(\frac{1 - \sin\delta\phi}{\cos\delta\phi}\right).
\tag{18}
$$

Therefore, if we have an estimate for the average (or maximum) roll attitude error, and there is a desired orbit radius, e.g., the minimum turn radius, then Eq. 18 indicates the appropriate altitude for minimizing the sensitivity of the geo-location estimate to errors in the roll attitude measurement.

In addition, the computer vision algorithm may require a specific number of pixels-on-target to effectively track the target. In order to talk more generally about the notion of pixels-on-target, we define the *pixel density* to be the number of pixels imaging a square meter of area on the ground. Let μ denote pixel density in units

Figure 11 A plot of the geo-location error as a function of altitude for a fixed radius and a fixed roll attitude error. The optimal altitude h^* in indicated by a *circle*.
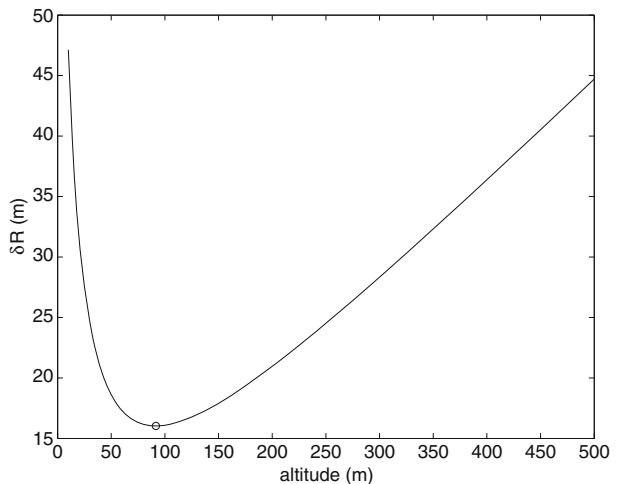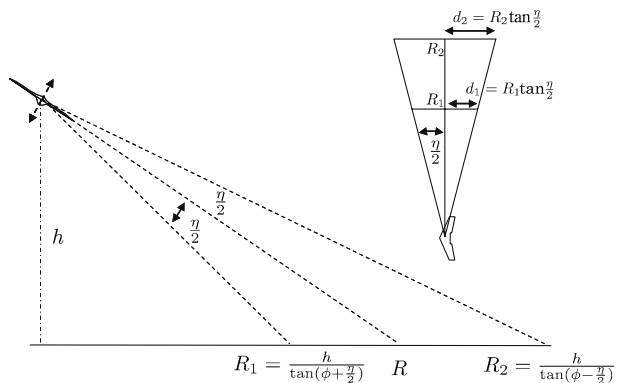
of pixels per meters squared. If η is the field of view of the lens (assumed equal in both directions) then the area imaged by the camera can be computed by referencing Figure 12. The total area is given by

$$
\begin{aligned}
\text{Area} &= (d_1 + d_2)(R_2 - R_1) \\
&= (R_2 + R_1)(R_2 - R_1)\tan\frac{\eta}{2} \\
&= h^2 \left(\frac{1}{\tan^2\left(\phi - \frac{\eta}{2}\right)} - \frac{1}{\tan^2\left(\phi + \frac{\eta}{2}\right)} \right)\tan\frac{\eta}{2} \\
&= 4h^2 \tan\phi \tan^2\frac{\eta}{2} \left[\frac{(1 + \tan^2\phi)\left(1 + \tan^2\frac{\eta}{2}\right)}{\left(\tan^2\phi - \tan^2\frac{\eta}{2}\right)^2} \right] \\
&= 4h^2 \frac{h}{R}\tan^2\frac{\eta}{2} \left[\frac{\left(1 + \frac{h^2}{R^2}\right)\left(1 + \tan^2\frac{\eta}{2}\right)}{\left(\frac{h^2}{R^2} - \tan^2\frac{\eta}{2}\right)^2} \right] \\
&= 4h^2 \frac{(1 - \sin\delta\phi)}{\cos\delta\phi}\tan^2\frac{\eta}{2} \left[\frac{\left(1 + \frac{(1 - \sin\delta\phi)^2}{\cos^2\delta\phi}\right)\left(1 + \tan^2\frac{\eta}{2}\right)}{\left(\frac{(1 - \sin\delta\phi)^2}{\cos^2\delta\phi} - \tan^2\frac{\eta}{2}\right)^2} \right] \\
&\triangleq h^2\mathcal{A},
\end{aligned}
$$

where we have used the relation $\tan\phi = \frac{h}{R}$ and Eq. 18. If P is the number of pixels on the camera, then the average pixel density is given by

$$
\mu = \frac{P}{h^2\mathcal{A}}. \tag{19}
$$

Figure 12 Assuming a *flat earth model*, the area imaged by the camera can be computed by knowledge of the roll angle ϕ the lens field-of-view η, and the altitude h.

$d_2 = R_2\tan\frac{\eta}{2}$

$d_1 = R_1\tan\frac{\eta}{2}$

$R_1 = \frac{h}{\tan(\phi + \frac{\eta}{2})}$ R $R_2 = \frac{h}{\tan(\phi - \frac{\eta}{2})}$

Suppose that the computer vision algorithm requires a desired average pixel density of μ^d, then using Eqs. 19 and 18 we get that the optimal altitude and orbit radius are given by

$$h^* = \sqrt{\frac{P}{\mu^d \mathcal{A}}}$$

$$R^* = h^* \left(\frac{\cos \delta\phi}{1 - \sin \delta\phi} \right).$$

3.4 Wind Estimation

For MAVs, winds that are a significant percentage of the airspeed are almost always present. Therefore, the airframe is typically required to "crab" into the wind, causing the course (direction of travel) to deviate from the heading (direction of the body frame x-axis). Since the camera is mounted to the body, the difference between course and heading, if it is not accounted for, will cause significant errors in the geo-location estimates. In this section, the heading angle is denoted by ψ and the course angle will be denoted by χ.

To illustrate the effect of wind, Figure 13 shows the error in the geo-location estimate generated by a simulation of a complete orbit in significant wind. The simulated MAV has an airspeed of 18 m/s and the wind is from the East at 9 m/s. Note that since the MAV must crab right into the wind, the geo-location errors shown in Figure 13 are significantly biased to the South. We note that wind does not introduce a constant bias in the estimate and can therefore not be removed by the techniques discussed in Section 3.2. To compensate for wind, the direction and magnitude of the wind is estimated on-line from flight data and is used to modify ψ in Eq. 13. We will

Figure 13 Effect of wind on geo-location estimates.
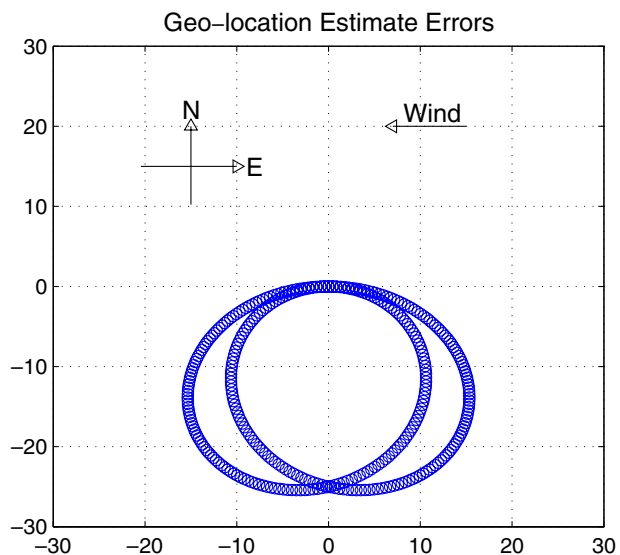
assume that GPS measurements are available but that the MAV is not equipped with magnetometers.

The relationship between windspeed, groundspeed, and airspeed is illustrated in Figure 14, where V_w is the windspeed, V_g is the groundspeed, V_a is the airspeed, and ξ is the wind direction, and can be expressed as

$$V_g = V_a \cos(\psi - \chi) + V_w \cos(\xi - \chi). \tag{20}$$

The GPS sensor measures V_g and χ, and a differential pressure sensor can be used to measure V_a.

From the law of cosines and Figure 14, we have

$$V_g^2 - V_a^2 + V_w^2 - 2V_g V_w \cos(\xi - \chi) = 0. \tag{21}$$

To estimate V_w and ξ we collect on-line measurements of V_g, V_a, and χ and use a quasi-Newton nonlinear equation solver to minimize the objective function

$$\sum_{i=0}^{n} \left(V_{g_i}^2 - V_{a_i}^2 + V_w^2 - 2V_{g_i} V_w \cos(\xi - \chi_i) \right)^2, \tag{22}$$

where the index i denotes a measurement sample.

To quantify the effectiveness of our wind estimation scheme, we flew a MAV in windy conditions in an orbit pattern. Since we do not have the instrumentation to measure true wind speed at the elevations that the MAV is flying (100–200 m), to measure the accuracy of our wind estimation method, we used the estimated windspeed and the measured airspeed to estimate ground speed, and compared the estimate with the measured GPS ground speed. Figure 15 shows actual flight

Figure 14 Relationship between ground, air, and wind velocities.
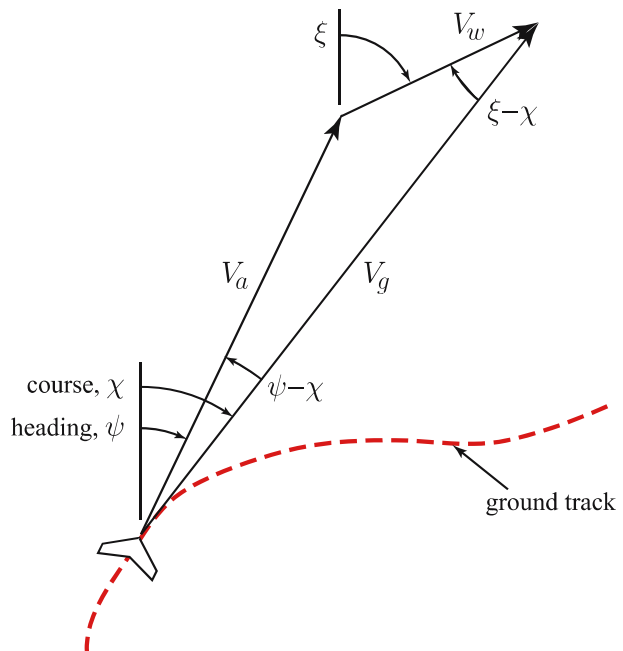
Figure 15 Wind solution for a dataset taken in high-wind conditions $V_w \approx 9$ m/s.
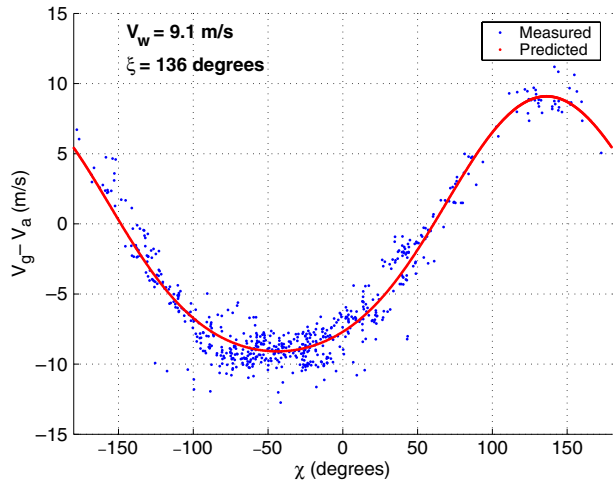
data recorded while the MAV was flying in winds of approximately 9 m/s from the north. Figure 15 shows the efficacy of our method by plotting the raw estimates of ground speed taken from GPS measurements (the scattered points), together with the ground speed predicted for a constant air speed and the estimated wind speed (the solid curve). Results demonstrating the efficacy of the wind correction scheme for geo-location are discussed in Section 4.2.

The wind estimation scheme discussed in this section estimates a constant wind and does not account for gusts. On the other hand, flight test data suggest that the gusts are essentially normally distributed about the constant wind and are therefore removed by the RLS filter.

4 Results

4.1 Hardware Testbed

BYU has developed a reliable and robust platform for testing unmanned air vehicles [1]. Figure 16 shows the key elements of the testbed. The first frame shows the Procerus[1] Kestrel autopilot (originally developed at BYU) which is equipped with a Rabbit 3400 29 MHz processor, rate gyros, accelerometers, absolute and differential pressure sensors. The autopilot measures $3.8 \times 5.1 \times 1.9$ cm and weighs 17 g.

The second frame in Figure 16 shows the airframe used for the flight tests reported in this paper. The airframe is a flying wing with expanded payload bay and servo-driven elevons designed at the BYU Magicc Lab. It has a wingspan of 152 cm, a length of 58 cm, and a width of 12 cm. It weighs 1.1 kg unloaded and 2.0 kg fully loaded. It is propelled by a brushless electric motor which uses an electronic speed control and is fueled by four multi-cell lithium polymer batteries. Typical speeds for the aircraft are between 15 and 20 m/s (33 and 45 miles/h). Maximum flight time for

---

[1]http://procerusuav.com/.

Figure 16 a Procerus' Kestrel autopilot. **b** MAV airframe. **c** Ground station components.

this aircraft is between 1 and 2 h depending on external conditions and the mission it is required to fly.

The third frame in Figure 16 shows the ground station components. A laptop runs the Virtual Cockpit software that interfaces through a communication box to the MAV. An RC transmitter is used as a stand-by fail-safe mechanism to ensure safe operations.

The gimbal and camera used for this paper are shown in Figure 17. The gimbal was designed and constructed at the BYU Magicc Lab. It weighs 150 g, and has a range of motion of 135° in azimuth (at 333°/s) and 120° in elevation (at 660°/s). The camera is a Panasonic KX-141 with 480 lines of resolution. The field of view of the lens is 60°.

4.2 Geo-location Accuracy

Using the MAV system described above, in conjunction with the geo-location techniques described in this paper, we have repeatedly (15–20 experiments in a variety of weather conditions) geo-located well defined visual objects, with errors ranging between 2 and 4 m. (The true value of the target is measured using the

Figure 17 The gimbal and camera used for the results in the paper are shown unmounted from the MAV and without its protective dome.

Figure 18 Localization results
in high-wind conditions.
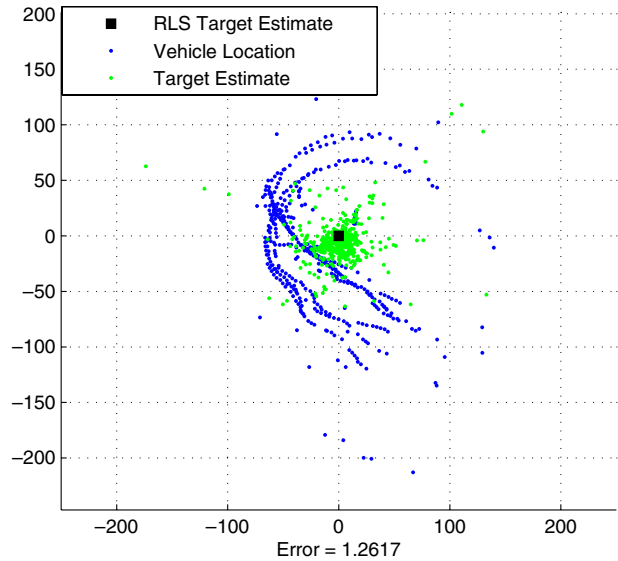
same commercial grade GPS receiver used on the MAV. Note that the geo-location techniques discussed in this paper do not remove GPS bias. A military grade GPS, or differential GPS would remove this bias.) The results of two particular flight tests are shown in Figures 8 and 18. The outer blue dots represent the GPS location of the MAV, while the inner green dots are the raw geo-location estimates. All location values are in reference to the true location of the target (as measured by GPS). The flight tests shown in Figure 8 were performed on a day with relatively little wind, while the flight tests shown in Figure 18 were performed in extremely high-wind conditions (>10 m/s). Note that the high-wind conditions cause the irregular flight pattern shown in Figure 18. In both Figures, the accuracy of the raw geo-location

Figure 19 Efficacy of RLS
algorithm.
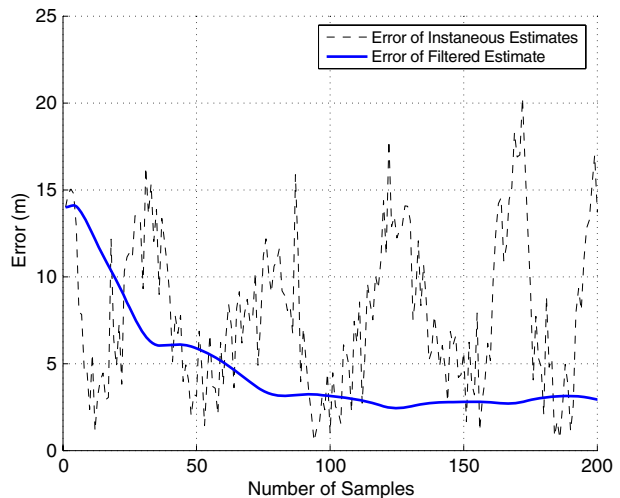
estimates is typically less than 20 m, although there are some outliers in the high-wind case. The black dot in the center of the figures represents the final geo-location estimate, and is approximately 3 m away from the target in the low-wind case, and 2 m away in the high-wind case.

In Figure 19, we show the effects of using the RLS system to derive the final geo-location estimate. In this plot, the x-axis denotes different raw estimates of geo-location (typically estimated about three times per second), while the y-axis denotes the magnitude of the localization error. The data in this graph corresponds with the low-wind experiment plotted in Figure 8. As illustrated in Figure 19, the raw estimates can be up to 20 m in error. However, the RLS filtered estimate quickly converges to less than 5 m of error.

5 Conclusions

This paper introduces a system for vision-based target geo-localization from a fixed-wing micro air vehicle. The geometry required to produce raw localization estimates is discussed in detail. The primary contribution of the paper is the description of four key techniques for mitigating the error in the raw estimates. These techniques include RLS filtering, bias estimation, flight path selection, and wind estimation. The algorithms were successfully flight tested on a micro air vehicle using Procerus' Kestrel autopilot and a BYU designed gimbal system. Geo-location errors below 5 m were repeatedly obtained under a variety of weather conditions. Throughout the paper we have assumed a flat earth model and a stationary target. Future research will include generalizing the techniques to non-flat terrain and to moving ground targets.

References

1. Beard, R., Kingston, D., Quigley, M., Snyder, D., Christiansen, R., Johnson, W., McLain, T., Goodrich, M.: Autonomous vehicle technologies for small fixed wing UAVs. AIAA J. Aero. Comput. Inform. Comm. **2**(1), 92–108 (2005)
2. Campbell, M.E., Wheeler, M.: A vision based geolocation tracking system for UAVs. In: Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, CO (2006) (Paper no. AIAA-2006-6246)
3. Chaimowicz, L., Grocholsky, B., Keller, J.F., Kumar, V., Taylor, C.J.: Experiments in multirobot air-ground coordination. In: Proceedings of the 2004 International Conference on Robotics and Automation, pp. 4053–4058. New Orleans, LA (2004)
4. Chroust, S.G., Vincze, M.: Fusion of vision and inertial data for motion and structure estimation. J. Robot. Syst. **21**, 73–83 (2003)
5. Dobrokhodov, V.N., Kaminer, I.I., Jones, K.D.: Vision-based tracking and motion estimation for moving targets using small UAVs. In: Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, CO (2006) (Paper no. AIAA-2006-6606)
6. Eldredge, A.M.: Improved state estimation for miniature air vehicles. Master's thesis, Brigham Young University (2006)
7. Frew, E., Rock, S.: Trajectory generation for monocular-vision based tracking of a constant-velocity target. In: Proceedings of the 2003 IEEE International Conference on Robotics and Automation, Taipei, Taiwan (2003)

8. Gibbins, D., Roberts, P., Swierkowski, L.: A video geo-location and image enhancement tool for small unmanned air vehicles (UAVs). In: Intelligent Sensors, Sensor Networks and Information Processing Conference, Proceedings of the 2004, pp. 469–473. Melbourne, Australia (2004)
9. Kumar, R., Samarasekera, S., Hsu, S., Hanna, K.: Registration of highly-oblique and zoomed in aerial video to reference imagery. In: Proceedings of the IEEE Computer Society Computer Vision and Pattern Recognition Conference, Barcelona, Spain (2000)
10. Kumar, R., Sawhney, H., Asmuth, J., Pope, A., Hsu, S.: Registration of video to geo-referenced imagery. In: Proceedings of the 14th International Conference on Pattern Recognition, vol. 2, pp. 1393–1400. Brisbane, Australia (1998)
11. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S.: An Invitation to 3-D Vision From Images to Geometric Models. Springer, Berlin Heidelberg New York (2003)
12. Moon, T.K. Stirling, W.C.: Mathematical Methods and Algorithms. Prentice-Hall, Englewood Cliffs, NJ (2000)
13. Office of the Secretary of Defense (ed.): Unmanned Aerial Vehicles Roadmap, pp. 2002–2027. United States Government, Washington DC, USA (2002)
14. Redding, J., McLain, T.W., Beard, R.W., Taylor, C.: Vision-based target localization from a fixed-wing miniature air vehicle. In: Proceedings of the American Control Conference, pp. 2862–2867. Minneapolis, MN (2006)
15. Redding, J.D.: Vision based target localization from a small fixed-wing unmanned air vehicle. Master's thesis, Brigham Young University (2005)
16. Rysdyk, R.: UAV path following for constant line-of-sight. In: 2nd AIAA Unmanned Unlimited Systems, Technologies and Operations Aerospace, Land and Sea Conference, San Diego, CA (2003)
17. Saeedi, P., Lowe, D.G., Lawrence, P.D.: 3D localization and tracking in unknown environments. In: Proceedings of the IEEE Conference on Robotics and Automation, vol. 1, pp. 1297–1303. Taipei, Taiwan (2003)
18. Schultz, H., Hanson, A., Riseman, E., Stolle, F., Zhu, Z.: A system for real-time generation of geo-referenced terrain models. SPIE Enabling Technologies for Law Enforcement, Boston, MA, USA (2000)
19. Stolle, S., Rysdyk, R.: Flight path following guidance for unmanned air vehicles with pan-tilt camera for target observation. In: 22nd Digital Avionics Systems Conference, Indianapolis, IN (2003)
20. Trucco, E., Verri, A.: Introduzctory Techniques for 3-D Computer Vision. Prentice-Hall, NJ, USA (2002)
21. Vidal, R., Sastry, S.: Vision-based detection of autonomous vehicles for pursuit-evasion games. In: IFAC World Congress on Automatic Control, Barcelona, Spain (2002)
22. Wang, I., Dobrokhodov, V., Kaminer, I., Jones, K.: On vision-based target tracking and range estimation for small UAVs. In: 2005 AIAA Guidance, Navigation, and Control Conference and Exhibit, pp. 1–11. San Francisco, CA (2005)
23. Whang, I.H., Dobrokhodov, V.N., Kaminer, I.I., Jones, K.D.: On vision-based tracking and range estimation for small UAVs. In: Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, San Francisco, CA (2005) (Paper no. AIAA-2005-6401)