Enabling Navigation of MAVs through Inertial, Vision, and Air Pressure Sensor Fusion

Clark N. Taylor

Abstract Traditional methods used for navigating miniature unmanned aerial vehicles (MAVs) consist of fusion between Global Positioning System (GPS) and Inertial Measurement Unit (IMU) information. However, many of the flight scenarios envisioned for MAVs (in urban terrain, indoors, in hostile (jammed) environments, etc.) are not conducive to utilizing GPS. Navigation in GPS-denied areas can be performed using an IMU only. However, the size, weight, and power constraints of MAVs severely limits the quality of IMUs that can be placed on-board the MAVs, making IMU-only navigation extremely inaccurate. In this paper, we introduce a Kalman filter based system for fusing information from two additional sensors (an electro-optical camera and differential air pressure sensor) with the IMU to improve the navigation abilities of the MAV. We discuss some important implementation issues that must be addressed when fusing information from these sensors together. Results demonstrate an improvement of at least 10x in final position and attitude accuracy using the system proposed in this paper.

Keywords Vision-aided navigation · GPS-denied narrigation · Sensor fusion

1 Introduction

Recently, Unmanned Aerial Vehicles (UAVs) have seen a dramatic increase in utilization for military applications. In addition, UAVs are being investigated for multiple civilian uses, including rural search and rescue, forest fire monitoring, and agricultural information gathering. Due to their small size, Miniature UAVs (MAVs) are an attractive platform for executing many of the missions traditionally performed by larger UAVs. Some of the primary advantages of MAVs include: (1) they are significantly less expensive to purchase than the large UAVs typically used by the

C.N. Taylor (✉)
459 CB, Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602
e-mail: clark.taylor@byu.edu

S. Lee et al. (eds.), *Multisensor Fusion and Integration for Intelligent Systems*, Lecture Notes in Electrical Engineering 35, DOI 10.1007/978-3-540-89859-7_11,

military, (2) their small size simplifies transport, launch, and retrieval, and (3) they are less expensive to operate than large UAVs.

To enable the utilization of MAVs, the ability to accurately navigate (estimate the location, attitude, and velocity of the MAV) is essential. For example, if an MAV is being utilized as part of a search and rescue operation, knowing the correct location and attitude of the MAV is critical to geo-locate an observed object of interest.

Navigation methods implemented in MAVs today are primarily based on the fusion of measurements from the Global Positioning System (GPS) and an inertial measurement unit (IMU) (e.g. [1–4]). However, there are many scenarios in which an MAV might prove useful but GPS is not available (e.g., indoors, urban terrain, etc.). Therefore, a number of methods have been proposed for fusing visual information with IMU measurements to enable navigation without GPS.

For the most accurate navigation possible, a batch method that analyzes all vision and IMU information from an entire flight was introduced in [5]. While accurate, this method cannot be utilized in real-time due to nature of batch optimization routines. Therefore, the paper also introduces a recursive method which is essentially a SLAM filter. Other implementations of SLAM-based filters for navigation can also be found in [6–8]. While SLAM-based methods are highly effective, there are two bottlenecks to SLAM that make them difficult to implement in the computationally-limited environments that characterize MAVs. First, visual SLAM requires that objects in the video be tracked for an extended period of time. Second, the size of the state grows with the number of landmarks that SLAM is attempting to find the location for, dramatically increasing the computation time required. In [9], the size of the state is limited, but tracking of features over a long period of time is still required.

While it is computationally expensive to track points for an extended period of time in video, it is relatively simple to track points over a small number of video frames. Therefore, we focus in this paper on a method that utilizes only the relationship between objects in two frames of video. Other methods that utilize only two frames of video ([10–12]) have been introduced previously. Ready and Taylor [11] and Andersen and Taylor [12], however, requires that the terrain being observed is planar, while we assume in this paper that the points being tracked from frame to frame are not planar (a better assumption for indoor or dense environments that would obscure GPS signals).

In this paper, we focus on utilizing the *epipolar constraint* for fusing visual measurements with the IMU as described in [10]. Using the epipolar constraint, however, has three significant weaknesses that must be addressed when performing fusion with IMU measurements for MAV navigation. First, the epipolar constraint biases movements of the camera toward the center of points in the image. Second, visual measurements always include a "scale ambiguity" – it is impossible to distinguish between the camera moving quickly and observing an object that is far away and the camera moving slowly and observing an object that is close. Third, when

using vision to navigate, the navigation state of the camera can only be determined relative to its previous navigation states.[1]

In this paper, we present methods to overcome each of these three weaknesses in utilizing the epipolar constraint. First, to overcome the bias in the epipolar constraint, we analyze the epipolar constraint equation and propose an alternate algorithm for computing deviations from the epipolar constraint. Second, to overcome the scale ambiguity of vision, we integrate a differential air pressure sensor into the fusion system. On a fixed-wing MAV, the differential air pressure sensor is capable of measuring the airspeed of the MAV. This airspeed can be treated as a direct measurement of the velocity magnitude, allowing the scale ambiguity of vision to be overcome. Third, because vision measurements of motion are relative, we propose using the *minimal* sampling rate at which vision can be effectively fused with IMU data. We demonstrate that sampling at the minimal, rather than maximal, rate increases the accuracy of the overall navigation system. We also discuss limitations on choosing the minimal sampling rate.

Once the weaknesses of epipolar-based fusion are overcome, it is possible to enable on-line estimation of inertial sensor biases. We prove this capability by performing an observability analysis of a simplified system, and demonstrate improved navigation results when estimating biases.

The remainder of this paper is organized as follows. In Sect. 2, we discuss the general framework used for fusing vision and IMU information. In Sect. 3, we describe our three modifications for improving the fusion of visual and IMU information. In Sect. 4, we describe our method for estimating the biases of the inertial sensors. Section 5 demonstrates the improvement in navigation state estimates achieved by implementing our modifications. Section 6 concludes the paper.

2 Epipolar Constraint-Based Fusion

In this section, we first describe the epipolar constraint which is used in our fusion setup. We then describe how the epipolar constraint can be used in a Kalman filter to enable fusion between visual and inertial information.

2.1 The Epipolar Constraint

The epipolar constraint can be utilized whenever a single fixed object is observed by a camera at two locations (or two cameras at different locations). Given that any three points in the world form a plane, a single world point and the two camera projection centers form a plane in the 3-D world – the *epipolar* plane. Similarly, when a world point is observed in two images, the two vectors representing where the point was observed in the image plane ($\mathbf{x}'$ and $\mathbf{x}$), and the translation vector

[1] Note that while it is possible to determine absolute position or attitude using vision, knowledge of pre-existing visual "landmarks" is required. We do not address these techniques in this paper as we are interested in using MAVs to explore new areas, not fly over pre-mapped areas.

between the two camera locations, will all lie on the epipolar plane.[2] By assigning a unique coordinate frame to each camera location, this constraint is represented by the formula

$$\mathbf{x}'[\mathbf{p}_{c1}^{c2}]_{\times}\mathbf{C}_{c1}^{c2}\mathbf{x} = 0, \tag{1}$$

where $\mathbf{C}_{c1}^{c2}$ is the direction cosine matrix between the camera coordinate frames and $[\mathbf{p}_{c1}^{c2}]_{\times}$ is the position of camera 1 in the second camera's coordinate frame, put into a skew-symmetric matrix. (In other words, we calculate the cross product between $\mathbf{p}_{c1}^{c2}$ and $\mathbf{C}_{c1}^{c2}\mathbf{x}$.) This constraint enforces that the three vectors $\mathbf{x}'$, $\mathbf{C}_{c1}^{c2}\mathbf{x}$, and $\mathbf{p}_{c1}^{c2}$ all lie within the same plane in the 3-D world.

2.2 Utilizing the Epipolar Constraint in a Fusion Environment

To utilize the epipolar constraint in a fusion environment, we created an Unscented Kalman Filter (UKF) framework. The state in the Kalman filter must contain enough information to generate both $\mathbf{C}_{c1}^{c2}$ and $\mathbf{p}_{c1}^{c2}$ during the measurement step. To represent general motion by an MAV between two different locations, the UKF state $\mathbf{X} = \begin{bmatrix} \chi_t \\ \chi_{t-1} \end{bmatrix}$ is used, where χ_t is the navigation state estimate at time t and χ_{t-1} is the navigation state estimate at the previous time. The navigation state at each time contains p_t, the position of the camera at time t in the inertial frame, $\mathbf{C}_t$, the direction cosine matrix relating the inertial frame to the current camera coordinate frame, and v_t, the velocity of the camera at time t. This method for setting up the UKF to enable vision and inertial information fusion was first introduced in [12].

2.2.1 Performing the Time Update

The time update for this UKF implementation takes two different forms. The first form updates χ_t (the first 10 elements of the current state) every time an IMU measurement occurs. This makes the first 10 elements of the state the most recent navigation state estimate. After each measurement update, the state is also updated using the formula

$$\mathbf{X}^{+} = \mathbf{A}\mathbf{X}^{-} \text{ where} \tag{2}$$

$$\mathbf{A} = \begin{bmatrix} I & 0 \\ I & 0 \end{bmatrix}, \tag{3}$$

[2] Note that image locations are typically in pixels, while the discussion of vectors so far assumes all vectors are in an unscaled, Euclidean space. In this paper, we assume that the camera has been calibrated a-priori and that the conversion from image to Euclidean vectors has already occurred using this calibration information.

causing the current state to become $\mathbf{X} = \begin{bmatrix} \chi_t \\ \chi_t \end{bmatrix}$. The first 10 elements of the state vector are then updated by IMU measurements to realize a new current state $\mathbf{X} = \begin{bmatrix} \chi_{t+1} \\ \chi_t \end{bmatrix}$. With this technique, the two most recent χ estimates, corresponding to the time of the two most recently captured images, are always stored as the current state.

2.2.2 Performing the Measurement Update

To utilize the epipolar constraint as a measurement in a UKF framework, the "Dynamic Vision" approach introduced in [10] is used. Assuming a feature has been detected in two images, the locations of the features are represented by $\mathbf{x}'$ and $\mathbf{x}$. Because the epipolar constraint should always be equal to zero, the "measurement" used by the UKF is a vector of zeros in length equal to the number of corresponding features found between the two images. The predicted measurement is $\mathbf{x}'[\mathbf{p}_{c1}^{c2}]_\times \mathbf{C}_{c1}^{c2}\mathbf{x}$ (from Eq. (1)) for each set of features $\mathbf{x}'$ and $\mathbf{x}$, where $\mathbf{p}_{c1}^{c2}$ and $\mathbf{C}_{c1}^{c2}$ are functions of χ_t and χ_{t-1}.

While the Dynamic Vision method yields good results in certain cases, it does exhibit some weaknesses that need to be addressed for use on an MAV. First, the translation direction estimates are biased in the direction the camera is pointing. Second, as with all vision-based approaches, it does not estimate the magnitude of translation. We propose methods for overcoming these weaknesses in the following section.

3 Improving the Fusion of Visual and IMU Sensors

In this section, we propose three modifications to baseline epipolar constraint-based fusion of IMU and visual information to significantly increase the accuracy of navigation state estimation on MAVs. These modifications overcome the centering bias, scale ambiguity, and sampling rate problems discussed in the introduction.

3.1 Overcoming Bias Toward the Center of Image Points

When using the epipolar constraint to fuse inertial and vision sensors together, the fusion system introduces a bias in estimated translation direction toward the center of the points observed from frame to frame. To understand the source of this error, let us analyze the epipolar constraint when the result of Eq. (1) is not zero. The value $x'[\mathbf{p}_{c1}^{c2}]_\times \mathbf{C}_{c1}^{c2}x$ can be rewritten as a cross product of two vectors followed by a dot product of two vectors. The final results of this computation will be

$$||x'||\ ||\mathbf{p}_{c1}^{c2}||\ ||x||\sin(\theta_{\mathbf{p}_{c1}^{c2}\to\mathbf{C}_{c1}^{c2}x})\cos(\theta_{x'\to[\mathbf{p}_{c1}^{c2}]_\times\mathbf{C}_{c1}^{c2}x}), \tag{4}$$

where $\theta_{\mathbf{p}_{c1}^{c2} \rightarrow \mathbf{C}_{c1}^{c2} x}$ is the angle between $\mathbf{p}_{c1}^{c2}$ and $\mathbf{C}_{c1}^{c2} x$ and $\theta_{x' \rightarrow [\mathbf{p}_{c1}^{c2}]_\times \mathbf{C}_{c1}^{c2} x}$ is the angle between x' and $[\mathbf{p}_{c1}^{c2}]_\times \mathbf{C}_{c1}^{c2} x$.

As mentioned previously, the correct magnitude for these values is 0. Therefore, the UKF will attempt to set χ_t and χ_{t-1} in its state vector such that the resulting $\mathbf{p}_{c1}^{c2}$ and $\mathbf{C}_{c1}^{c2}$ minimizes the set of all measurements. However, there are two ways to push the set of measurements toward zero: (1) the epipolar constraint can be met by setting $[\mathbf{p}_{c1}^{c2}]_\times \mathbf{C}_{c1}^{c2} x$ to be orthogonal to x' (i.e., set $\cos(\theta_{x' \rightarrow [\mathbf{p}_{c1}^{c2}]_\times \mathbf{C}_{c1}^{c2} x}) = 0$), or (2) $\mathbf{p}_{c1}^{c2}$ can be set parallel to $\mathbf{C}_{c1}^{c2} x$ (i.e., set $\sin(\theta_{\mathbf{p}_{c1}^{c2} \rightarrow \mathbf{C}_{c1}^{c2} x}) = 0$). To meet the first condition, $\mathbf{C}_{c1}^{c2} x$, p_{c1}^{c2}, and x' must all lie in a plane, the original justification behind the epipolar constraint. The second condition, however, can be met by setting the direction of p_{c1}^{c2} equal to x'. Because of this second condition, the results of the epipolar constraint can be pushed to zero by setting $\mathbf{p}_{c1}^{c2}$ to be as close to parallel to the set of $\mathbf{C}_{c1}^{c2} x$ vectors as possible. Therefore, the translation direction after the UKF measurement update is biased toward the center of the feature points that have been tracked in the second image.

To overcome this biasing of the translation direction, we propose modifying the measurement step of the UKF to eliminate the $\sin(\theta_{\mathbf{p}_{c1}^{c2} \rightarrow \mathbf{C}_{c1}^{c2} x})$ term from the measurement. To eliminate the effect of $\sin(\theta_{\mathbf{p}_{c1}^{c2} \rightarrow \mathbf{C}_{c1}^{c2} x})$, the term $[\mathbf{p}_{c1}^{c2}]_\times \mathbf{C}_{c1}^{c2} x$ is first computed and then normalized to be of length one. The inner product of this term with x' is then taken and returned as the predicted measurement.

To determine the results of this modification, we simulated a 16 s flight of an MAV traveling 200 m in a straight line. (More details on our simulation environment can be found in Sect. 5.) In Table 1, we show the average and standard deviation of the error in the final estimated position of the MAV, both before and after the sin removal modification. Note that before sin removal, the p_z location error mean is a large positive number. This is a result of the bias inherent in the unmodified epipolar fusion environment. After removing the sin as discussed above, the z error is dramatically decreased.

Despite the fact that the p_z error has been decreased by removing the sin term from the epipolar constraint, the p_x error is still quite significant. This error is an artifact of vision-based techniques where there is always a scale ambiguity in the direction of travel (in this case, along the x axis). In the next subsection, we discuss how to reduce the error present in the direction of travel of the MAV due to the scale ambiguity.

Table 1 This table demonstrates the accuracy improvements achieved by removing the sin term from the epipolar constraint computation. Each entry lists the mean and standard deviation of the error, in meters. Note that the average p_z error has decreased from 758 to 12 m, demonstrating the effect of removing the bias from the epipolar computation

	p_x error	p_y error	p_z error
With sin (μ, σ)	$(-607.0, 136.6)$	$(2.83, 76.5)$	$(757.5, 318.1)$
sin removed (μ, σ)	$(23.1, 85.9)$	$(-1.77, 16.0)$	$(11.8, 9.71)$

3.2 Removing the Scale Ambiguity

To remove the large amount of error present in the direction of travel of the MAV, we propose integrating another sensor into the UKF framework discussed above. On a fixed-wing MAV, a pitot tube designed for measuring airspeed can be utilized to measure the current velocity of the MAV. To integrate this measurement in the UKF, we utilize the property discussed in [13] that if measurements are uncorrelated, they can be applied during separate measurement updates of the Kalman Filter. Therefore, whenever the pitot tube is read (approximately 10 Hz), the current magnitude of the velocity in the state is computed as a predicted measurement, with the air speed measured by the pitot tube used as a measurement to the UKF.

By applying this measurement at 10 Hz, significant gains in accuracy were achieved. In Table 2, we present the results of this modification using the same simulation setup as described for Table 1. Note that both the mean and standard deviation of error has decreased in all three location parameters, demonstrating the importance of overcoming the scale ambiguity in visual measurement.

Table 2 This table demonstrates the accuracy improvements achieved by fusing the pitot tube measurements with vision and IMU information. Note that the standard deviation of error on the p_x term has decreased from 86 to 2

	p_x error	p_y error	p_z error
Without pitot tube (μ, σ)	(23.1, 85.9)	$(-1.77, 16.0)$	(11.8, 9.71)
With pitot tube (μ, σ)	$(-0.47, 1.92)$	$(-1.01, 11.2)$	(2.85, 1.08)

3.3 Determining the Optimal Image Sampling Rate

Typically, when fusing information together, the more information that is available, the more accurate the final result will be. However, with epipolar based visual and inertial fusion, this is not the case. In this subsection, we show that it best to sample the imaging sensors at the *minimal* sampling rate allowed. We also discuss what limits the minimal possible sampling rate.

In Figure 1, we show plots of the mean squared error (MSE) in the estimated final location of the MAV for different image sampling rates. The MSE represents the error in estimated location after a 15 s, straight-line flight. Note that the lowest MSE point does *not* lie at the maximal sampling rate. This can be explained by noting that fusing with the epipolar constraint helps to reduce the amount of error present between two navigation state estimates (i.e., relative error). The total error at the end of flight is going to be the summation of all the relative estimation errors during the flight. Therefore, if the same relative error is achieved by each measurement of the epipolar constraint, but fewer measurements occur, the total error will be reduced. This leads to the counter-intuitive fact that the minimal, as opposed to maximal sampling rate, is ideal for epipolar constraint-based information fusion.

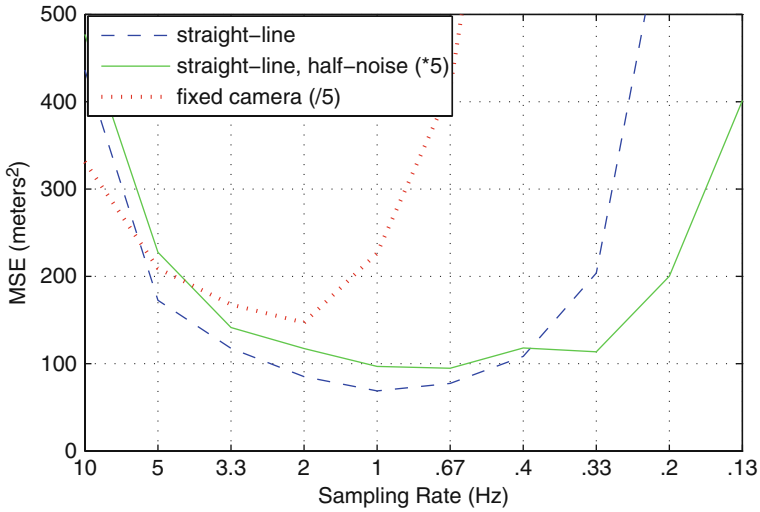

Fig. 1 Results of fusion for different image sampling rates (x-axis) and different fusion setups (different plots). Results are mean squared error (in m^2). To make all plots appear on the same axes, two of the plots have been scaled by 5 (up and down) as denoted in the legend

Despite the general rule that the minimal sampling rate is ideal, there are secondary considerations that must be taken into account when deciding on a sampling rate. Note that in Figure 1, the MSE is *not* monotonically decreasing as the sampling rate decreases. There are two principal causes for the increase in MSE at lower sampling rates.

First, the time update of the IMU introduces error into the estimated navigation states, which the UKF attempts to correct using the epipolar constraint. This correction applied by the UKF is a linear correction. As the distance between the estimated and measured navigation states increase however, the linear assumption becomes invalid. Therefore, if too much noise has been added by the IMU, it will not be possible for the linear update from the epipolar constraint to correct the IMU-introduced noise. This is demonstrated by the "straight-line" and "straight-line, half-noise" plots in Figure 1. The only difference between the simulation setup of the different plots is that the IMU noise was halved for the "straight-line, half noise" plot. Note that as the IMU noise is decreased, the "optimal" sampling rate becomes lower (moving from 1 to .67 Hz), demonstrating that there is a limit placed on the minimal sampling rate by the noise present in the IMU.

Second, the minimal sampling rate is limited by how long the camera can track features in the image. To demonstrate this fact, a simulation was run where, rather than tracking a set of objects throughout the entire flight, as in the "straight-line" simulations of Figure 1, a fixed camera was used so that objects would leave the field of view more quickly. This is the "straight-line, fixed-camera" plot in Figure 1. Note that the sampling rate with the minimal MSE is higher (2 Hz) for this plot than the "straight-line" plot (1 Hz) where the same world points are observed throughout the MAV flight. Therefore, when applying epipolar constraint-based fusion, it is best

to apply the minimal sampling rate that is allowed by (1) the noise present in the IMU and (2) the persistence of features across the images.

4 Estimating Inertial Sensor Biases

Using the modifications proposed in the prior section, it is possible to *overcome* a significant amount of error introduced by inertial sensors when navigating an MAV. In this section, we discuss how fusion based on the epipolar constraint can be used to estimate biases of the inertial sensors, thereby *reducing* the noise from those sensors.

Typical low-cost inertial sensors will have several different *types* of noise [14,15]. The Kalman Filter setup described in Sect. 2 essentially assumes that all noise from the inertial sensors is uncorrelated over time. While the existence of correlated noise (*bias* for the remainder of this paper) in the inertial sensors can be overcome by simply increasing the uncorrelated noise covariance estimates, it is preferable to estimate and remove the bias, thereby reducing the amount of noise present in the inertial sensor measurements.

Before describing our method for estimating the bias of the inertial sensors, it is important to demonstrate the feasibility of estimating biases despite the fact that epipolar constraint measurements yield only relative navigation information. In this section, we first demonstrate that it is possible to estimate biases using relative navigation measurements. We then describe our modifications to the Kalman filter framework described in Sect. 2 to enable estimation of the inertial sensor biases.

4.1 Proof of Ability to Estimate Biases

To prove the feasibility of estimating inertial sensor biases, we will perform an observability analysis of a simplified system with a setup that is very similar to our complete MAV navigation system. Our simplified system consists of a state vector with two locations, l_t and l_{t-1}. Similar to the situation where the accelerometers are used to update the current velocity estimates, we use an external rate measurement ($\hat{\dot{l}}_t$) to update the current locations. To estimate the bias on this measurement, we modify the state vector to include the bias, obtaining a state vector of:

$$\mathbf{x}_t = \begin{bmatrix} l_t & l_{t-1} & b \end{bmatrix}^T, \tag{5}$$

where b is the bias of the sensor.

The time update for this state over time Δt is

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{G}\hat{\dot{l}}_t \tag{6}$$

where

$$\mathbf{G} = [\Delta t \quad 0 \quad 0] \tag{7}$$

and

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & -\Delta t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{8}$$

If the measurement of the system provides relative measurements (like vision does for MAV motion), then the observation matrix is:

$$\mathbf{H} = [1 \quad -1 \quad 0]. \tag{9}$$

With this simplified system, we can analyze the observability of b for this Kalman Filter setup. To prove observability, the rank and null vectors of the matrix

$$\mathcal{O} = \begin{bmatrix} \mathbf{H} \\ \mathbf{HF} \\ \mathbf{HFF} \\ \cdots \end{bmatrix} \tag{10}$$

must be found. Substituting in for $\mathbf{H}$ and $\mathbf{F}$ in the first three rows of $\mathcal{O}$, we obtain:

$$\mathcal{O} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & -\Delta t \\ 0 & 0 & -\Delta t \end{bmatrix} \tag{11}$$

By inspection, we find that this system can observe two modes of the system, namely $[1 \quad -1 \quad 0]$ and $[0 \quad 0 \quad 1]$. Because of the second mode, we conclude that biases are observable when relative measurements of the state are used in a Kalman Filter framework.

4.2 Filter Setup for Inertial Bias Estimation

Knowing that biases are observable when relative navigation measurements are used, we can modify the fusion framework developed in Sects. 2 and 3 to estimate biases. First, we modify the state vector to include biases for all sensors, yielding

$$\mathbf{X} = \left[\chi_t, \chi_{t-1}, b_{ax}, b_{ay}, b_{az}, b_{gx}, b_{gy}, b_{gz}\right]^T, \tag{12}$$

where b_{nm} is the bias estimate with the n subscript denoting a for accelerometer and g for gyro, and the m subscript denoting which axis the sensor is measuring (x, y, or z). The $\mathbf{A}$ matrix described in Eq. (3) is modified to

$$\mathbf{A} = \begin{bmatrix} I_{10\times 10} & 0 & 0 \\ I_{10\times 10} & 0 & 0 \\ 0 & 0 & I_{6\times 6} \end{bmatrix}, \tag{13}$$

to maintain bias estimates between measurements from the visual sensor. The time update that utilizes the inertial sensors to update the first 10 elements of the state is modified to subtract out the bias estimate from the inertial measurements. The measurement step for the filter remains unchanged. In the following section, we present results demonstrating the improved navigation performance obtained from estimating the biases of the inertial sensors.

5 Results

To demonstrate the results of fusing visual, air pressure, and inertial sensors together as proposed in this paper, we developed a detailed simulator that generated synthetic MAV flights, together with uncorrupted and corrupted sensor measurements for that flight. In the subsections that follow, we first describe the simulation environment in more detail, followed by results demonstrating the efficacy of the fusion system proposed in this paper.

5.1 Simulation Environment

To enable an evaluation of our epipolar constraint-based fusion environment, we first need to generate the true navigation states of the MAV over time. To generate true location data about the MAV, a Bézier curve representing the true path of the MAV was created. A Bézier curve was chosen due to its inherent flexibility in representing many different types of curves in 3-D space. In addition, Bézier curves are a polynomial function of a single scalar t, yielding two significant advantages. First, the location at any time can be easily determined. Second, by differentiating the polynomial with respect to t, the velocity and acceleration at any point on the curve can be computed in closed form. All quantities are assumed to be in a "navigation frame" which has North as its x axis, East as its y axis, and straight down as the z axis. The origin of this frame was arbitrarily chosen as a location on the ground in Utah, near Brigham Young University (close to our MAV flight test area).

In addition to generating the location, velocity, and acceleration of the MAV, we also need to generate the angular orientation (attitude) of the MAV camera. We have used two basic approaches to generating the attitude of the MAV camera. First, for a "fixed" camera, the angular orientation is always constant within the MAV body frame. Second, we set the attitude of the camera such that a specified world location will always be in the center of the image, representing a gimbaled camera that remains pointed at a specific location. We utilize the second approach for the results presented in this section.

Once the true location and attitude of the camera are known, the inputs to the fusion algorithm are generated. We assume the inputs from the IMU consist of 3-axis accelerometer and gyroscope (gyro) readings. To generate accelerometer readings, the acceleration of the camera is computed from the Bézier curve. The effects of gravity, Coriolis, and the rotation of the earth are then added to the accelerometer readings as described in [16], yielding noise-free accelerometer readings. To generate gyro readings, the attitude at two locations on the Bézier curve is computed. The locations on the curve are separated by the gyro sample time. The difference in attitude is then used to compute the angular rates of the camera, yielding noise-free gyro readings. Noise-free pitot tube readings are computed as the magnitude of the velocity at a point on the Bézier curve.

Once the noise-free readings have been computed, two types of noise are added to the sensor readings. First, Gaussian, zero-mean white noise is added to the computed readings. The variance of the noise values were chosen to approximate measurement errors observed on a Kestrel autopilot [3]. Second, a constant bias is added to the gyro and accelerometer readings. For each run of the simulator, biases were randomly selected from a Gaussian distribution with twice the standard deviation of the white noise for that sensor.

To simulate inputs from the camera, a set of random world points to be imaged are created. Using the locations of the world points and the location and attitude of the MAV over time, a set of feature locations corresponding with time along its flight path are created. Features locations for a specific MAV location and attitude are computed using the formula

$$\lambda \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \mathbf{K}\mathbf{C}_n^c(\mathbf{X}^n - \mathbf{p}^n), \tag{14}$$

where $\mathbf{X}^n$ was the location of the world point (in the navigation frame), $\mathbf{p}^n$ is the position of the camera in navigation frame coordinates (determined from its point on the Bézier curve), $\mathbf{C}_n^c$ is the direction cosine matrix from the navigation frame to the camera frame (also a function of location on the Bézier curve), $\mathbf{K}$ is the calibration matrix of the camera, mapping from Euclidean to pixel locations, λ is a scale factor used for normalizing the third element of the image frame vector to 1, and x_i and y_i are the image coordinates of the point.

After determining the location of the object in the image space, Gaussian white zero-mean noise is added to the image location. We set the standard deviation of the noise equal to a single pixel in the image plane. After adding noise, the pixel values are then "de-calibrated" (multiplied by $\mathbf{K}^{-1}$) to obtain vectors in the same Euclidean space as the MAV navigation state.

5.2 Fusion Results

To test the efficacy of our proposed fusion environment, we use two different "flight scenarios." In the first scenario, the MAV moves in a straight line starting at 100 m above the ground and 100 m south of the navigation frame origin. The camera then moves in a straight line to 100 m north of the navigation frame origin, holding a constant altitude. In the East-West (y) direction, the MAV is always at 0. Along this path, 161 images were captured at a rate of 10 Hz, requiring 16 s to fly this path. These values were chosen to achieve an airspeed (12.5 m/s) typical of MAVs. In addition, 1,601 samples of the gyro and accelerometer readings were collected. Note that while this flight scenario may seem like an overly simplistic maneuver (flying in a straight line), it was chosen because it actually exacerbates one of the fundamental problem of vision, the universal scale ambiguity. Therefore, this scenario is one of the most difficult scenarios for vision-aided navigation. Results for this scenario are shown in Table 3. Note that this scenario was used for the partial results presented earlier in this paper.

The second scenario represents a more generic flight of an MAV. It starts at −100 m north, 100 m in altitude. It then flies an "S" pattern, going northeast before turning to go northwest. While flying northwest, it passes directly over the navigation frame origin, after which it turns back to head northeast, arriving at −30 m east, 100 m north. During the course of the flight, the altitude also drops from 100 m to 60 m. This entire flight takes 19 s. We refer to this scenario as "The S Pattern", with results shown in Table 4.

In both scenarios described above, the world points being observed were distributed using a three-dimensional Gaussian distribution centered about the navigation frame origin. To keep the objects in view, the camera is continuously rotated to "look at" the origin.

To determine the overall accuracy of each fusion technique, we ran each UKF filter setup with each flight path scenario 100 times. In Tables 3 and 4, the mean and standard deviation of the errors across 100 runs of the filter are shown. The mean and standard deviation achieved using only the IMU is also shown for each flight scenario as a reference. The units for the final position error (p_x, p_y, and p_z) are in meters, while the final attitude errors are in degrees. The attitude errors are the amount of yaw (ψ), pitch (θ) and roll (ϕ) that would be required to move from the true location to the estimated locations.

For each of these flight scenarios, five different setups of our UKF environment were used. First, we ran epipolar constraint-based fusion without any of the modifications introduced in Sect. 3 (*Baseline*). Second, we remove the bias in the direction the camera is pointed as discussed in Sect. 3.1 (sin *Removed*). Third, the measurements from the pitot tube are included in the UKF framework (*Pitot Added*). Fourth, a slower sampling rate (2 Hz, as opposed to 10 Hz) is used in addition to all the other modifications (*Min. Sampling*). Finally, the "Min. Sampling" filter is modified to estimate the inertial sensor biases (*Est. Bias*).

As shown in these tables, each modification proposed in this paper significantly reduces the mean and standard deviation of the error. By including all four proposed

Table 3 Mean and standard deviation of error in the final estimated location and attitude of the MAV when flying a straight-line path with a gimbaled camera. Note that the errors in attitude are all under one degree when all four modifications proposed in this paper are implemented

	IMU only	Baseline	sin Removed	Pitot Added	Min. Sampling	Est. Bias
p_x **Error** (μ,σ)	(−4.48, 115.4)	(−607.0, 136.6)	(23.1, 85.9)	(−0.47, 1.92)	(−0.39, 1.19)	(0.03, 1.13)
p_y **Error** (μ,σ)	(8.27, 95.3)	(2.83, 76.5)	(−1.77, 16.0)	(−1.01, 11.2)	(−0.24, 3.05)	(−.08, 1.96)
p_z **Error** (μ,σ)	(15.1, 12.7)	(757.5, 318.1)	(11.8, 9.71)	(2.85, 1.08)	(11.1, 1.17)	(2.19, 0.82)
ψ **Error** (μ,σ)	(−0.69, 14.0)	(2.26, 17.1)	(0.20, 2.52)	(0.21, 2.71)	(0.09, 1.07)	(0.02, 0.58)
θ **Error** (μ,σ)	(0.16, 15.93)	(−143.5, 61.3)	(1.57, 6.59)	(−0.31, 1.70)	(1.23, 1.78)	(0.27, 0.61)
ϕ **Error** (μ,σ)	(−1.05, 12.9)	(2.90, 18.1)	(0.65, 5.67)	(0.26, 4.13)	(0.08, 1.13)	(0.02, 0.72)

Table 4 Mean and standard deviation of error in the final estimated location and attitude of the MAV when flying the "S-curve" path with a gimbaled camera

	IMU only	Baseline	sin Removed	Pitot Added	Min. Sampling	Est. Bias
p_x **Error** (μ,σ)	(0.26, 110.4)	(−128.6, 130.9)	(−22.8, 88.1)	(−1.89, 23.1)	(0.79, 2.39)	(0.08, 1.73)
p_y **Error** (μ,σ)	(−15.4, 132.4)	(78.0, 354.0)	(−11.4, 92.7)	(0.91, 11.8)	(6.20, 2.75)	(2.09, 1.82)
p_z **Error** (μ,σ)	(18.8, 33.6)	(528.3, 278.9)	(41.4, 102.0)	(−2.90, 4.04)	(5.18, 2.79)	(−1.58, 1.36)
ψ **Error** (μ,σ)	(0.91, 14.5)	(39.3, 31.4)	(5.01, 19.4)	(0.38, 3.20)	(−0.23, 2.67)	(0.26, 1.40)
θ **Error** (μ,σ)	(−0.07, 11.5)	(−27.1, 26.9)	(−3.84, 23.0)	(−1.18, 4.91)	(0.78, 1.87)	(−0.41, 0.99)
ϕ **Error** (μ,σ)	(2.32, 13.2)	(−55.7, 47.9)	(−1.40, 25.6)	(1.85, 5.59)	(0.71, 1.28)	(0.65, 0.73)

modifications, more than an order of magnitude decrease in error is achieved from both IMU-only navigation and the baseline fusion approach. This demonstrates the necessity of including the proposed modifications when considering epipolar constraint based fusion for navigation. It also demonstrates the advantages of estimating the inertial sensor biases to help reduce noise.

6 Conclusion

In this paper, we have proposed a system for fusing IMU and visual information together in a UKF framework utilizing the epipolar constraint. However, a naive implementation of this approach yields sub-optimal results. Therefore, we propose three modifications to the baseline fusion setup that significantly improve the overall performance of this system. These include: (1) removing the bias toward the center of tracked points when using the epipolar constraint, (2) using a pitot tube to measure the velocity of th MAV, and (3) minimizing the sampling rate of the image data to achieve minimal error growth. After implementing these improvements it is possible to estimate the biases of the inertial sensors, decreasing the amount of noise present in the system. Gains in accuracy of at least 10X in both the location and attitude estimates were achieved using these improvements.

Acknowledgments This work was funded by an AFOSR Young Investigator Award, number FA9550-07-1-0167. This work began while the author was a summer faculty fellow with the Air Force Research Labs under the supervision of Mikel Miller in the Munitions Directorate. Assistance was also provided by Mike Veth of the Air Force Institute of Technology.

References

1. Beard R, Kingston D, Quigley M, Snyder D, Christiansen R, Johnson W, McLain T, and Goodrich M (2005) Autonomous vehicle technologies for small fixed wing UAVs. *AIAA Journal of Aerospace Computing, Information, and Communication* 2(1): 92–108.
2. Kingston DB and Beard RW (2004) Real-time attitude and position estimation for small uav's using low-cost sensors. In: *AIAA 3rd Unmanned Unlimited Systems Conference and Workshop*, Chicago, IL.
3. Procerus Technologies (2008) Procerus technologies URL http://www.procerusuav.com
4. Micropilot (2008) Micropilot, URL http://www.micropilot.com
5. Strelow D and Singh S (2004) Motion Estimation from Image and Inertial Measurements. *The International Journal of Robotics Research* 23(12): 1157–1195.
6. Veth MJ, Raquet JF, and Pachter M (2006) Stochastic constraints for efficient image correspondence search. *IEEE Transactions on Aerospace and Electronic Systems* 42(3): 973–982.
7. Veth M and Raquet J (2007) Fusing low-cost image and inertial sensors for passive navigation. *Journal of the Institute of Navigation* 54(1): 11–20.
8. Langelaan J (2006) State estimation for autonomous flight in cluttered environments. PhD thesis, Stanford University.
9. Prazenica R, Watkins A, Kurdila A, Ke Q, and Kanade T (2005) Vision-based kalman filtering for aircraft state estimation and structure from motion. In: *2005 AIAA Guidance, Navigation, and Control Conference and Exhibit*, pp. 1–13.

10. Soatto S, Frezza R, and Perona P (1996) Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control* 41(3): 393–413.
11. Ready BB and Taylor CN (2007) Improving accuracy of mav pose estimation using visual odometry. In: *American Control Conference, 2007*. ACC '07, pp. 3721–3726.
12. Andersen ED and Taylor CN (2007) Improving mav pose estimation using visual information. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.3745–3750.
13. Sorenson HW (1966) Kalman filtering techniques. In: Leondes CT (ed) *Advances in Control Systems, Theory and Applications*, vol 3, pp. 218–292.
14. Xing Z and Gebre-Egziabher D (2008) Modeling and bounding low cost inertial sensor errors. In: *Proceedings of IEEE/ION Position Location and Navigation Symposium*, pp. 1122–1132.
15. IEEE Std 952 (1997) IEEE Standard specification formate guide and test procedures for single-axis interferometric fiber optic gyros. *IEEE Standard* 952–1997.
16. Titterton D and Weston J. (1997) *Strapdown Inertial Navigation Technology*. Lavenham, United Kingdom: Peter Peregrinus Ltd.
17. Cloudcap (2008) Cloud cap technology. URL http://www.cloudcaptech.com