State Estimation for Indoor and Outdoor Operation with a Micro-Aerial Vehicle

Shaojie Shen and Nathan Michael

Abstract. In this work, we detail a methodology for estimating the state of a micro-aerial vehicle (MAV) as it transitions between different operating environments with varying applicable sensors. We ensure that the estimate is smooth and continuous throughout and provide an associated quality measure of the state estimate. We address the challenge of maintaining consistency between local and global measurements and propose a strategy to recursively estimate the transform between different coordinate frames. We close with experiments that validate the approach and the resulting performance as a MAV navigates between mixed indoor and outdoor environments.

1 Introduction

In this paper, we focus on the problem of estimating the state of a micro-aerial vehicle (MAV) while operating in and transitioning between indoor and outdoor environments. Aerial vehicles offer mobility and perspective advantages over ground platforms and can transition through small openings such as windows and doors. This fact makes MAVs particularly applicable to problems such as search-and-rescue and surveillance. In recent work, we focused on the problem of autonomous indoor navigation in multi-floor environments using only onboard sensing and computation [1]. Through our previous work, we developed a methodology that enables mapping, localization, planning, and control of a MAV in complex indoor environments. This paper builds upon this previous work to permit operation in both indoor and outdoor environments, including transitions between these mixed environments.

The problem of autonomous navigation with an aerial vehicle in outdoor or indoor environments is thoroughly studied in the literature. Julier and Uhlmann [2] and Merwe et al. [3] propose state estimation methods using IMU and GPS sensors

Shaojie Shen · Nathan Michael
GRASP Laboratory
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: {shaojie,nmichael}@grasp.upenn.edu

J.P. Desai et al. (Eds.): *Experimental Robotics*, STAR 88, pp. 273–288.
DOI: 10.1007/978-3-319-00065-7_20

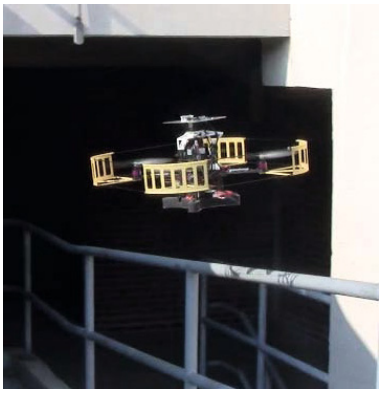


Fig. 1 We address the estimation of the state of a micro-aerial vehicle as it transitions between indoor and outdoor environments. While the robot is equipped with a laser, GPS, magnetometer, and IMU, only a subset of those sensors may be operational at any time given the robot's state and the sensor operating conditions.

for navigation in outdoor environments. Recent developments toward autonomous MAV operation in indoor environments using onboard sensors (IMU, laser, and camera) include the works of Grzonka et al. [4], Bachrach [5], and Blosch et al. [6] as well as our own work [1].

A challenge in operating in both indoor and outdoor environments is the fact that the performance of sensors can vary between environments; sensors can provide meaningful data in one environment while becoming compromised in another environment. In this work, we consider a vehicle equipped with a standard GPS unit, laser range finder, magnetometer, and IMU (Fig. 1). The GPS provides relatively inaccurate and latent information compared to the other sensors and is only operational when outdoors and in view of satellites. Similarly, the magnetometer only provides accurate information when outdoors. The laser scanner provides accurate information indoors but is compromised by direct sunlight and provides limited to no information in large open spaces. The IMU is consistent and operational in both indoor and outdoor environments but is too inaccurate to enable stable feedback control as the only source of information. Conditions that challenge the sensor capabilities extend beyond just indoor and outdoor transitions and include negotiating GPS-shadowing due to buildings and trees. A summary of sensor characteristics and limitations is provided in Table 1.

Table 1 Characteristics and limitations of the onboard sensors on our MAV platform

Sensor	Indoor	Transition	Outdoor	Frame	Accuracy	Relevant Failure Modes
GPS	×	√	√	Global	Low	Obstructed view of satellites
Magnetometer	×	√	√	Global	Low	Magnetic interference
Laser	√	√	×	Body	High	Direct sunlight, open spaces
IMU	√	√	√	Body	Low	None

Consistency between local and global reference frames also becomes a challenge when operating in environments where the primary position sensors provide either relative or global state information. The fusion of global observations from GPS and relative observations from sensors such as lasers and cameras is addressed by Carlson [7] and Schleicher et al. [8]. These authors propose strategies to introduce additional global constraints to the existing simultaneous localization and mapping (SLAM) formulation to handle the correspondence between local and global frames. Although these methods yield reasonable results, there is no guarantee that the resulting pose estimate is low-latency and smooth, which is of vital importance for stable feedback control of MAVs. Moore et al. [9] investigate the problem of maintaining smoothness in the state estimate and handling position discontinuity due to GPS interference, but these methods do not directly extend to a MAV as the approach relies on information only available to a ground vehicle (wheel odometry). Additionally, we require a notion of the quality of the estimate as we wish to leverage robust control methods that require both a mean and covariance measure of the state estimate [10].

While there are similarities in terminology in our presentation to submap-based SLAM approaches [11, 12], there are fundamental differences in the problem definition and approach. For submap-based SLAM, local maps are created via a sparse sampling of the robot trajectory and local maps are assumed to be accurate. Local maps are linked by incremental motion constraints or loop-closure correspondence. Pose graph optimization techniques can be applied to obtain globally consistent maps [13, 14]. However, for this work, local maps are created based on sensor availability rather than distance, we do not limit the size of local maps, and we allow for the existence of estimation errors within each local map. Further, our local maps are not linked by any incremental motion constraints, and thus the above optimization techniques are no longer applicable.

2 Approach

In this section we discuss our approach to estimating the state of the MAV during indoor and outdoor operations. We discuss the necessity of using different reference frames for state estimation in Sect. 2.1. In Sect. 2.2, we detail an Unscented-Kalman Filter (UKF) formulation that includes process and measurement models for the sensors under consideration. We discuss globally consistent map building and management, as well as loop closure detection and map correction in Sect. 2.3. We conclude by integrating planning and control modules into our proposed approach to form a complete autonomous aerial navigation system (Sect. 2.4).

We build upon our previous work [1] and therefore assume the existence of a laser-based SLAM solution running in real-time onboard the robot. The output of the laser-based SLAM is a pose estimate of the robot in three dimensions with respect to the map generated via the laser observations.

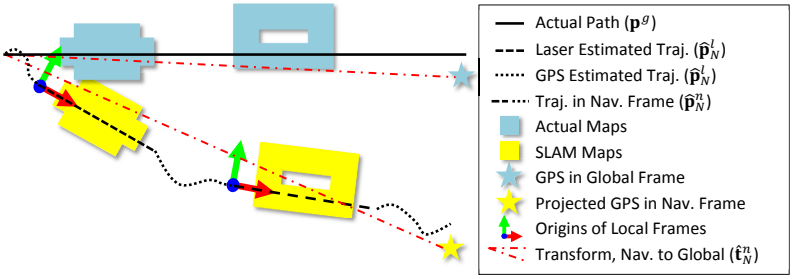


Fig. 2 Graphical illustration of the use of different coordinate frames in our work. The robot travels along a straight line (black) between two buildings (cyan) in the global frame. When navigating outdoors and away from buildings, only GPS is available as the primary position sensor, and the resulting pose estimate is inaccurate (dotted line). Laser-based SLAM provides an accurate position estimate (dashed line) in a local map (yellow) when navigating indoors or close to the buildings. We address the problem of ensuring consistency between global and local information sources by creating a navigation frame and estimate its transform to the global frame using a recursive formulation.

2.1 Reference Frames

We define four reference frames for this work.

- Body frame: defined with respect to the body of the robot and denoted by $(\cdot)^b$. All laser and IMU measurements are made in the body frame.
- Global frame: defined with respect to the inertial frame and denoted by $(\cdot)^g$. All global measurements (GPS and magnetometer) are made in the global frame.
- Local frame: defined with respect to the origin of the laser-based localization solution when transitioning between regions of mixed sensor information (clarified below) and denoted by $(\cdot)^l$.
- Navigation frame: defines the reference frame used by the onboard feedback control (clarified below) and denoted by $(\cdot)^n$.

Multiple local frames may be defined through the course of an experiment (see Fig. 2). A new local frame is created when the covariance matrix associated with the laser-based localization, which is computed via an inversion of the Fisher's Information Matrix of the laser scan [15], changes from singular to nonsingular. Such events correspond to when the laser sensor provides sufficiently salient information that we may again include this information in the state estimate (i.e. the sensor is no longer in a failure mode). In experimentation, this event usually occurs when the robot starts to observe strong geometric structures (e.g. buildings) or escapes from strong direct sunlight. Note that even if the laser-based localization fails to provide a valid solution, the state of the robot in the current local frame can still be estimated via sensor measurements made in other frames (IMU, GPS and magnetometer) and transformed into the local frame.

As noted above, maintaining a consistent, smooth, and continuous state estimate for use by the onboard feedback control is of particular importance when working with a MAV and considering operation in mixed environments. We define the frame associated with this state estimate as the navigation frame and now detail its definition. For the sake of brevity, we only consider the pose of the robot, $\mathbf{p}$, in this section and simplify the estimation problem to consider a 2D system (x and y) where the vertical position and velocity of the robot is estimated by downward facing laser beams and a pressure sensor. Therefore, the GPS and magnetometer sensors provide 2D pose and velocity observations while the laser-based localization system provides a 2D pose estimate. Roll and pitch angles are obtained directly from the IMU.

Denote the true global pose of the robot as $\mathbf{p}^g$ and the global pose estimate of the robot as $\hat{\mathbf{p}}^g$. Similarly, the pose in the k^{th} local frame is given by $\mathbf{p}_k^l$. The direct output of the laser-based localization system, which is a noisy estimate of $\mathbf{p}_k^l$, is defined as $\hat{\mathbf{p}}_k^l$. To simplify the presentation, we use the pose compounding ($\oplus$) and inverse ($\ominus$) operators to capture the transformation between different coordinate frames [16].

Consider a robot, initialized at the global origin, that moves into and out of regions with available laser-based localization information such that it now considers the definition N local frames. The true global pose of the robot is:

$$\mathbf{p}^g = \mathbf{p}_1^{l_f} \oplus \mathbf{p}_2^{l_f} \oplus \cdots \oplus \mathbf{p}_{N-1}^{l_f} \oplus \mathbf{p}_N^l \tag{1}$$

where $\mathbf{p}_k^{l_f}$ is the final pose of the vehicle in the k^{th} local frame. As $\mathbf{p}_k^{l_f}$ is unavailable, we consider the estimated value:

$$\begin{aligned} \hat{\mathbf{p}}_k^{l_f} &= \mathbf{p}_k^{l_f} \oplus \mathbf{e}_k^{l_f} \quad k = \{1, \cdots, N-1\} \\ \hat{\mathbf{p}}_N^l &= \mathbf{p}_N^l \oplus \mathbf{e}_N^l \end{aligned}$$

where $\mathbf{e}_k^{l_f}$ and $\mathbf{e}_N^l$ are estimation errors. We define the pose in the navigation frame as follows:

$$\begin{aligned} \hat{\mathbf{p}}_N^n &= \hat{\mathbf{p}}_1^{l_f} \oplus \hat{\mathbf{p}}_2^{l_f} \oplus \cdots \oplus \hat{\mathbf{p}}_{N-1}^{l_f} \oplus \hat{\mathbf{p}}_N^l \\ &= \left(\mathbf{p}_1^{l_f} \oplus \mathbf{e}_1^{l_f} \oplus \mathbf{p}_2^{l_f} \oplus \mathbf{e}_2^{l_f} \oplus \cdots \oplus \mathbf{p}_{N-1}^{l_f} \oplus \mathbf{e}_{N-1}^{l_f} \oplus \mathbf{p}_N^l\right) \oplus \mathbf{e}_N^l \\ &= \mathbf{p}_N^n \oplus \mathbf{e}_N^l \end{aligned} \tag{2}$$

where the subscript $(\cdot)_N$ in $\mathbf{p}_N^n$ indicates the number of local frames integrated into the navigation frame. Hence, the navigation frame serves a similar role as wheel odometry for ground robots. Clearly $\mathbf{p}_N^n \neq \mathbf{p}^g$ as $\mathbf{p}_N^n$ incorporates the accrual of error in pose estimates and is therefore suitable for applications that mainly require only local accuracy, such as feedback control and obstacle avoidance. Thus, we define $\mathbf{p}_N^n$ as the robot pose in the navigation frame (see Fig. 2). The estimated pose in the navigation frame, $\hat{\mathbf{p}}_N^n$, follows a similar error model as the pose estimate in the current local frame $\hat{\mathbf{p}}_N^l$.

We must now consider the transform between the global frame and the navigation frame. For notational convenience, we first define $\mathbf{q}_k$ as:

$$\mathbf{q}_k = \hat{\mathbf{p}}_1^{l_f} \oplus \hat{\mathbf{p}}_2^{l_f} \oplus \cdots \oplus \hat{\mathbf{p}}_k^{l_f}. \tag{3}$$

Therefore, if there are N local frames, then $\mathbf{q}_1, \cdots, \mathbf{q}_{N-1}$ are all known and readily computed constant transforms.

Define the time-varying rigid body transform from the navigation frame to the global frame as $\mathbf{t}_N^n$ such that:

$$\mathbf{p}^g = \mathbf{t}_N^n \oplus \mathbf{p}_N^n. \tag{4}$$

As $\mathbf{t}_N^n$ is not directly accessible, we estimate this transform (with error $\mathbf{e}$):

$$\hat{\mathbf{t}}_N^n = \mathbf{t}_N^n \oplus \mathbf{e}.$$

The resulting global pose estimate may be written as:

$$\hat{\mathbf{p}}^g = \hat{\mathbf{t}}_N^n \oplus \hat{\mathbf{p}}_N^n = \mathbf{t}_N^n \oplus \mathbf{e} \oplus \mathbf{p}_N^n \oplus \mathbf{e}_N^l.$$

For this work, we consider the state estimation problem in the navigation frame, transforming all global sensor information into this frame via $\mathbf{t}_N^n$. We now discuss a recursive filtering formulation that permits estimation of this transform along with the state of the robot.

2.2 UKF-Based Sensor Fusion

We employ a UKF framework with delayed measurement compensation to estimate the pose and velocity of the robot, hidden sensor bias parameters, and the transformation between the global and navigation frames [3]. All quantities, unless otherwise specified, are defined with respect to the navigation frame. We also assume that N local frames have been created since the beginning of the experiment. The subscript and superscript $(\cdot)_N^n$ will be omitted for the remainder of the paper. The system state is defined as:

$$\mathbf{x} = \left[\mathbf{r}, \dot{\mathbf{r}}, \mathbf{\Phi}, \mathbf{a}_b, \mathbf{\Psi}_b, \mathbf{t}_N^l\right]^{\mathrm{T}} \tag{5}$$

where $\mathbf{r} = [x, y, z]^{\mathrm{T}}$ is the position of the robot and $\mathbf{\Phi} = [\phi, \theta, \psi]^{\mathrm{T}}$ is the roll, pitch, and yaw Euler angles that represent the 3D orientation of the robot. $\mathbf{a}_b = \left[a_{b_x}, a_{b_y}, a_{b_z}\right]^{\mathrm{T}}$ is the bias of the 3D accelerometer in the body frame and $\mathbf{\Psi}_b = [\phi_b, \theta_b]^{\mathrm{T}}$ is the bias of the roll and pitch estimate from the IMU. $\mathbf{t}_N^l = \left[\Delta x^l, \Delta y^l, \Delta \psi^l\right]^{\mathrm{T}}$ is a 2D rigid body transformation between the global frame and navigation frame and related to $\mathbf{t}_N^n$. The definition of $\mathbf{t}_N^l$ is discussed later in this section. As the UKF formulation follows a similar structure to existing works [3], we only briefly discuss the process model and provide specific details pertinent to the measurement update step.

Process Model

We consider an IMU-based process model:

$$\begin{aligned} \mathbf{x}_{t+1} &= f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) \\ \mathbf{u} = [\omega, \mathbf{a}]^{\mathrm{T}} &= [\omega_x, \omega_y, \omega_z, a_x, a_y, a_z]^{\mathrm{T}} \\ \mathbf{v} &= [\mathbf{v}_\omega, \mathbf{v}_\mathbf{a}, \mathbf{v}_{\mathbf{a_b}}, \mathbf{v}_{\mathbf{\Psi_b}}]^{\mathrm{T}} \end{aligned}$$

where $\mathbf{u}$ is the body frame angular velocities and linear accelerations from the IMU. $\mathbf{v}$ represents additive Gaussian noise associated with the gyroscope, accelerometer, accelerometer bias, and IMU attitude bias.

Measurement Model – Laser-Based SLAM

The measurement from the laser-based localization relative to the local frame is:

$$\tilde{\mathbf{z}}_N^l = \left[\tilde{x}_N^l, \tilde{y}_N^l, \tilde{\psi}_N^l\right]^{\mathrm{T}}.$$

We transform this measurement into the navigation frame prior to the measurement update. The transformed laser measurement is:

$$\tilde{\mathbf{z}}_N^n = \mathbf{q}_{N-1} \oplus \tilde{\mathbf{z}}_N^l$$

where $\mathbf{q}_{N-1}$ is defined in (3). The measurement model is linear and may be written as:

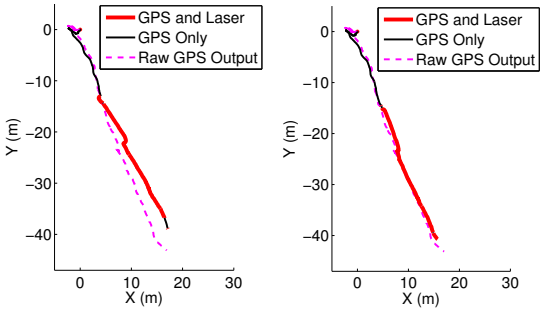
$$\mathbf{z}_N^n = H^n \mathbf{x} + \mathbf{n}_N^n, \quad H^n = \begin{bmatrix} 1 & 0 & & 0 & \\ 0 & 1 & \mathbf{0}_{3\times 6} & 0 & \mathbf{0}_{3\times 8} \\ 0 & 0 & & 1 & \end{bmatrix}$$

where H^n extracts the 3-DOF pose in the state and $\mathbf{n}_N^n$ is additive Gaussian noise.

Note that the laser-based localization error covariance matrix is obtained in the local frame, not the navigation frame. We must transform this covariance matrix to the navigation frame. Assuming independence between the noise for position and heading, the covariance matrix of $\mathbf{n}_N^n$ is:

$$\Sigma_N^n = \begin{bmatrix} \Sigma_{xy}^n & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{1\times 2} & \sigma_\psi^{n2} \end{bmatrix}, \quad \Sigma_{xy}^n = R_{\psi_{q_{N-1}}} \Sigma_{xy}^l R_{\psi_{q_{N-1}}}^{\mathrm{T}}$$

where $R_{\psi_{q_{N-1}}}$ is the 2D rotation matrix obtained from the heading component in $\mathbf{q}_{N-1}$, and Σ_{xy}^l is a laser-based covariance measure and computed using the methods proposed in [15]. The standard deviation in heading remains the same in both the local frame and the navigation frame (e.g. $\sigma_\psi^n = \sigma_\psi^l$). The measurement update for the laser-based localization is linear and can be performed via a KF update step.



(a) State estimation via $\hat{\mathbf{t}}_N^n$ (b) State estimation via $\hat{\mathbf{t}}_N^l$

Fig. 3 The effects of accrued error in the estimate of the transformation from the navigation frame into the global frame, $\mathbf{t}_N^n$, are amplified as the robot moves away from the global origin. We choose to estimate $\mathbf{t}_N^l$ rather than $\mathbf{t}_N^n$ directly in our recursive filter formulation. Here, we show experimental data collected when transitioning between local frames. The red and black lines show the measurement prediction, $\mathbf{z}^g$, and the dashed magenta lines show the raw GPS position measurement, $\tilde{\mathbf{z}}^g$. In Fig. 3(a), we try to estimate $\mathbf{t}_N^n$ directly. As the robot moves away from the origin, the UKF is unable to directly estimate $\mathbf{t}_N^n$, resulting in inconsistency between the measurement prediction and the actual GPS measurement. Figure 3(b) shows the result of the proposed approach (estimating $\mathbf{t}_N^l$), where the measurement prediction is consistent with the actual measurement.

Measurement Model – GPS and Magnetometer

GPS provides information about the horizontal vehicle position and velocity in the global frame. We ignore GPS altitude information as it is generally inaccurate. The magnetometer provides a measure of global orientation. The assembled measurement vector is:

$$\tilde{\mathbf{z}}^g = \left[\tilde{x}^g,\ \tilde{y}^g,\ \tilde{\dot{x}}^g,\ \tilde{\dot{y}}^g,\ \tilde{\psi}^g\right]^{\mathrm{T}}.$$

We now motivate the estimation of $\mathbf{t}_N^l$ in the system state (5). Recall that $\mathbf{t}_N^n$ represents the transformation from the navigation frame into the global frame and includes the accrual of estimation errors during the evolution of the vehicle operation. As the robot moves away from the global origin, the effects due to error in the estimate of the transformation are amplified. Consequently, the estimator performance decreases. Therefore, rather than estimate the transform directly, we treat the transform as a compound operation on the current and prior transforms and pose estimates:

$$\mathbf{t}_N^n \oplus \mathbf{q}_{N-1} = \hat{\mathbf{t}}_{N-1}^{n_f} \oplus \mathbf{q}_{N-1} \oplus \mathbf{t}_N^l$$

where $\hat{\mathbf{t}}_{N-1}^{n_f}$ is the estimated transformation between the global frame and the navigation frame at the time of the initialization of the current local frame. This approach moves the origin of the frame transform from the global origin to $\mathbf{q}_{N-1}$.

Based on the above discussion, the GPS measurement model is defined as:

$$\mathbf{t}_N^n = \hat{\mathbf{t}}_{N-1}^{n_f} \oplus \mathbf{q}_{N-1} \oplus \mathbf{t}_N^l \ominus \mathbf{q}_{N-1} = [\Delta x,\, \Delta y,\, \Delta\psi]^{\mathrm{T}}$$

$$\mathbf{z}^g = h^g(\mathbf{x}) + \mathbf{n}^g = \begin{bmatrix} R_{\Delta\psi} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \\ R_{\Delta\psi} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \\ \psi + \Delta\psi \end{bmatrix} + \mathbf{n}^g$$

and can be performed via a nonlinear UKF measurement update procedure. The consequence of estimating $\mathbf{t}_N^l$ and indirectly computing $\mathbf{t}_N^n$ is a significant improvement in overall estimator accuracy and consistency (see Fig. 3).

Measurement Model – Altitude, Roll, and Pitch

The vertical position and velocity of the robot is observed by downward facing laser beams and a pressure sensor while the roll and pitch angles are obtained from the IMU. Therefore, the third measurement is assembled as:

$$\tilde{\mathbf{z}}^a = \left[\tilde{z},\, \tilde{\dot{z}},\, \tilde{\phi},\, \tilde{\theta}\right]^{\mathrm{T}}.$$

As $\mathbf{z}^a$ is a subset of the system state $\mathbf{x}$, this measurement model follows a linear form:

$$\mathbf{z}^a = H^a\mathbf{x} + \mathbf{n}^a, \quad H^a = \begin{bmatrix} & 1 & & 0\,0\,0 & & 0\,0 & \\ & 0 & & 1\,0\,0 & & 0\,0 & \\ \mathbf{0}_{4\times2} & 0 & \mathbf{0}_{4\times2} & 0\,1\,0 & \mathbf{0}_{4\times4} & 1\,0 & \mathbf{0}_{4\times3} \\ & 0 & & 0\,0\,1 & & 0\,1 & \end{bmatrix}$$

where H^a extracts the required elements in the system state and $\mathbf{n}^a$ is additive Gaussian noise.

2.3 3D Map Generation

We associate a 3D local map with each local frame. If a valid laser-based localization solution is available, the laser scan is transformed into the local frame and the 3D map is created and updated via a multi-volume occupancy grid data structure [17]. We may transform the local map associated with local frame k into the global frame via its origin in the navigation frame $\mathbf{q}_{k-1}$ and the corresponding estimated frame transform $\hat{\mathbf{t}}_k^n$.

We match the current local map against all previous local maps using a multi-resolution scan matching algorithm [18]. Detection of correspondences between the current and previous local maps results in the alignment and merger of these maps.

2.4 Planning and Control

While the focus of this work is on estimation, we validate our approach by integrating it with the planning and control components of the autonomous aerial navigation system. All experimental results consider the overall performance of the navigation system.

We use an RRT-based planner for online trajectory generation and obstacle avoidance given the current 3D map. The desired goals for the planner are provided by high-level commands from the human operator.

We employ a gain-scheduled LQR controller based on pre-computed optimal gains at various levels of state estimate accuracy [10]. To establish the expected state estimate accuracy across different environments, we experimentally obtain a set of representative covariance matrices by flying the robot in different environments and determine LQR gains offline. In practice, we find that this step is only necessary once and the pre-computed gains are applicable to all other experimental environments.

All high-level goals are provided by the operator in the global frame and transformed into the navigation frame prior to being sent to the planner and controller. Note that the transform $\mathbf{t}_N^n$ is time varying and therefore the desired waypoint in the navigation frame changes over time. This subtle point highlights a key contribution and goal of this work, that despite the drift and non-smooth nature of GPS and magnetometer information along with changing reference frames, we defer these considerations to the goal definition in the navigation frame, not the vehicle state estimate.

3 Experimental Results

3.1 Experiment Design and Implementation Details

We present experimental results to demonstrate the performance of the proposed algorithm in mixed indoor and outdoor environments. We first detail a representative experiment that investigates the validity of the methodology detailed in this work. We then briefly outline experimental trials in other environments to demonstrate repeatable performance.

The robot platform is sold by Ascending Technologies, GmbH [19] and equipped with an IMU (accelerometer, gyroscope), magnetometer, and pressure sensor. We developed custom firmware to run at the embedded level to address feedback control and estimation requirements. The other computation unit onboard is a 1.6 GHz Atom processor with 1 GB of RAM. All sensing and processing is done onboard the robot without need for external infrastructure. We outfitted the robot with a Hokuyo UTM-30LX (laser) and a uBlox LEA-5T GPS module. A custom 3D printed mount is attached to the laser that houses mirrors and redirects a small number of laser beams upward and downward. Communication with the robot for monitoring experiment progress is via 802.11n networking. Figure 1 shows a picture of our robot

platform. All algorithm development is in C++ using ROS [20] as the interfacing robotics middleware.

There are two modes of operation for our robot (1) high-level waypoint control (discussed above) and (2) kinematic control using a desired velocity command provided by the operator. Switching between control modes can be done seamlessly without aborting the experiment. In either operation mode, the robot is autonomously controlled using the onboard state feedback and operator guidance (via desired waypoint or velocity inputs). As operator input, planning, and control are not the focus of this work, we only note that we switch between these modes during the experiments based on operator preference.

3.2 Autonomous Flight with Indoor/Outdoor Transitions

We now detail a representative field experiment conducted at an urban operations (UO) testing facility. Aerial satellite imagery of the experimental site and the resulting 3D map generated online by our system are shown in Fig. 4(a). The experimental site contains indoor environments, outdoor environments with building structure, and open outdoor spaces.

The experiment begins as the robot takes off in an open courtyard where laser-based localization fails to provide any valid solution (Fig. 5(a)), The robot flies inside a nearby building via an open window (Fig. 5(b)) and navigates the various rooms and corridors of the building (Fig. 5(c)) prior to exiting the building through a door located on the second story of the building (Fig. 5(d)). The robot then flies near a sparse wooded area (Fig. 5(e)). The robot is intentionally oriented away from the building so that the laser scanner does not observe any strong geometric structures. The experiment concludes as the robot returns to the courtyard, yielding loop closure across multiple local maps (Fig. 5(f)). Three local maps are created during the experiment and aligned into a globally consistent map when the robot returns to the courtyard and detects loop closure. An overlay of the resulting globally consistent 3D map with the corresponding satellite imagery is detailed in Figs. 4(b) and 4(c).

The estimated trajectory of the robot in the navigation frame is shown in Fig. 6(a). We are always able to obtain the position estimate in the global frame, $\hat{\mathbf{p}}^g$, via the current transform estimate $\hat{\mathbf{t}}_N^n$ (according to (4)) as shown in Fig. 6(b). A comparison between the position estimate in the navigation frame and the global frame (with reference GPS measurements) is shown in Fig. 6(c). It is clear that when valid GPS measurements are available, the global position estimate, $\hat{\mathbf{p}}^g$, is consistent with these global measurements, showing the validity of the recursively estimated frame transform $\hat{\mathbf{t}}_N^n$. While the navigation frame drifts over the length of the vehicle trajectory, the role of the navigation frame is to maintain a locally consistent state estimate. The smoothness and continuity of the position and velocity estimate in the navigation frame is presented in Fig. 7.

To highlight the performance of the position estimate and associated estimator quality resulting from switching sensors, we provide the position estimate and corresponding 3σ error bounds for a subset of the trajectory in Fig. 8. We can clearly

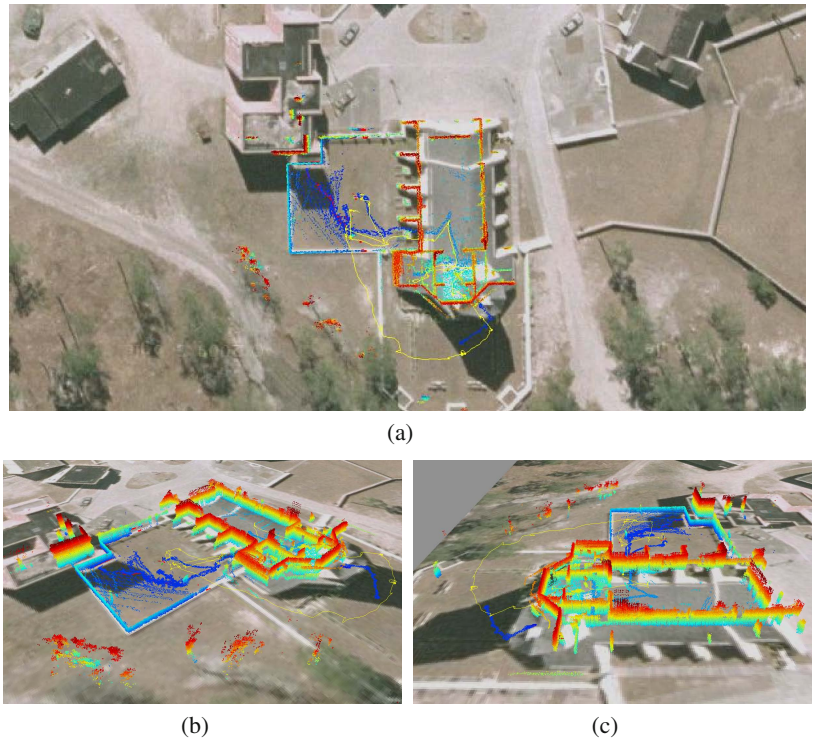


Fig. 4 Aerial imagery of the experiment site with an overlay of the 3D map generated online during the experiment. Both large open spaces and confined indoor spaces are observed in the experimental setting. Perspective views of the map are visible in Figs. 4(b)-4(c). Videos of the experiments are available at http://mrsl.grasp.upenn.edu/shaojie/ISER2012.m4v.

see changes in the estimation quality when adding or removing the laser-based localization from the system. However, note that even during rapid sensor addition and removal, the position estimate remains smooth and continuous.

3.3 Operation in Different Environments

To demonstrate that our approach is applicable in a variety of environments, we pursued two additional field tests at the Franklin Field at the University of Pennsylvania (Fig. 9(b)) and a training building at the Philadelphia Fire Academy (Fig. 9(c)). In both cases, the experiment considered at least one outdoor to indoor transition and in the latter case, also included operation in a multi-story building. Due to space constraints, we do not detail the full analysis of the experimental results here and only show an overlay of the resulting global trajectories on the corresponding satellite imagery. Although there is no ground truth information available, we can empirically

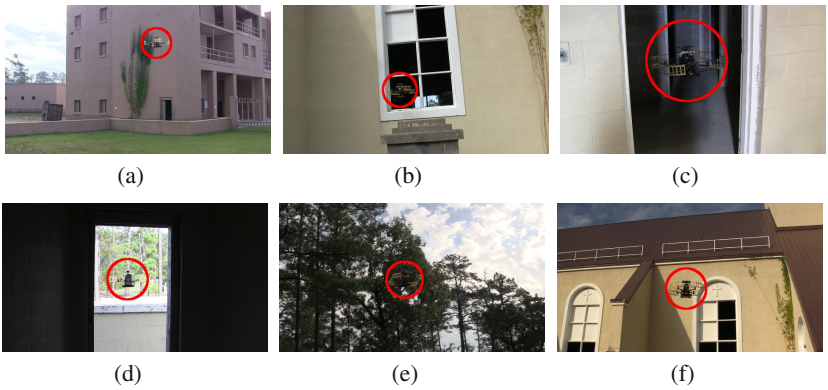


Fig. 5 The MAV flies between indoor and outdoor environments during an experimental trial. We highlight the position of the robot with a red circle.

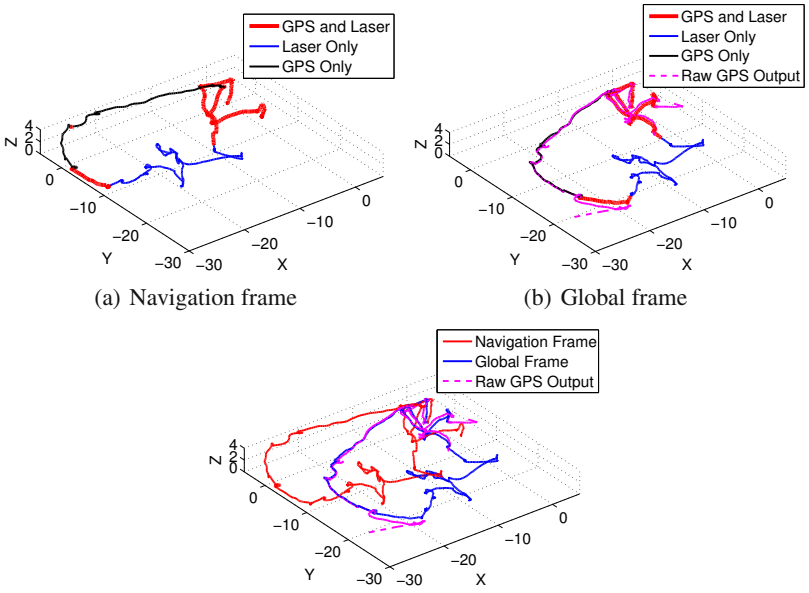
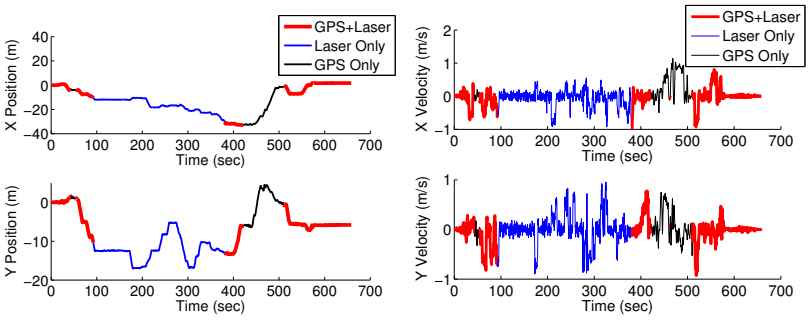


Fig. 6 Trajectory of the robot in the navigation frame (Fig. 6(a)) and the global frame (Fig. 6(b)). Figure 6(c) shows a comparison between the state estimate in the navigation and global frame.



(a) Position (b) Velocity

Fig. 7 Position and velocity estimate in the navigation frame in X-Y directions. Note that there are no discontinuities or jumps in the position estimate despite adding and removing sensors.

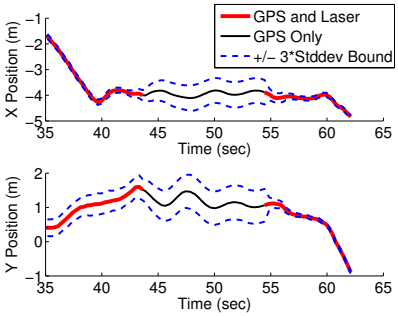
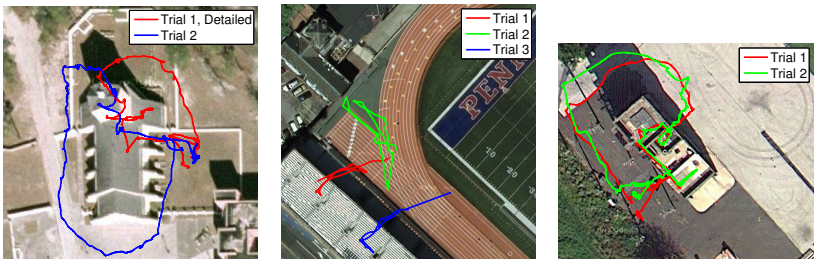


Fig. 8 A detail of the position estimate with associated error bounds corresponding to Fig. 7



(a) UO Site (b) Franklin Field (c) Fire Academy

Fig. 9 Autonomous flight experiments in multiple environments. The red trajectory in Fig. 9(a) is detailed in Sect. 3.2.

verify that the there is no conflict between the estimated trajectories of the robot and the satellite imagery. All analysis of the resulting experimental data yields similar commentary to the discussion in Sect. 3.2.

4 Conclusion and Future Work

In this work, we detail a methodology for estimating the state of a micro-aerial vehicle as it transitions between different operating environments with varying applicable sensors. We propose using a separate coordinate frame, called the navigation frame, to ensure smoothness in the state estimate and provide robust handling of sensor failures. We address the challenge of maintaining consistency between local and global measurements and propose a strategy to recursively estimate the transform between different coordinate frames. We present field experiments of a MAV navigating between mixed indoor and outdoor environments using the onboard state estimate for feedback control. We close by analyzing experimental results, showing local smoothness, global consistency, as well as repeatability of the proposed approach.

We are interested in moving forward by enhancing our current methods with vision-based state estimation methods so that our system can handle more general environments where laser and GPS information may be unavailable or unusable. We are also interested in adapting our current map alignment and loop closure methods to incorporate information from multiple robots.

References

1. Shen, S., Michael, N., Kumar, V.: Autonomous multi-floor indoor navigation with a computationally constrained MAV. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Shanghai, China, pp. 20–25 (May 2011)
2. Julier, S.J., Uhlmann, J.K.: A new extension of the kalman filter to nonlinear systems. In: Kadar, I. (ed.) Proc. of SPIE, vol. 3068, pp. 182–193 (July 1997)
3. Merwe, R.V.D., Wan, E.A., Julier, S.I.: Sigma-point kalman filters for nonlinear estimation: Applications to integrated navigation. In: Proc. of AIAA Guidance, Navigation, and Controls Conf., Providence, RI (August 2004)
4. Grzonka, S., Grisetti, G., Burgard, W.: Towards a navigation system for autonomous indoor flying. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Kobe, Japan, pp. 2878–2883 (May 2009)
5. Bachrach, A.G.: Autonomous flight in unstructured and unknown indoor environments. Master's thesis, MIT, Cambridge, MA (September 2009)
6. Blösch, M., Weiss, S., Scaramuzza, D., Siegwart, R.: Vision based MAV navigation in unknown and unstructured environments. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Anchorage, AK, pp. 21–28 (May 2010)
7. Carlson, J.: Mapping large urban environments with GPS-aided SLAM. Ph.D. dissertation, CMU, Pittsburgh, PA (July 2010)
8. Schleicher, D., Bergasa, L.M., Ocaña, M., Barea, R., López, E.: Real-time hierarchical GPS aided visual SLAM on urban environments. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Kobe, Japan, pp. 4381–4386 (May 2009)

9. Moore, D.C., Huang, A.S., Walter, M., Olson, E.: Simultaneous local and global state estimation for robotic navigation. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Kobe, Japan, pp. 3794–3799 (May 2009)
10. Weinmann, A.: Uncertain Models and Robust Control. Springer, New York (1991)
11. Bosse, M., Newman, P., Leonard, J., Soika, M., Feiten, W., Teller, S.: An atlas framework for scalable mapping. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Taipei, Taiwan, vol. 2, pp. 1899–1906 (September 2003)
12. Estrada, C., Neira, J., Tardos, J.D.: Hierarchical SLAM: Real-time accurate mapping of large environments. IEEE Trans. Robot. 21(4), 588–596 (2005)
13. Dellaert, F., Kaess, M.: Square root SAM: Simultaneous localization and mapping via square root information smoothing. Intl. J. Robot. Research 25(12), 1181–1203 (2006)
14. Kaess, M., Ranganathan, A., Dellaert, F.: iSAM: Incremental smoothing and mapping. IEEE Trans. Robot. 24(6), 1365–1378 (2008)
15. Censi, A.: On achievable accuracy for pose tracking. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Kobe, Japan, pp. 1–7 (May 2009)
16. Smith, R., Self, M., Cheeseman, P.: Estimating uncertain spatial relationships in robotics. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Rayleigh, NC, vol. 4, p. 850 (March 1987)
17. Dryanovski, I., Morris, W., Jizhong, X.: Multi-volume occupancy grids: An efficient probabilistic 3d mapping model for micro aerial vehicles. In: Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots and Syst., Taipei, Taiwan, pp. 1553–1559 (October 2010)
18. Olson, E.: Real-time correlative scan matching. In: Proc. of the IEEE Intl. Conf. on Robot. and Autom., Kobe, Japan, pp. 4387–4393 (May 2009)
19. Ascending Technologies, GmbH (February 2012), `http://www.asctec.de/`
20. Robot Operating System (February 2012), `http://www.ros.org/`