Multi-Sensor Fusion for Robust Autonomous Flight in Indoor and Outdoor Environments with a Rotorcraft MAV

Shaojie Shen, Yash Mulgaonkar, Nathan Michael, and Vijay Kumar

Abstract— We present a modular and extensible approach to integrate noisy measurements from multiple heterogeneous sensors that yield either absolute or relative observations at different and varying time intervals, and to provide smooth and globally consistent estimates of position in real time for autonomous flight. We describe the development of algorithms and software architecture for a new $1.9\,\mathrm{kg}$ MAV platform equipped with an IMU, laser scanner, stereo cameras, pressure altimeter, magnetometer, and a GPS receiver, in which the state estimation and control are performed onboard on an Intel NUC 3rd generation i3 processor. We illustrate the robustness of our framework in large-scale, indoor-outdoor autonomous aerial navigation experiments involving traversals of over 440 meters at average speeds of 1.5 m/s with winds around 10 mph while entering and exiting buildings.

Fig. 1. Our 1.9 kg MAV platform equipped with an IMU, laser scanner, stereo cameras, pressure altimeter, magnetometer, and GPS receiver. All the computation is performed onboard on an Intel NUC computer with 3rd generation i3 processor.

I. INTRODUCTION

Rotorcraft micro aerial vehicles (MAVs) are ideal platforms for surveillance and search and rescue in confined indoor and outdoor environments due to their small size, superior mobility, and hover capability. In such missions, it is essential that the MAV is capable of autonomous flight to minimize operator workload. Robust state estimation is critical to autonomous flight especially because of the inherently fast dynamics of MAVs. Due to cost and payload constraints, most MAVs are equipped with low cost *proprioceptive* sensors (e.g. MEMS IMUs) that are incapable for long term state estimation. As such, *exteroceptive* sensors, such as GPS, cameras, and laser scanners, are usually fused with proprioceptive sensors to improve estimation accuracy.

Besides the well-developed GPS-based navigation technology [1, 2]. There is recent literature on robust state estimation for autonomous flight in GPS-denied environments using laser scanners [3, 4], monocular camera [5, 6], stereo cameras [7, 8], and RGB-D sensors [9]. However, all these approaches rely on a single exteroceptive sensing modality that is only functional under certain environment conditions. For example, laser-based approaches require structured environments, vision based approaches demand sufficient lighting and features, and GPS only works outdoors. This makes them prone to failure in large-scale environments involving indoor-outdoor transitions, in which the environment can change

significantly. It is clear that in such scenarios, multiple measurements from GPS, cameras, and lasers may be available, and the fusion of all these measurements yields increased estimator accuracy and robustness. In practice, however, this extra information is either ignored or handled as switching between sensor suites [10].

The main goal of this work is to develop a modular and extensible approach to integrate noisy measurements from multiple heterogeneous sensors that yield either absolute or relative observations at different and varying time intervals, and to provide smooth and globally consistent estimates of position in real time for autonomous flight. The first key contribution, that is central to our work, is a principled approach, building on [11], to fusing relative measurements by augmenting the vehicle state with copies of previous states to create an augmented state vector for which consistent estimates are obtained and maintained using a filtering framework. A second significant contribution is our Unscented Kalman Filter (UKF) formulation in which the propagation and update steps circumvent the difficulties that result from the semi-definiteness of the covariance matrix for the augmented state. Finally, we demonstrate results with a new experimental platform (Fig. 1) to illustrate the robustness of our framework in large-scale, indoor-outdoor autonomous aerial navigation experiments involving traversals of over 440 meters at average speeds of 1.5 m/s with winds around 10 mph while entering and exiting two buildings.

Next, we present previous work on which our work is based. In Section III, we outline the modeling framework before presenting the key contributions of UKF-based sensor fusion scheme in Section IV. We bring all the ideas together in our description of the experimental platform and the experimental results in Section VI.

S. Shen, Y. Mulgaonkar, and V. Kumar are with the GRASP Laboratory, University of Pennsylvania, Philadelphia, PA 19104, USA. {shaojie, yashm, kumar}@grasp.upenn.edu

N. Michael is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA. nmichael@cmu.edu

II. PREVIOUS WORK

We are interested in applying constant computation complexity filtering-based approaches, such as nonlinear variants of the Kalman filter, to fuse all available sensor information. We stress that although SLAM-based multi-sensor fusion approaches [12, 13] yield optimal results, they are computationally expensive for real-time state feedback for the purpose of autonomous control.

While it is straightforward to fuse multiple *absolute* measurements such as GPS, pressure/laser altimeter in a recursive filtering formulation, the fusion of multiple *relative* measurements obtained from laser or visual odometry are more involved. It is common to accumulate the relative measurements with the previous state estimates fuse them as pseudo-absolute measurements [5, 14]. However, such fusion is sub-optimal since the resulting global position and yaw covariance is inconsistently small compared to the actual estimation error. This violates the observability properties [6], which suggests that such global quantities are in fact unobservable. As such, we develop our method based on state augmentation techniques [11] to properly account for the state uncertainty when applying multiple relative measurements from multiple sensors.

We aim to develop a modular framework that allows easy addition and removal of sensors with minimum coding and mathematical derivation. We note that in the popular EKF-based formulation [5, 8], the computation of Jacobians can be problematic for complex systems like MAVs. As such, we employ a loosely coupled, derivative-free Unscented Kalman Filter (UKF) framework [1]. Switching from EKF to UKF poses several challenges, which will be detailed and addressed in Sect. IV-A. [15] is similar to our work. However, the EKF-based estimator in [15] does not support fusion of multiple relative measurements.

III. MULTI-SENSOR SYSTEM MODEL

We define vectors in the world and body frames as $(\cdot)^w$ and $(\cdot)^b$ respectively. For the sake of brevity, we assume that all onboard sensors are calibrated and are attached to the body frame. The *main* states of the MAV is defined as:

$$\mathbf{x} = \left[\mathbf{p}^w, \; \boldsymbol{\Phi}^w, \; \dot{\mathbf{p}}^b, \; \mathbf{b}_a^b, \; \mathbf{b}_\omega^b, \; \mathbf{b}_z^w \right]^{\mathrm{T}}$$

where $\mathbf{p}^w = [x^w, \; y^w, \; z^w]^{\mathrm{T}}$ is the 3D position in the world frame, $\boldsymbol{\Phi}^w = [\psi^w, \; \theta^w, \; \phi^w]^{\mathrm{T}}$ is the yaw, pitch, and roll Euler angles that represent the 3-D orientation of the body in the world frame [1], from which a matrix $\mathbf{R}_b^w$ that represent the rotation of a vector from the body frame to the world frame can be obtained. $\dot{\mathbf{p}}^b$ is the 3D velocity in the body frame. $\mathbf{b}_a^b$ and $\mathbf{b}_\omega^b$ are the bias of the accelerometer and gyroscope, both expressed in the body frame. $\mathbf{b}_z^w$ models the bias of the laser and/or pressure altimeter in the world frame.

[1] It is straightforward to formulate the filter with quaternion-based rotation representation [8, 15], We present the direct formulation for the brevity of presentation in Sect. IV.

We consider an IMU-based state propagation model:

$$\mathbf{u}_t = \left[\mathbf{a}^b, \; \omega^b \right]^{\mathrm{T}}$$
$$\mathbf{v}_t = [\mathbf{v}_a, \; \mathbf{v}_\omega, \; \mathbf{v}_{\mathbf{b}_a}, \; \mathbf{v}_{\mathbf{b}_\omega}, \; \mathbf{v}_{\mathbf{b}_z}]^{\mathrm{T}} \tag{1}$$
$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \; \mathbf{u}_t, \; \mathbf{v}_t)$$

where $\mathbf{u}$ is the measurement of linear accelerations and angular velocities from the IMU in the body frame. $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t) \in \mathbb{R}^m$ is the process noise. $\mathbf{v}_a$ and $\mathbf{v}_\omega$ represent additive noise associated with the gyroscope and the accelerometer. $\mathbf{v}_{\mathbf{b}_a}$, $\mathbf{v}_{\mathbf{b}_\omega}$, $\mathbf{v}_{\mathbf{b}_z}$ model the Gaussian random walk of the gyroscope, accelerometer and altimeter bias. The function $f(\cdot)$ is a discretized version of the continuous time dynamical equation [6].

Exteroceptive sensors are usually used to correct the errors in the state propagation. Following [11], we consider measurements as either being *absolute* or *relative*, depending on the nature of the underlying sensor. We allow an arbitrary number of either absolute or relative measurement models.

A. Absolute Measurements

All absolute measurements can be modeled in the form:

$$\mathbf{z}_{t+m} = h_a(\mathbf{x}_{t+m}, \mathbf{n}_{t+m}) \tag{2}$$

where $\mathbf{n}_{t+m} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t) \in \mathbb{R}^p$ is the measurement noise that can be either additive or not. $h_a(\cdot)$ is in general a nonlinear function. An absolute measurement connects the *current* state with the sensor output. Examples are shown in in Sect. V-B.

B. Relative Measurements

A relative measurement connects the *current* and the *past* states with the sensor output, which can be written as:

$$\mathbf{z}_{t+m} = h_r(\mathbf{x}_{t+m}, \mathbf{x}_t, \mathbf{n}_{t+m}) \tag{3}$$

The formulation accurately models the nature of odometry-like algorithms (Sect. V-C and Sect. V-D) as odometry measures the incremental changes between two time instants of the state. We also note that, in order to avoid temporal drifting, most state-of-the-art laser/visual odometry algorithms are *keyframe* based. As such, we allow multiple future measurement ($m \in \mathcal{M}$, $|\mathcal{M}| > 1$) that corresponds to the same past state $\mathbf{x}_t$.

IV. UKF-BASED MULTI-SENSOR FUSION

We wish to design a modular sensor-fusion filter that is easily extensible even for inexperienced users. This means that amount of coding and mathematical deviation for the addition/removal of sensors should be minimal. One disadvantage of the popular EKF-based filtering framework is the requirement of computing the Jacobian matrices, which is proven to be difficult and time consuming for a complex MAV system. As such, we employ the derivative-free UKF-based approach [1]. The key of UKF is the approximation of the propagation of Gaussian random vectors through nonlinear functions via the propagation of sigma points. Let $\mathbf{x} \sim \mathcal{N}(\hat{\mathbf{x}}, \mathbf{P^{xx}}) \in \mathbb{R}^n$ and consider the nonlinear function:

$$\mathbf{y} = g(\mathbf{x}), \tag{4}$$

and let:

$$\mathcal{X} = \left[\hat{\mathbf{x}}, \ \hat{\mathbf{x}} \pm \left(\sqrt{(n+\lambda)\mathbf{P}^{\mathbf{xx}}} \right)_i \right] \quad \text{for } i = 1, ..., n$$
$$\mathcal{Y}_i = g(\mathcal{X}_i), \tag{5}$$

where $g(\cdot)$ is a nonlinear function, λ is a UKF parameter. $\left(\sqrt{(n+\lambda)\mathbf{P}^{\mathbf{xx}}} \right)_i$ is the ith column of the square root covariance matrix, which is usually computed via Cholesky decomposition. And $\mathcal{X}$ are called the sigma points. The mean, covariance of the random vector $\mathbf{y}$, and the cross-covariance between $\mathbf{x}$ and $\mathbf{y}$, can be approximated as:

$$\hat{\mathbf{y}} = \sum_{i=0}^{2n} w_i^m \mathcal{Y}_i$$

$$\mathbf{P}^{\mathbf{yy}} = \sum_{i=0}^{2n} w_i^c (\mathcal{Y}_i - \hat{\mathbf{y}})(\mathcal{Y}_i - \hat{\mathbf{y}})^{\mathrm{T}} \tag{6}$$

$$\mathbf{P}^{\mathbf{yx}} = \sum_{i=0}^{2n} w_i^c (\mathcal{Y}_i - \hat{\mathbf{y}})(\mathcal{X}_i - \hat{\mathbf{x}})^{\mathrm{T}}$$

where w_i^m and w_i^c are weights for the sigma points. This unscented transform can be used to keep track of the covariance in both the state propagation and measurement update, thus avoiding the need of a Jacobian-based covariance approximation.

A. State Augmentation for Multiple Relative Measurements

Since a relative measurement depends both the *current* and *past* states, it is a violation of the fundamental assumption in the Kalman filter that the measurement should *only* depend on the *current* state. One way to deal with this is through state augmentation [11], where a copy of the past state is maintained in the filter. Here we present an extension of [11] to handle arbitrary number of relative measurement models with the possiblity that multiple measurements correspond to the same augmented state. Our generic filtering framework allows convenient setup, and facilitates addition and removal of absolute and relative measurement models.

Note that a measurement may not affect all components in the state $\mathbf{x}$. For example, a visual odometry only affects the 6-DOF pose, not the velocity or the bias terms. We define the ith augmented state as $\mathbf{x}_i \in \mathbb{R}^{n_i}$, $n_i \leq n$. $\mathbf{x}_i$ is an arbitrary subset of $\mathbf{x}$. We define a binary selection matrix $\mathbf{B}_i$ of size $n_i \times n$, such that $\mathbf{x}_i = \mathbf{B}_i \mathbf{x}$. Consider a time instant, there are I augmented states in the filter, along with the covariance:

$$\check{\mathbf{x}} = [\hat{\mathbf{x}}, \ \hat{\mathbf{x}}_1, \ ..., \ \hat{\mathbf{x}}_I]^{\mathrm{T}}$$

$$\check{\mathbf{P}} = \begin{bmatrix} \mathbf{P}^{\mathbf{xx}} & \mathbf{P}^{\mathbf{xx}_1} & \cdots & \mathbf{P}^{\mathbf{xx}_I} \\ \mathbf{P}^{\mathbf{x}_1\mathbf{x}} & \mathbf{P}^{\mathbf{x}_1\mathbf{x}_1} & \cdots & \mathbf{P}^{\mathbf{x}_1\mathbf{x}_I} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}^{\mathbf{x}_I\mathbf{x}} & \mathbf{P}^{\mathbf{x}_I\mathbf{x}_1} & \cdots & \mathbf{P}^{\mathbf{x}_I\mathbf{x}_I} \end{bmatrix}. \tag{7}$$

The addition of a new augmented state $\mathbf{x}_{I+1}$ can be done by:

$$\check{\mathbf{x}}^+ = \mathbf{M}^+ \check{\mathbf{x}}, \quad \mathbf{M}^+ = \begin{bmatrix} \mathbf{I}_{n+\sum_I n_i} \\ \mathbf{B}_{I+1} \end{bmatrix} \tag{8}$$

Similarly, the removal of an augmented state $\mathbf{x}_j$ is given as:

$$\check{\mathbf{x}}^- = \mathbf{M}^- \check{\mathbf{x}}, \quad \mathbf{M}^- = \begin{bmatrix} \mathbf{I}_a & \mathbf{0}_{a \times n_j} & \mathbf{0}_{a \times b} \\ \mathbf{0}_{b \times n} & \mathbf{0}_{b \times n_j} & \mathbf{I}_b \end{bmatrix},$$

where $a = n + \sum_{i=1}^{j-1} n_i$ and $b = \sum_{i=j+1}^{I} n_i$. The updated augmented state covariance is given as:

$$\check{\mathbf{P}}^{\pm} = \mathbf{M}^{\pm} \check{\mathbf{P}} \mathbf{M}^{\pm\mathrm{T}}.$$

The change of keyframes in an odometry-like measurement model is simply the removal of an augmented state $\mathbf{x}_i$ followed by the addition of another augmented state with the same $\mathbf{B}_i$. Since we allow multiple relative measurements that correspond to the same augmented state, contrast to [11], augmented states are *not* deleted after measurement updates (Sect. IV-D).

This state augmentation formulation works well in an EKF setting, however, it poses issues when we try to apply it to the UKF. Since the addition of a new augmented state (8) is essentially a copy of the *main* state. The resulting covariance matrix $\check{\mathbf{P}}^+$ will not be positive definite, and the Cholesky decomposition (5) for state propagation will fail (non-unique). We now wish to have something that is similar to the Jacobian matrices for EKF, but *without* explicitly computing the Jacobians.

B. Jacobians for UKF

In [16], the authors present a new interpretation of the UKF as a Linear Regression Kalman Filter (LRKF). In LRKF, we seek to find the optimal linear approximation $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} + \mathbf{e}$ of the nonlinear function (4) given a weighted discrete (or sigma points (6)) representation of the distribution $\mathcal{N}(\hat{\mathbf{x}}, \mathbf{P}^{\mathbf{xx}})$. The objective is to find the regression matrix $\mathbf{A}$ and vector $\mathbf{b}$ that minimize the linearizion error $\mathbf{e}$:

$$\min_{\mathbf{A},\mathbf{b}} \sum_{i=0}^{2n} w_i (\mathcal{Y}_i - \mathbf{A}\mathcal{X}_i - \mathbf{b})(\mathcal{Y}_i - \mathbf{A}\mathcal{X}_i - \mathbf{b})^{\mathrm{T}}.$$

As shown in [16], the optimal linear regression is given by:

$$\mathbf{A} = \mathbf{P}^{\mathbf{yx}} \mathbf{P}^{\mathbf{xx}^{-1}}, \quad \mathbf{b} = \hat{\mathbf{y}} - \mathbf{A}\hat{\mathbf{x}} \tag{9}$$

The linear regression matrix $\mathbf{A}$ in (9) serves as the linear approximation of the nonlinear fuction (4). It is similar to the Jacobian in the EKF formulation. As such, the propagation and update steps in UKF can be performed in a similar fashion as EKF.

C. State Propagation

Observing the fact that during state propagation only the *main* state changes, we start off by partitioning the augmented state and the covariance (7) into:

$$\check{\mathbf{x}}_{t|t} = \begin{bmatrix} \hat{\mathbf{x}}_{t|t} \\ \hat{\mathbf{x}}_{\mathcal{I}|t} \end{bmatrix}, \quad \check{\mathbf{P}}_{t|t} = \begin{bmatrix} \mathbf{P}^{\mathbf{xx}}_{t|t} & \mathbf{P}^{\mathbf{xx}_{\mathcal{I}}}_{t|t} \\ \mathbf{P}^{\mathbf{x}_{\mathcal{I}}\mathbf{x}}_{t|t} & \mathbf{P}^{\mathbf{x}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}}_{t|t} \end{bmatrix}.$$

The linear approximation of the nonlinear state propagation (1), applied on the augmented state (7), is:

$$\check{\mathbf{x}}_{t+1|t} = f(\check{\mathbf{x}}_{t|t}, \mathbf{u}_t, \mathbf{v}_t)$$
$$= \begin{bmatrix} \mathbf{F}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{|\mathcal{I}|} \end{bmatrix} \check{\mathbf{x}}_{t|t} + \begin{bmatrix} \mathbf{J}_t & \mathbf{G}_t \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix} + \mathbf{b}_t + \mathbf{e}_t, \tag{10}$$

from which we can see that the propagation of the full augmented state is actually unnecessary since the only nontrivial regression matrix corresponds to the main state. We can propagate only the main state $\mathbf{x}$ via sigma points generated from $\mathbf{P}_{t|t}^{\mathbf{xx}}$ and use the UKF Jacobian $\mathbf{F}_t$ to update the cross-covariance $\mathbf{P}_{t|t}^{\mathbf{xx}_{\mathcal{I}}}$. Since the covariance matrix of the main state $\mathbf{P}_{t|t}^{\mathbf{xx}}$ is always positive definite, we avoid the Cholesky decomposition failure problem.

Since the process noise is not additive, we augment the main state with the process noise and generate sigma points from:

$$\bar{\mathbf{x}}_{t|t} = \begin{bmatrix} \hat{\mathbf{x}}_{t|t} \\ \mathbf{0} \end{bmatrix}, \quad \bar{\mathbf{P}}_{t|t} = \begin{bmatrix} \mathbf{P}_{t|t}^{\mathbf{xx}} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_t \end{bmatrix}. \tag{11}$$

The state is then propagated forward by substituting (11) into (1), (5) and (6). We obtain $\hat{\mathbf{x}}_{t+1|t}$, the estimated value of $\mathbf{x}$ at time $t+1$ given the measurements up to t, as well as $\mathbf{P}_{t+1|t}^{\mathbf{xx}}$ and $\mathbf{P}_{t+1|t}^{\mathbf{x}\bar{\mathbf{x}}}$. Following (9), we know that:

$$\mathbf{P}_{t+1|t}^{\mathbf{x}\bar{\mathbf{x}}} \bar{\mathbf{P}}_{t|t}^{-1} = [\mathbf{F}_t, \mathbf{G}_t].$$

The propagated augmented state and its covariance is updated according to (10):

$$\check{\mathbf{x}}_{t+1|t} = \begin{bmatrix} \hat{\mathbf{x}}_{t+1|t} \\ \hat{\mathbf{x}}_{\mathcal{I}|t} \end{bmatrix}, \quad \check{\mathbf{P}}_{t+1|t} = \begin{bmatrix} \mathbf{P}_{t+1|t}^{\mathbf{xx}} & \mathbf{F}_t \mathbf{P}_{t|t}^{\mathbf{xx}_{\mathcal{I}}} \\ \mathbf{P}_{t|t}^{\mathbf{x}_{\mathcal{I}}\mathbf{x}} \mathbf{F}_t^{\mathrm{T}} & \mathbf{P}_{t|t}^{\mathbf{x}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}} \end{bmatrix}. \tag{12}$$

D. Measurement Update

Let there be m state propagations between two measurements, and we maintain $\check{\mathbf{x}}_{t+m|t}$ and $\check{\mathbf{P}}_{t+m|t}$ as the newest measurement arrives. Consider a relative measurement (3) that depends on the jth augmented state, the measurement prediction and its linear regression approximation can be written as:

$$\hat{\mathbf{z}}_{t+m|t} = h_r(\hat{\mathbf{x}}_{t+m|t}, \mathbf{B}_j^{\mathrm{T}} \hat{\mathbf{x}}_{j_{t+m|t}}, \mathbf{n}_{t+m})$$
$$= \mathbf{H}_{t+m|t} \check{\mathbf{x}}_{t+m|t} + \mathbf{L}_{t+m} \mathbf{n}_{t+m} + \mathbf{b}_{t+m} + \mathbf{e}_{t+m}$$
$$\mathbf{H}_{t+m|t} = \begin{bmatrix} \mathbf{H}_{t+m|t}^{\mathbf{x}}, \mathbf{0}, \mathbf{H}_{t+m|t}^{\mathbf{x}_j}, \mathbf{0} \end{bmatrix}.$$

Again, since only the main state and one augmented state are involved in each measurement update, we can construct another augmented state together with the possibly non-additive measurement noise:

$$\dot{\mathbf{x}}_{t+m|t} = \begin{bmatrix} \hat{\mathbf{x}}_{t+m|t} \\ \hat{\mathbf{x}}_{j_{t+m|t}} \\ \mathbf{0} \end{bmatrix}, \dot{\mathbf{P}}_{t+m|t} = \begin{bmatrix} \mathbf{P}_{t+m|t}^{\mathbf{xx}} & \mathbf{P}_{t+m|t}^{\mathbf{xx}_j} & \mathbf{0} \\ \mathbf{P}_{t+m|t}^{\mathbf{x}_j\mathbf{x}} & \mathbf{P}_{t+m|t}^{\mathbf{x}_j\mathbf{x}_j} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_{t+m} \end{bmatrix}.$$

After the state propagation (12), $\dot{\mathbf{P}}_{t+m|t}$ is guaranteed to be positive definite, thus it is safe to perform sigma point

propagation as in (5) and (6). We obtain $\hat{\mathbf{z}}_{t+m|t}$, $\mathbf{P}_{t+m|t}^{\mathbf{zz}}$, $\mathbf{P}_{t+m|t}^{\mathbf{z}\dot{\mathbf{x}}}$, and:

$$\mathbf{P}_{t+m|t}^{\mathbf{z}\dot{\mathbf{x}}} \dot{\mathbf{P}}_{t+m|t}^{-1} = \begin{bmatrix} \mathbf{H}_{t+m|t}^{\mathbf{x}}, \mathbf{H}_{t+m|t}^{\mathbf{x}_j}, \mathbf{L}_{t+m} \end{bmatrix}.$$

We can apply the measurement update similar to an EKF:

$$\check{\mathbf{K}}_{t+m} = \check{\mathbf{P}}_{t+m|t} \mathbf{H}_{t+m|t}^{\mathrm{T}} \mathbf{P}_{t+m|t}^{\mathbf{zz}^{-1}}$$
$$\check{\mathbf{x}}_{t+m|t+m} = \check{\mathbf{x}}_{t+m|t} + \check{\mathbf{K}}_{t+m} (\mathbf{z}_{t+m} - \hat{\mathbf{z}}_{t+m|t})$$
$$\check{\mathbf{P}}_{t+m|t+m} = \check{\mathbf{P}}_{t+m|t} - \check{\mathbf{K}}_{t+m} \mathbf{H}_{t+m|t} \check{\mathbf{P}}_{t+m|t},$$

where $\mathbf{z}_{t+m}$ is the actual sensor measurement. Both the main and augmented states will be corrected during measurement update. We note that entries in $\mathbf{H}_{t+m|t}$ that correspond to inactive augmented states are zero. This can be utilized to speed up the matrix multiplication.

The fusion of absolute measurements can simply be done by setting $\hat{\mathbf{x}}_{j_{t+m|t}} = \emptyset$ and applying the corresponding absolute measurement model (2).

As shown in Fig. 9, fusion of multiple relative measurements results in slow growing, but *unbounded* covariance in the global position and yaw. This is consistent with results in [6] that these global quantities are unobservable.

E. Delayed, Out-of-Order Measurement Update

When fusing multiple measurements, it is possible that the measurements arrive out-of-order to the filter, that is, a measurement that corresponds to an *earlier* state arrives *after* the measurement that corresponds to a *later* state. This violates the Markov assumption of the Kalman filter. Also, due to the sensor processing delay, measurements may lag behind the state propagation.

We address these two issues by storing measurements in a priority queue, where the top of the queue corresponds to the oldest measurement. A pre-defined a maximum allowable sensor delay t_d of $100\,\mathrm{ms}$ was set for our MAV platform. Newly arrived measurements that correspond to a state older than t_d from the current state (generated by state propagation) are directly discarded. After each state propagation, we check the queue and process all measurements in the queue that are older than t_d. The priority queue essentially serves as a measurement reordering mechanism (Fig. 2) for all measurements that are not older than t_d from the current state. In the filter, we always utilize the most recent IMU measurement to propagate the state forward. We, however, only propagate the covariance on demand. As illustrated in Fig. 2, the covariance is only propagated from the time of the last measurement to the current measurement.

F. An Alternative Way for Handling Global Pose Measurements

As the vehicle moves through the environment, global pose measurements from GPS and magnetometer may be available. It is straightforward to fuse the GPS as a global pose measurement and generate the optimal state estimate. However, this may not be the best for real-world applications. A vehicle that operates in a GPS-denied environment may
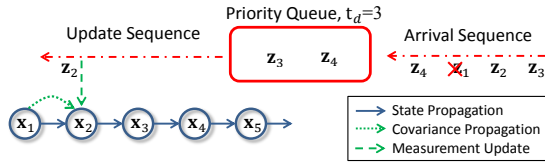
Fig. 2. Delayed, out-of-order measurement with priority queue. While $\mathbf{z}_4$ arrives before $\mathbf{z}_2$, $\mathbf{z}_2$ is first applied to the filter. $\mathbf{z}_4$ is temporary stored in the queue. $\mathbf{z}_1$ is discarded since it is older than t_d from the current state. The covariance is only propagated up to time where the most recent measurement is applied to the filter. The state is propagated till the most recent IMU input.
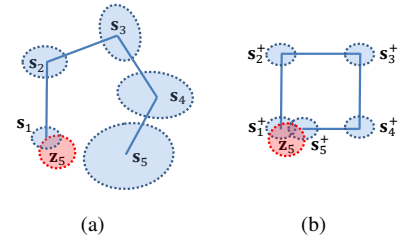

Fig. 3. In Fig. 3(a), GPS signal is regained at $k = 5$, resulting in large discrepancies between the measurement $\mathbf{z}_5$ and the state $\mathbf{s}_5$. Pose graph SLAM produces a globally consistent graph (Fig. 3(b)).
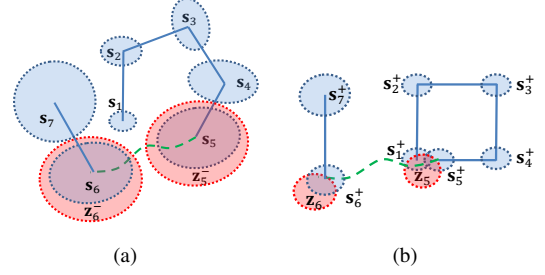

Fig. 4. Fig. 4(a) illustrates the alternative GPS fusion, the discrepancy between transformed GPS measurement $\mathbf{z}_5$ and the non-optimized state $\mathbf{s}_5$ is minimized. Fusion of such indirect GPS measurement will lead to smooth state estimate (green dashed line).

suffer from accumulated drift. When the vehicle gains GPS signal, as illustrated in Fig. 3(a), there maybe large discrepancies between the GPS measurement and the estimated state ($\mathbf{z}_5 - \mathbf{s}_5$). Directly applying GPS as global measurements will result in undesirable behaviors in both estimation (large linearizion error) and control (sudden pose change).

This is not a new problem and it has been studied for ground vehicles [17] under the term of local frame-based navigation. However, [17] assumes that a reasonably accurate local estimate of the vehicle is always available (e.g. wheel odometry). This is not the case for MAVs since the state estimates with only the onboard IMUs drifts away vastly within a few seconds. The major difference between dead reckoning with IMU and wheel odometry is that the former drifts temporally, while the latter only drifts spatially. However, we have relative exteroceptive sensors that are able to produce temporally drift-free estimates. As such, we only need to deal with the case that all relative exteroceptive sensors have failed. Therefore, our goal is to properly transform the global GPS measurement into the local frame to bridge the gap between relative sensor failures.

Consider a pose-only graph SLAM formulation with $\mathbf{s}_k = [x_k^w, y_k^w, \psi_k^w]^\mathrm{T} \in \Theta$ being 2D poses. The SLAM module may run at a much lower rate than the UKF-based estimator. We optimize the pose graph given incremental motion constraints $\mathbf{d}_k$ from laser/visual odometry, spatial loop closure constraints $\mathbf{l}_k$, and absolute pose constraints $\mathbf{z}_k$ from GPS:

$$\min_\Theta \left\{ \sum_{k=1}^M \|h_i(\mathbf{s}_{k-1}, \mathbf{d}_k) - \mathbf{s}_k\|_{\mathbf{P}_k^\mathbf{d}} \right.$$
$$\left. + \sum_{k=1}^L \|h_l(\mathbf{s}_k, \mathbf{l}_k) - \mathbf{s}_{l(k)}\|_{\mathbf{P}_k^\mathbf{l}} + \sum_{k=1}^N \|\mathbf{z}_k - \mathbf{s}_k\|_{\mathbf{P}_k^\mathbf{z}} \right\}.$$

The optimal pose graph configuration can be found with available solvers [18], as shown in Fig. 3(b). The pose graph is disconnected if there are no relative exteroceptive measurements between two nodes. Let two pose graphs be disconnected between $k - 1$ and k.

The pose graph SLAM provides the transformation between the non-optimized $\mathbf{s}_{k-1}$ and the SLAM-optimized $\mathbf{s}_{k-1}^+$ state. This transform can be utilized to transform the

global GPS measurement to be aligned with $\mathbf{s}_{k-1}$:

$$\Delta_{t-1} = \mathbf{s}_{k-1} \ominus \mathbf{s}_{k-1}^+$$
$$\mathbf{z}_{k-1}^- = \Delta_{t-1} \oplus \mathbf{z}_{k-1},$$

where $\oplus$ and $\ominus$ are pose compound operations as defined in [19]. The covariance $\mathbf{P}_{t-1}^\Delta$ of Δ_{t-1} and subsequently the covariance $\mathbf{P}_{t-1}^z$ of $\mathbf{z}_{k-1}^-$ can be computed following [19]. This formulation minimizes the discrepancies between $\mathbf{z}_{k-1}^-$ and $\mathbf{s}_{k-1}$, and thus maintains smoothness in the state estimate. The transformed GPS $\mathbf{z}_{k-1}^-$, is still applied as an absolute measurement to the UKF (Fig. 4(a)).

However, despite the large scale in our field experiments (Sect. VI), we hardly find a case where the accumulated drift is large enough to cause issues with direct GPS fusion. In the future, we will seek for even larger scale experiments to verify the necessity of the above local frame-based approach.

V. IMPLEMENTATION DETAILS

A. Experimental Platform

The experimental platform shown in Fig. 1 is based on the Pelican quadrotor from Ascending Technologies, GmbH[2]. This platform is natively equipped with an AutoPilot board consisting of an IMU and a user-programmable ARM7 microcontroller. The main computation unit onboard is an Intel NUC with a 1.8 GHz Core i3 processor with 8 GB of RAM and a 120 GB SSD. The sensor suite includes a u-blox LEA-6T GPS module, a Hokuyo UTM-30LX LiDAR and two mvBlueFOX-MLC200w grayscale HDR cameras

[2]Ascending Technologies, GmbH, http://www.asctec.de/

with fisheye lenses that capture 752×480 images at $25\,\text{Hz}$. We use hardware triggering for frame synchronization. The onboard auto exposure controller is fine tuned to enable fast adaption during rapid light condition changes. A 3-D printed laser housing redirects some of the laser beams for altitude measurement. The total mass of the platform is 1.87kg. The entire algorithm is developed in C++ using ROS[3] as the interfacing robotics middleware.

B. Absolute Measurements

Some onboard sensors are capable of producing absolute measurements (Sect. III-A), here are their details:

1) GPS And Magnetometer:

$$\mathbf{z}_t = \begin{bmatrix} \begin{pmatrix} x_t^w \\ y_t^w \end{pmatrix} \\ \mathbf{R}_b^w \begin{pmatrix} \dot{x}_t^b \\ \dot{y}_t^b \end{pmatrix} \\ \psi_t^w \end{bmatrix} + \mathbf{n}_t.$$

2) Laser/Pressure Altimeter:

$$\mathbf{z}_t = z_t^w + \mathbf{b}_{z_t}^w + \mathbf{n}_t.$$

3) Pseudo Gravity Vector: If the MAVs is near hover or moving at approximately constant speed, we may say that the accelerometer output provides a pseudo measurement of the gravity vector. Let $\mathbf{g} = [0, 0, g]^{\mathrm{T}}$, we have:

$$\mathbf{z}_t = \mathbf{R}_b^{w\mathrm{T}} \mathbf{g}^w + \mathbf{b}_{a_t}^b + \mathbf{n}_t.$$

C. Relative Measurement - Laser-Based Odometry

We utilize the laser-based odometry that we developed in our earlier work [4]. Observing that man-made indoor environments mostly contains vertical walls, we can make a 2.5-D environment assumption. With this assumption, we can make use of the onboard roll and pitch estimates to project the laser scanner onto a common ground plane. As such, 2D scan matching can be utilized to estimate the incremental horizontal motion of the vehicle. We keep a local map to avoid drifting while hovering.

$$\mathbf{z}_{t+m} = \ominus_{2d} \begin{bmatrix} x_t^w \\ y_t^w \\ \psi_t^w \end{bmatrix} \oplus_{2d} \begin{bmatrix} x_{t+m}^w \\ y_{t+m}^w \\ \psi_{t+m}^w \end{bmatrix} + \mathbf{n}_{t+m},$$

where $\mathbf{p}_{2d_t} = [x_t^w, y_t^w, \psi_t^w]^{\mathrm{T}}$, $\oplus_{2d}$ and $\ominus_{2d}$ are the 2-D pose compound operations as defined in [19].

D. Relative Measurement - Visual Odometry

We implemented a classic keyframe-based visual odometry algorithm. Keyframe-based approaches have the benefit of being temporally drift-free. We choose to use light-weight corner features but run the algorithm at a high-rate ($25\,\text{Hz}$). Features are tracked across images via KLT tracker. Given a keyframe with a set of triangulated feature points, we run a robust iterative 2D-3D pose estimation [8] to estimate the 6-DOF motion of the vehicle with respect to the keyframe. New
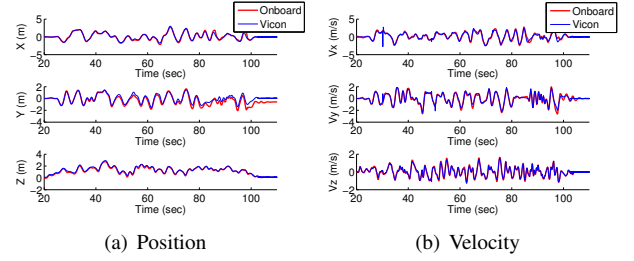
---

[3]Robot Operating System, http://www.ros.org/

(a) Position   (b) Velocity

Fig. 5. The MAV maneuvers aggressively with a maximum speed of $3.5\,\text{m/s}$ (Fig. 5(b)). The horizontal position also compares well with the ground truth with slight drift (Fig. 5(a)).

keyframes are inserted depending on the distance traveled and the current number of valid 3D points.

$$\mathbf{z}_{t+m} = \ominus \begin{bmatrix} \mathbf{p}_t^w \\ \boldsymbol{\Phi}_t^w \end{bmatrix} \oplus \begin{bmatrix} \mathbf{p}_{t+m}^w \\ \boldsymbol{\Phi}_{t+m}^w \end{bmatrix} + \mathbf{n}_{t+m}$$

E. Feedback Control

To achieve stable flight across different environments with possibly large orientation changes, we choose to use a position tracking controller with a nonlinear error metric [20]. The $100\,\text{Hz}$ filter output (Sect. IV) is used directly as the feedback for the controller. In our implementation, the attitude controller runs at 1 kHz on the ARM processor on the MAV's AutoPilot board, while the position tracking control operates at 100 Hz on the main computer. We implemented both setpoint trajectory tracking and velocity control to allow flexible operations.

VI. EXPERIMENTAL RESULTS

Multiple experiments are conducted to demonstrate the robustness of our system. We begin with an quantitative evaluation in a lab environment equipped with a motion capture systems. We then test our system in two real-world autonomous flight experiments, including an industrial complex and a tree-lined campus.

A. Evaluation of Estimator Performance

We would like to push the limits of our onboard estimator. Therefore, we have a professional pilot to aggressively fly the quadrotor with a $3.5\,\text{m/s}$ maximum speed and large attitude of up to $40°$. The onboard state estimates are compared the ground truth from the motion capture system. Since there is no GPS measurement indoor, our system relies on a fusion of relative measurements from laser and vision. We do observe occasional laser failure due to large attitude violating the 2.5-D assumption (Sect. V-C). However, the multi-sensor filter still tracks the vehicle state throughout (Fig. 5). We do not quantify the absolute pose error since it is unbounded. However, the body frame velocity (Fig. 5(b)) compares well with the ground truth with standard deviations of $\{0.1021, 0.1185, 0.0755\}^{\mathrm{T}}$ (m/s) in x, y, and z, respectively.
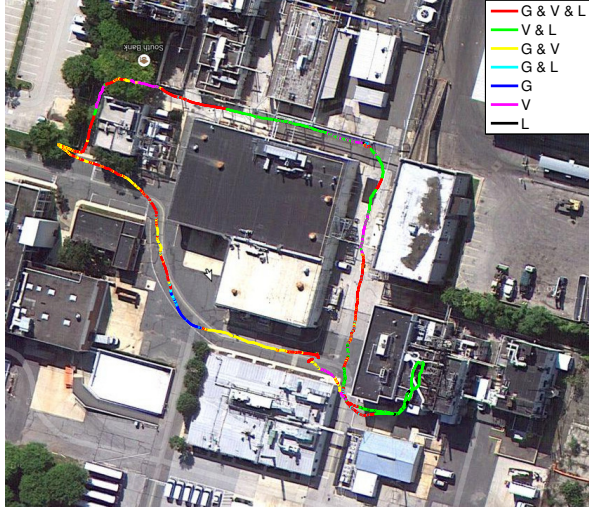
Fig. 7. Vehicle trajectory aligned with satellite imagery. Different colors indicate different combinations of sensing modalities. G=GPS, V=Vision, and L=Laser.
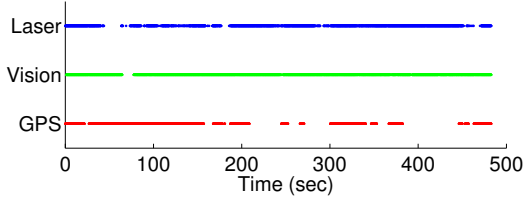

Fig. 8. Sensor availability over time. Note that failures occurred to all sensors. This shows that multi-sensor fusion is a must for this kind of indoor-outdoor missions.

B. Autonomous Flight in Large-Scale Indoor and Outdoor Environments

We tested our system in a challenging industrial complex. The testing site spans a variety of environments, including outdoor open space, densely filled trees, cluttered building area, and indoor environments (Fig. 6). The MAV is autonomously controlled using the onboard state estimates. However, a human operator always has the option of sending high level waypoints or velocity commands to the vehicle. The total flight time is approximately 8 minutes, and the vehicle travels 445 meters with an average speed of 1.5 m/s. As shown in the map-aligned trajectory (Fig. 7), during the experiment, frequent sensor failures occurred (Fig. 8), indicating the necessity of multi-sensor fusion. Fig. 9 shows the evolution of covariance as the vehicle flies through a GPS shadowing area. The global x, y and yaw error is bounded by GPS measurement, without which the error will grow unbounded. This matches the observability analysis results. It should be noted that the error on body frame velocity does not grow, regardless of the availability of GPS. The spike in velocity covariance in Fig. 9 is due to the camera facing direct sunlight.
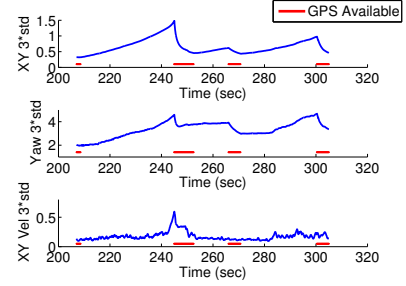

Fig. 9. Covariance changes as the vehicle flies through a dense building area (between 200s - 300s, top of Fig. 7, green line). The GPS comes in and out due to building shadowing. The covariance of x, y, and yaw increases as GPS fails and decreases as GPS resumes. Note that the body frame velocity are observable regardless of GPS measurements, and thus its covariance remains small. The spike in the velocity covariance is due to the vehicle directly facing the sun. The X-Y covariance is calculated from the Frobenius norm of the covariance submatrix.
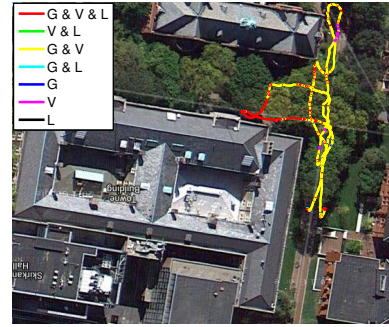

Fig. 10. Vehicle trajectory overlaid on a satellite map. The vehicle operates in a tree-lined campus environment, where there is high risk of GPS failure during operation.

C. Autonomous Flight in Tree-Lined Campus

We also conduct experiments in a tree-lined campus environment, as shown in Fig. 10. Autonomous flight in this environment is challenging due to nontrivial light condition changes as the vehicle moves in and out of tree shadows. The risk of GPS failure is also very high due to the trees above the vehicle. Laser-based odometry only works when close to buildings. The total trajectory length is 281 meters.

VII. CONCLUSION AND FUTURE WORK

In this work, we present a modular and extensible approach to integrate noisy measurements from multiple het-

(a)   (b)

Fig. 11. Onboard (Fig. 11(a)) and external (Fig. 11(b)) camera images as the MAV autonomously flies through a tree-lined campus environment. Note the nontrivial light condition.
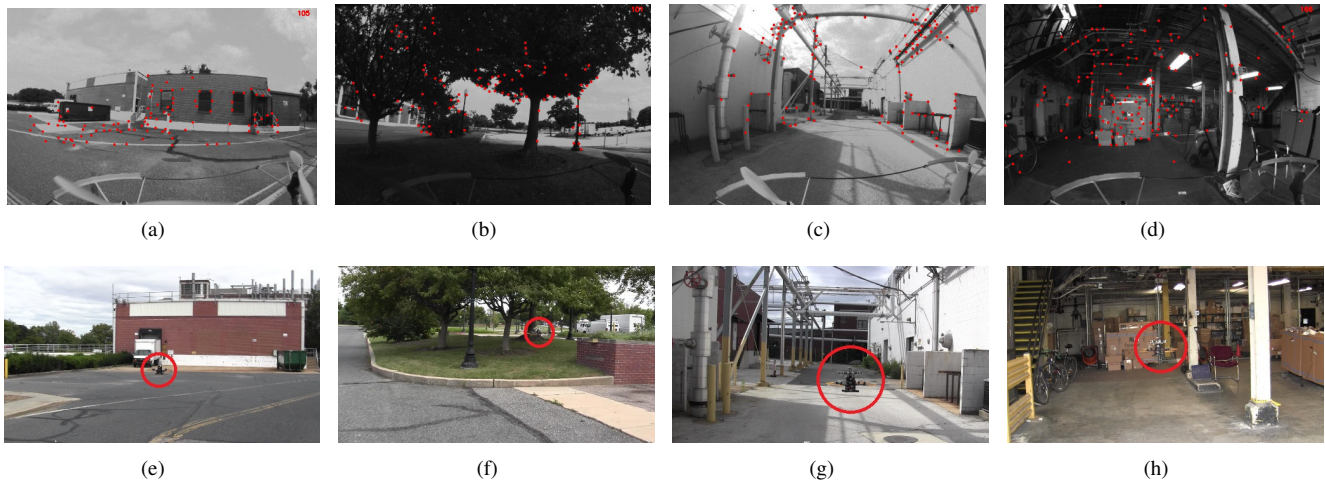
Fig. 6. Images from the onboard camera (Figs. 6(a)- 6(d)) and an external camera (Figs. 6(e)- 6(h)). Note the vast variety of environments, including open space, trees, complex building structures, and indoor environments. We highlight the position of the MAV with a red circle. Videos of the experiments are available in the video attachment and at http://mrsl.grasp.upenn.edu/shaojie/ICRA2014.mp4.

erogeneous sensors that yield either absolute or relative observations at different and varying time intervals. Our approach generates high rate state estimates in real-time for autonomous flight. The proposed approach runs onboard our new 1.9 kg MAV platform equipped with multiple heterogeneous sensors. We demonstrate the robustness of our framework in large-scale, indoor and outdoor autonomous flight experiments that involves traversal through a industrial complex and a tree-lined campus.

In the near future, we would like to integrate higher level planning and situational awareness on our MAV platform to achieve fully autonomous operation across large-scale complex environments.

REFERENCES

[1] S. J. Julier and J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," in *Proc. of SPIE*, I. Kadar, Ed., vol. 3068, July 1997, pp. 182–193.

[2] R. V. D. Merwe, E. A. Wan, and S. I. Julier, "Sigma-point kalman filters for nonlinear estimation: Applications to integrated navigation," in *Proc. of AIAA Guidance, Navigation, and Controls Conf.*, Providence, RI, Aug. 2004.

[3] A. Bachrach, S. Prentice, R. He, and N. Roy, "RANGE-robust autonomous navigation in gps-denied environments," *J. Field Robotics*, vol. 28, no. 5, pp. 644–666, 2011.

[4] S. Shen, N. Michael, and V. Kumar, "Autonomous multi-floor indoor navigation with a computationally constrained MAV," in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Shanghai, China, May 2011, pp. 20–25.

[5] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time onboard visual-inertial state estimation and self-calibration of mavs in unknown environments," in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Saint Paul, MN, May 2012, pp. 957–964.

[6] D. G. Kottas, J. A. Hesch, S. L. Bowman, and S. I. Roumeliotis, "On the consistency of vision-aided inertial navigation," in *Proc. of the Intl. Sym. on Exp. Robot.*, Quebec, Canada, June 2012.

[7] F. Fraundorfer, L. Heng, D. Honegger, G. H. Lee, L. Meier, P. Tanskanen, , and M. Pollefeys, "Vision-based autonomous mapping and exploration using a quadrotor MAV," in *Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots and Syst.*, Vilamoura, Algarve, Portugal, Oct. 2012.

[8] K. Schmid, T. Tomic, F. Ruess, H. Hirschmuller, and M. Suppa, "Stereo vision based indoor/outdoor navigation for flying robots," in *Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots and Syst.*, Tokyo, Japan, Nov. 2013.

[9] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy, "Visual odometry and mapping for autonomous flight using an RGB-D camera," in *Proc. of the Intl. Sym. of Robot. Research*, Flagstaff, AZ, Aug. 2011.

[10] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grixa, F. Ruess, M. Suppa, and D. Burschka, "Autonomous UAV: Research platform for indoor and outdoor urban search and rescue," *IEEE Robot. Autom. Mag.*, vol. 19, no. 3, pp. 46–56, 2012.

[11] S. I. Roumeliotis and J. W. Burdick, "Stochastic cloning: A generalized framework for processing relative state measurements," in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Washington, DC, May 2002, pp. 1788–1795.

[12] J. Carlson, "Mapping large urban environments with GPS-aided SLAM," Ph.D. dissertation, CMU, Pittsburgh, PA, July 2010.

[13] D. Schleicher, L. M. Bergasa, M. Ocaa, R. Barea, and E. Lpez, "Real-time hierarchical GPS aided visual SLAM on urban environments," in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Kobe, Japan, May 2009, pp. 4381–4386.

[14] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, "Vision-based state estimation and trajectory control towards high-speed flight with a quadrotor," in *Proc. of Robot.: Sci. and Syst.*, Berlin, Germany, 2013.

[15] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to mav navigation," in *Proc. of the IEEE/RSJ Intl. Conf. on Intell. Robots and Syst.*, Tokyo, Japan, Nov. 2013.

[16] T. Lefebvre, H. Bruyninckx, and J. D. Schuller, "Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators"," *IEEE Trans. Autom. Control*, vol. 47, no. 8, pp. 1406–1409, 2002.

[17] D. C. Moore, A. S. Huang, M. Walter, and E. Olson, "Simultaneous local and global state estimation for robotic navigation," in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Kobe, Japan, May 2009, pp. 3794–3799.

[18] R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimizations," in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, Shanghai, China, May 2011, pp. 3607–3613.

[19] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Proc. of the IEEE Intl. Conf. on Robot. and Autom.*, vol. 4, Rayleigh, NC, Mar. 1987, p. 850.

[20] T. Lee, M. Leoky, and N. McClamroch, "Geometric tracking control of a quadrotor uav on SE(3)," in *Proc. of the Intl. Conf. on Decision and Control*, Atlanta, GA, Dec. 2010, pp. 5420–5425.