Adaptive Multi-Sensor Localization Information Fusion For Autonomous Urban Air Mobility Operations

Vahram Stepanyan *
Universities Space Research Association, Columbia, MD 21046

Thomas Lombaerts†
NASA Ames Research Center, KBR Wyle Services, Moffett Field, CA 94035

Kimberlee Shish ‡ and Nicholas Cramer§
NASA Ames Research Center, Moffett Field, CA 94035

An adaptive method is developed to iteratively fuse the information provided by multiple sensors to enable autonomous urban air mobility type operations. First, noisy and bias corrupted IMU readings are processed as soon as they arrive using kinematic equations represented in the vehicle's body frame. To correct the systems drift resulting from the integration, an information content measure is introduced to decide on the environment. For the cluttered environment the information provided by environmental sensors is counted as reliable and the drift correction as accurate. For the open space, the GPS data is counted as reliable, and the drift correction is done based on the GPS readings. The measurement noise effects are minimize using Iterated Extended Kalman Filter framework. The algorithm is implemented in the in-house developed FlightDeckz simulation environment using an IMU model, simulated video recorded from a camera mounted on the vehicle (for the purpose of this study, outside scenery was generated with XPlane), which flies in an urban environment, and GPS data generated from the environment's digital map.

I. Introduction

Research interests in increasingly autonomous systems have been growing in resent years along with the industry push to make the self-driving cars [25], package delivery drones [1, 39] and air-taxis [2, 36] reality. These research efforts span across many disciplines, which can be summarized in the following way: sensing, perception, decision making, control technologies [20].

The focus of this paper is perception, in particular, that is fusion of information provided by the sensors to estimate the state of an autonomous system and the environment around it. Generally, these sensors can be divided into two groups: sensors that provide data about the autonomous system, such as accelerometers, gyroscopes, magnetometers, altimeters, GPS, etc. and sensors that sense the environment (exteroceptive), such as radars, cameras, lidars, ultrasound, sonars, etc. Some of these sensors include signal processing algorithms for object detection, image segmentation, labeling, classification, feature point selection, etc. A systematic review of sensors range, accuracy, data rate, etc. can be found in Ref. [27]. A specific survey of the capabilities and gaps in the exteroceptive sensors for autonomous urban air mobility applications is given in Ref. [28]

To reduce the uncertainty of sensor data and to increase the system's accuracy, fusion algorithms or techniques are used, which can be classified according to different criteria such as relations between the input data sources, input/output data types and their nature, abstraction level of the employed data, different data fusion levels, and architecture types [6]. Another categorization is established in [27] based on data processing methods such as Gaussian filters, probabilistic inference approach, or machine learning algorithms.

Data fusion algorithms are also characterized based on the way information from multiple sources is processed as tightly or fully-coupled [3, 7], partly-coupled, and loosely-coupled or decoupled [24, 38]. The tightly-coupled

*Principal Scientist, NASA Ames Research Center/Mail Stop 269/1, AIAA Senior Member, email: vahram.stepanyan@nasa.gov
†Aerospace Research Engineer, KBR Wyle Services, Intelligent Systems Division, Mail Stop 269-1, AIAA Associate Fellow, email: thomas.lombaerts@nasa.gov
‡Aerospace Research Engineer, Intelligent Systems Division, Mail Stop 269-1, AIAA member, email: kimberlee.h.shish@nasa.gov
§Aerospace Research Engineer, Intelligent Systems Division, Mail Stop 269-1, AIAA member, email: nicholas.b.cramer@nasa.gov

algorithms combine all sensors data into a single optimization routine. This is the most accurate, but computationally expensive approach and is prone to delays. On the other hand, loosely-coupled methods process data separately and fuse the results afterwards. These are computationally effective methods, but decoupling leads to information loss, which may affect the accuracy.

The main task performed by data fusion algorithms is to map the environment and localize the vehicle with respect to the map, which is commonly called simultaneous localization and mapping (SLAM) task. It has been in the focus of autonomous driving research community for decades. A survey of some SLAM approaches in autonomous driving can be found in [4]. These SLAM approaches use IMU, GPS, LIDAR and camera data all together or in some combinations to only estimate the vehicle position or perform both mapping and localization. In [41], a modular approach is proposed to sequentially estimate the vehicle's motion coarse-to-fine manner, where first IMU data is used to predict the motion, then visual and inertial data are coupled to improve the estimates, and finally laser points scan matching is used to further refine the estimates and register laser points to build a map. In [14], Adaptive Extended Kalman Filter framework is used to process IMU data, and LIDAR and GPS data are used to update the observation vector as they become available. Meanwhile, error covariance matrices are adapted based on the LIDAR scan matching error and GPS precision dilution measurements. In [32], a laser scanner itegration method with INSm is presented for navigation in GPS-denied urban environments. Line features are extracted from the scan images and exploited for navigation, as lines are computationally efficient to extract, and are common in man-made environments. In [29], for steady-state estimation error reduction, GPS and IMU data are fused in the filtered innovations Kalman filter framework using classical proportional-integral control approach. Other approaches use only LIDAR to achieve low-drift SLAM [30], conduct rigid body attitude estimation via a descriptor system formulation using measurements from IMU and magnetometers [23], fuse data from a stationary stereo camera and a mobile platform mounted laser range finder to track an object in a lab environment [9] , integrate radar and visual sensors data to improve object recognition accuracy in the vehicle surrounding [19], just to mention few of them.

Most frequently used fusion algorithms combine IMU measurements and camera images in visual-inertial navigation systems (VINS) because of decreasing costs and size of these sensors. VINS algorithms can be divided into two groups: visual-inertial simultaneous localization and mapping (VSLAM), and visual-inertial odometry (VIO). Recent VSLAM approaches are reviewed in [5] , and a nice review of VIO approaches can be found in [15]. Both group of algorithms use estimation techniques, which employ filter-based approaches such as extended Kalman filter (EKF) [18, 24] the unscented Kalman filter (UKF) [10, 17, 22], Information Filter [37, 40] , Particle Filter [8], batch or incremental smoother [16], and optimization-based approaches such as window optimization [11, 13, 21, 26], Bundle adjustment [42] and graph representation [35]. An effort to compare filter-based techniques and optimization approaches for a SLAM applications has been proposed in [33] with the outcome that the optimization tends to give better results than filters due to linearization issues. However, the EKF-based methods are still popular because of computational efficiency. The drawback is that these methods may diverge when the underlying dynamic model becomes unobservable from the provided measurements. As it is shown in [12], it can happen when the flight vehicle is not accelerating linearly and is not rotating. Additionally, singularities in the resulting covariance matrices arise when using quaternion-based kinematics for state propagation in the Kalman Filter framework because of additional unit constraints on the quaternions [34]. This issue is handled by using error state equations in the Kalman filter prediction stage by expressing 4D-quaternion error through 3D-rotation angle error [15, 31].

In this paper we take a similar to [14, 31] approach and process the information as soon as it becomes available from any sensor. While in [14] LIDAR and GPS data are used to update the observation vector in the Adaptive Extended Kalman Filter framework as they become available, and covariance matrices are adapted based on the LIDAR scan matching error and GPS precision dilution, here we assume that the available environmental sensor package data is preprocessed and 3D feature points and corresponding covariances are available, which is the case in [31]. However, our approach differs by adaptively using sensors data for the state estimation correction depending on the environment being open or cluttered, which is determined based on the analysis of acquired images. First, we process noisy and bias corrupted IMU readings as they arrive using kinematic equations represented in the vehicle's body frame. Then, information content measures are introduced to decide on the environment. For the cluttered environment the information provided by environmental sensors is counted as reliable for the drift correction. For the open space, the GPS data is counted as reliable, and the drift correction is done based on the GPS readings. The measurement noise effects are minimize using Iterated Extended Kalman Filter (IEKF) framework. The algorithm is implemented in the in-house developed FlightDeckz simulation environment using an IMU model, simulated video recorded from a camera mounted on the vehicle (for the purpose of this study, outside scenery was generated with XPlane), which flies in an urban environment, and GPS data generated from the environment's digital map.

II. Problem Formulation

We assume that the flight vehicle is equipped with an inertial measurement unit (IMU), environmental sensors such as cameras, LIDAR, radar, barometer, etc. and Global Positioning System (GPS) receivers. The information provided by these sensors except for GPS is relative to the vehicle's body frame denoted by F_B which is assumed to be centered at the vehicle's center of mass. The IMU provides high-frequency measurements

$$\begin{aligned} \boldsymbol{a}_m &= \mathbb{R}^\top(\boldsymbol{a} - \boldsymbol{g}) + \boldsymbol{b}_a + \boldsymbol{n}_a \\ \boldsymbol{\omega}_m &= \boldsymbol{\omega} + \boldsymbol{b}_\omega + \boldsymbol{n}_\omega \end{aligned} \tag{1}$$

expressed in the frame F_B, where $\boldsymbol{a}_m$ is the accelerometer reading (subscript m here and in what follows indicates measured quantities), $\boldsymbol{a}$ is the vehicles inertial acceleration, $\boldsymbol{g}$ is the gravity acceleration, $\mathbb{R}$ is the rotation matrix from F_B to F_I, $\boldsymbol{b}_a$ is the acceleration bias, $\boldsymbol{n}_a$ is the measurement noise, $\boldsymbol{\omega}_m$ is gyroscope reading, $\boldsymbol{\omega}$ is the vehicle's angular rate expressed in the frame F_B, $\boldsymbol{b}_\omega$ is the rate bias, and $\boldsymbol{n}_\omega$ is the measurement noise. The measurement noises are assumed to be isotropic and Gaussian in all directions with standard deviations σ_a and σ_ω respectively.

Environmental sensors are assumed to provide mid-frequency data (measured or estimated by means of some signal processing algorithms) of noise corrupted range R_k, azimuth γ_k and elevation λ_k and their first order derivatives of $k-th$ feature point with respect to the sensor frame

$$\begin{aligned} R_{km} &= R_k + n_R \\ \lambda_{km} &= \lambda_k + n_\lambda \\ \gamma_{km} &= \gamma_k + n_\gamma \\ \dot{R}_{km} &= \dot{R}_k + n_{\dot{R}} \\ \dot{\lambda}_{km} &= \dot{\lambda}_k + n_{\dot{\lambda}} \\ \dot{\gamma}_{km} &= \dot{\gamma}_k + n_{\dot{\gamma}}\,, \end{aligned} \tag{2}$$

where the measurement/estimation uncertainties $n_R,\ n_\lambda,\ n_\gamma,\ n_{\dot{R}},\ n_{\dot{\lambda}},\ n_{\dot{\gamma}}$ are assumed to be Gaussian noises with standard deviations $\sigma_R,\ \sigma_\lambda,\ \sigma_\gamma,\ \sigma_{\dot{R}},\ \sigma_{\dot{\lambda}},\ \sigma_{\dot{\gamma}}$ respectively for all feature points. Since the orientation and the position of origin of sensor frames are generally known in the frame F_B, we assume that the equations (2) are given in F_B. It is also assumed that the environmental sensor data has been pre-processed with some algorithm to select essential/feature points from the corresponding images or point clouds.

In some cases the aircraft may also be equipped whit an inertial navigation system (INS) that provides vehicle orientation angles with some uncertainties

$$\begin{aligned} \phi_m &= \phi + n_\phi \\ \theta_m &= \theta + n_\theta \\ \psi_m &= \psi + n_\psi\,, \end{aligned} \tag{3}$$

where the uncertainties $n_\phi,\ n_\theta,\ n_\psi$ are assumed to be Gaussian noises with standard deviations $\sigma_\phi,\ \sigma_\theta,\ \sigma_\psi$ respectively.

Finally, we assume that the GPS provides, generally at the slowest rate, the inertial coordinates of the vehicle's center of mass with some uncertainty as

$$\boldsymbol{p}_m = \boldsymbol{p} + \boldsymbol{n}_p\,, \tag{4}$$

where the uncertainty $\boldsymbol{n}_p$ is assumed to be a Gaussian noise with standard deviations σ_p in all directions.

The objective is to estimate the vehicle state as accurately as possible for all flight conditions fusing data from given sensor readings, assuming that the static map of the environment is available.

III. Adaptive mechanism

To correct the systems drift, exteroceptive sensors or GPS data are used depending on the reliability of the information they provide, which depends on the flight environment. Namely, it is assumed that in the cluttered environment exteroceptive sensors provide reliable information, whereas GPS data is reliable in the open environment. Therefore we derive criteria to characterize the environment based on the available exteroceptive sensors data. To this end, we first derive the criteria for visual sensor information to determine the openness of flight environment, then extend the

approach to other exteroceptive sensors. For this purpose we use the image information content such as image intensity and the gradient. One way to use the image intensity is to compute the entropy of one dimensional image histogram as it is done in Ref. [43] for infrared images. However, as it was pointed out in Ref. [44], this approach is totally blind to image spatial structure in the sense that visibly more complex images such as uniformly distributed random ones give the same entropy value as structured images with linear gradients. To overcome this issue, the image partial derivatives were used to define a measure of image entropy. In Ref. [45] a magnitude weighted and doubled image gradient orientation was used for spatial information consideration in 3-D tomographic data-sets alignment. In Ref. [46] the probability distribution in the Shannon entropy formula (see for example Ref. [47] for details) was replaced by the spatial distribution of the image gradient itself. In Ref. [48] the directional distribution of the gradient intensity was used for image registration. In Ref. [44], the joint probability density function of the image partial derivatives was used to introduce a new information content measure for 2D images, where it is called delentropy and claimed that it captures the underlying spatial image structure better. However, this approach increases the computational cost.

Here, we first determine the local horizontal plane, which passes through the optical center of the visual sensor and consider only part of the image above that plane, taking into account the camera tilt angle with respect to vehicle frame and the prior pitch angle from the Kalman filter. The rationale behind is that only the structures above that horizontal plane affect the GPS signal. Hence, the less informativeness of the mentioned upper part of the visual image corresponds to reliable GPS signal. Next, assuming a gray scale image, we derive an information content measure of the image upper part, which comprises if three factors: normalized average intensity I_{na}, normalized maximum gradient magnitude J_{mag} and entropy of the gradient direction distribution. For the $n \times m$ 2D gray scale image the normalized average intensity I_{na} is computed as

$$I_{na} \quad = \quad \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{n} f(i,j)\,, \tag{5}$$

where $f(i,j)$ is the intensity of (i, j)-th pixel. We take it into account in the information content measure via a Gaussian function centered at the midpoint of normalized gray spectrum as

$$J_{int} \quad = \quad \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(I_{na}-0.5)^2}{2\sigma^2}\right), \tag{6}$$

which effectively washes out both black and white ends of the gray spectrum. Next, for the (i, j)-th pixel we compute the gradient vector $\boldsymbol{g}(i,j) = [g_x(i,j)\ g_y(i,j)]^\top$, its magnitude and direction

$$g_m(i,j) = \sqrt{g_x^2(i,j) + g_y^2(i,j)}, \quad g_\phi(i,j) = \arctan\left(\frac{g_y(i,j)}{g_x(i,j)}\right). \tag{7}$$

The information content measure component for the normalized maximum gradient magnitude J_{mag} is defined as

$$J_{mag} \quad = \quad c \max_{i=1,..,n,\ j=1,...,m} g_m(i,j)\,, \tag{8}$$

where c is a normalization factor set to $255\sqrt{2}$, and the component for the gradient direction J_ϕ is defined via a sigmoid function

$$J_\phi \quad = \quad \frac{1}{1+\exp(a(b-E_\phi))}\,, \tag{9}$$

following Ref. [43] to compensate for too strong down-rating of low entropy values, where E_ϕ is the entropy of gradient direction intensity defined as

$$E_\phi \quad = \quad -\sum_{i=-\pi}^{\pi} \sum_{j=1}^{n} p_i \log_2(p_i)\,, \tag{10}$$

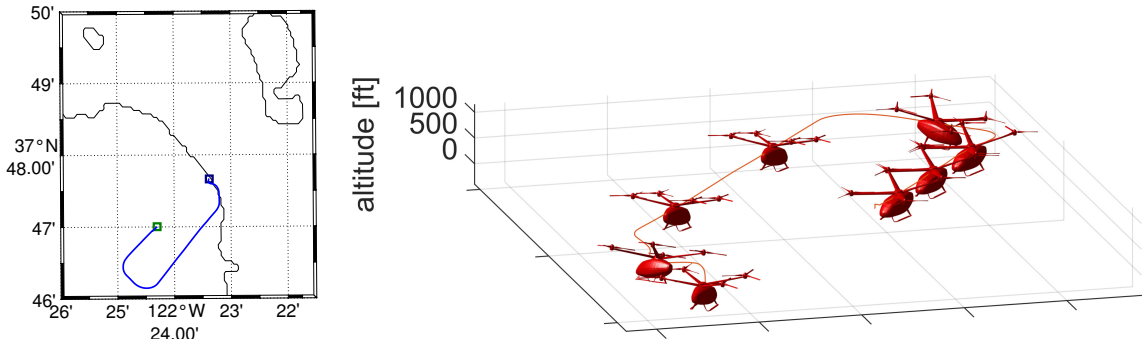
where p_i represents 1D histogram of $g_\phi(i,j)$. The complete information content measure is

$$J \quad = \quad J_{int} J_{mag} J_\phi\,. \tag{11}$$

Smaller values of J indicate that the image is mainly toward the black or white end of the gray scale spectrum, and there are not essential abrupt or gradual intensity changes. That is, if $J \le J_*$, where the threshold J_* can be set to 1 by selecting proper values for $a,\ b,\ c,\ \sigma$, then the flight environment is considered open and GPS signal reliable, otherwise the image has a rich enough content for the drift correction.

IV. Application Example

This example considers a localization task in simulations using in-house developed FlightDeckZ environment. Fig. 1 shows the map and the 3D trajectory of a sample demonstration mission, flown by means of a pre-programmed flight plan in the flight management system. The vehicle takes of vertically from the tip of pier 14 near the ferry building in San Francisco and flies a circuit after which it lands vertically on the rooftop of the parking structure at Mission and 5th. A more detailed discussion about the control aspects of this sample demonstration mission is given in Ref. [50].



(a) Map of flight plan

(b) 3D trajectory of flight

Fig. 1 Map and 3D trajectory of flight plan, source: Ref. [50]

A. Description of simulated vehicle

For this study, a quadrotor concept aircraft is used with a payload of 6 occupants at 6480 lbs total, with a large fuselage and very large rotors, depicted in Fig. 2. This is a battery-powered electrically driven concept aircraft. Ref. [49] provides more information about this concept aircraft. A Quasi-Linear Parameter Varying (qLPV) model was generated with NDARC/FlightCODE. This qLPV model is used as the simulation model in this study.

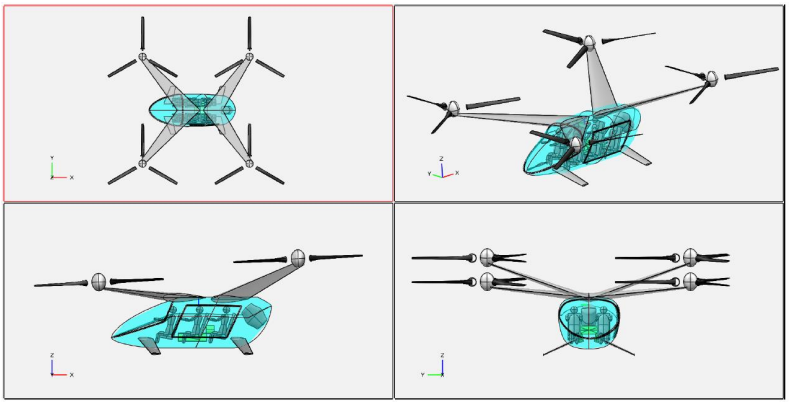


Fig. 2 6 passenger quadrotor concept vehicle, source: [49]

Standard conventional available sensor information for fixed wing aircraft is classified in three categories. First there are the air data sensors (ADS), providing true airspeed V_{TAS}, angle of attack α and angle of side-slip β. A second class are the data from the inertial navigation system (INS, consisting of inertial and magnetic equipment) giving measurement values for the specific forces A_x, A_y, A_z, the rotational rates p, q, r and aircraft attitude angles ϕ, θ, ψ. The third and last category is a combination of INS and GPS measurements leading to data for three dimensional position x, y, z and inertial velocity components u, v, w. However, for the vehicle in this application example, an unconventional sensor suite is considered. Since the presence of reliable air data sensor measurements on a quadrotor vehicle with significant downwash from the rotor blades is not guaranteed, these are omitted here. Furthermore, typical UAM operations will be in an urban environment, where the GPS signal could become unreliable due to temporarily

loss of satellite signal. As a complement, exteroceptive sensors are used (in this simulation study only visual camera is used) to determine the range R_k, azimuth γ_k and elevation λ_k, together with their respective time derivatives $\dot{R}_k$, $\dot{\lambda}_k$ and $\dot{\gamma}_k$ ($k = 1, \ldots, N$) of a static feature points with respect to the vehicle's body frame.

Table 1 gives information about the instrumentation errors which occur for each kind of measuring equipment mentioned above. As verification of sensor accuracy, the accelerometer and rate gyro biases have been estimated during the state estimation procedure. By making use of the kinematic and observation model of the aircraft, it is possible to estimate these biases, which will be discussed in more detail below.

sensor	variables	bias error	noise error
translational accelerometer	a_x, a_y, a_z	✓	✓
rate gyro	p, q, r	✓	✓
INS	ϕ, θ, ψ		✓
GPS	x, y, z		✓
exteroceptive sensor for feature point i	R_i, λ_i, γ_i		✓
	$\dot{R}_i, \dot{\lambda}_i, \dot{\gamma}_i$		✓

Table 1 Instrumentation error information for measuring equipment

The research presented in this paper considers flight trajectory reconstruction with a loosely coupled flight instrumentation system using processed observations, as opposed to the tightly coupled counterpart with raw measurement data.

B. Construction of measured sensor data with respect to feature point

Since the FlightDeckZ environment does not yet include the sensor model of a visual system that tracks a feature point, the first step for the application example was to develop this sensor model. This model generates noise perturbed range R_i, azimuth γ_i and elevation λ_i readings and their first order derivatives of an observed feature point, based on its known position and the ownship position, which is part of the state vector. As a simple test, a centrally located static feature point is defined with respect to this typical intra-city UAM demonstration vehicle route, as shown in Fig. 3(a). The on-board sensor detects the range, azimuth and elevation, as well as range rate, azimuth rate and elevation rate of this centrally located static feature point, of which the position is also known. Based on the geometric information of all positions along the trajectory, it is possible to calculate these measured variables as follows. Define the vehicle's own position in earth fixed coordinates as (x, y, z) and the corresponding position of the aforementioned feature point i as (x_i, y_i, z_i), then the relative distance between both in body fixed coordinates is defined as:

$$\begin{bmatrix} \Delta x_{b_i} \\ \Delta y_{b_i} \\ \Delta z_{b_i} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i - x \\ y_i - y \\ z_i - z \end{bmatrix} \tag{12}$$

And the range, azimuth and elevation of feature point i as measured from the vehicle are then defined as follows:

$$R_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \tag{13}$$

$$\gamma_i = \arctan \frac{\Delta y_{b_i}}{\Delta x_{b_i}} \tag{14}$$

$$\lambda_i = \arctan \frac{-\Delta z_{b_i}}{R_{\text{hor}_i}} \tag{15}$$

for which the horizontal range R_{hor_i} is defined as:

$$R_{\text{hor}_i} = \sqrt{(x_i - x)^2 + (y_i - y)^2} \tag{16}$$

Similarly, the time derivative of the relative distance between both in body fixed coordinates is calculated by means of the chain rule:

$$\begin{bmatrix}\Delta\dot{x}_{b_i}\\ \Delta\dot{y}_{b_i}\\ \Delta\dot{z}_{b_i}\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & \cos\phi & \sin\phi\\ 0 & -\sin\phi & \cos\phi\end{bmatrix}\begin{bmatrix}\cos\theta & 0 & -\sin\theta\\ 0 & 1 & 0\\ \sin\theta & 0 & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\psi & \sin\psi & 0\\ -\sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\dot{x}_i-\dot{x}\\ \dot{y}_i-\dot{y}\\ \dot{z}_i-\dot{z}\end{bmatrix} +$$
$$\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\phi & \sin\phi\\ 0 & -\sin\phi & \cos\phi\end{bmatrix}\begin{bmatrix}\cos\theta & 0 & -\sin\theta\\ 0 & 1 & 0\\ \sin\theta & 0 & \cos\theta\end{bmatrix}\begin{bmatrix}-\sin\psi & \cos\psi & 0\\ -\cos\psi & -\sin\psi & 0\\ 0 & 0 & 0\end{bmatrix}\dot{\psi}\begin{bmatrix}x_i-x\\ y_i-y\\ z_i-z\end{bmatrix} +$$
$$\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\phi & \sin\phi\\ 0 & -\sin\phi & \cos\phi\end{bmatrix}\begin{bmatrix}-\sin\theta & 0 & -\cos\theta\\ 0 & 0 & 0\\ \cos\theta & 0 & -\sin\theta\end{bmatrix}\dot{\theta}\begin{bmatrix}\cos\psi & \sin\psi & 0\\ -\sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_i-x\\ y_i-y\\ z_i-z\end{bmatrix} +$$
$$\begin{bmatrix}0 & 0 & 0\\ 0 & -\sin\phi & \cos\phi\\ 0 & -\cos\phi & -\sin\phi\end{bmatrix}\dot{\phi}\begin{bmatrix}\cos\theta & 0 & -\sin\theta\\ 0 & 1 & 0\\ \sin\theta & 0 & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\psi & \sin\psi & 0\\ -\sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_i-x\\ y_i-y\\ z_i-z\end{bmatrix} \tag{17}$$

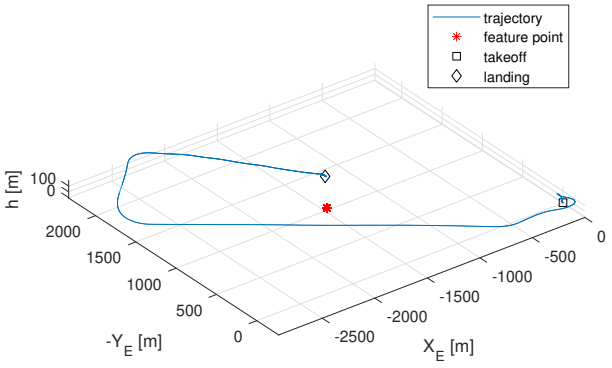
And the time derivatives of the range, azimuth and elevation of feature point i as measured from the vehicle are then defined as follows:

$$\dot{R}_i = \frac{\Delta x_{b_i}\Delta\dot{x}_{b_i} + \Delta y_{b_i}\Delta\dot{y}_{b_i} + \Delta z_{b_i}\Delta\dot{z}_{b_i}}{R_i} \tag{18}$$

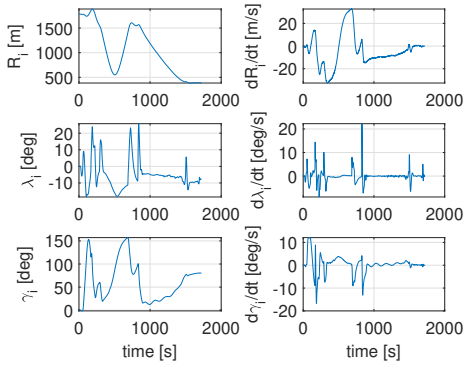
$$\dot{\gamma}_i = \frac{\Delta\dot{y}_{b_i}\Delta x_{b_i} - \Delta y_{b_i}\Delta\dot{x}_{b_i}}{R^2_{\mathrm{hor}_i}} \tag{19}$$

$$\dot{\lambda}_i = \frac{-\Delta\dot{z}_{b_i}R_{\mathrm{hor}_i} + \Delta z_{b_i}\left(\Delta x_{b_i}\Delta\dot{x}_{b_i} + \Delta y_{b_i}\Delta\dot{y}_{b_i}\right)}{R_{\mathrm{hor}_i}R_i^2} \tag{20}$$

Based on the geometry defined in Fig. 3(a) and the flight path of the vehicle, the measured variables range R, azimuth γ, elevation λ, and their time derivatives $\dot{R}$, $\dot{\gamma}$ and $\dot{\lambda}$, as calculated with Eq. (12)–(20), are shown in Fig. 3(b). These measured variables are used in the observation model of an Extended Kalman Filter, instead of the usual GPS coordinates, for reconstructing the vehicle position when the flight environment is determined to be clattered.



(a) 3D trajectory of flight with centrally located feature point



(b) Measured variables range R, azimuth γ, elevation λ, and their time derivatives

Fig. 3 3D trajectory and measured variables with respect to feature point

C. Nonlinear Aircraft kinematics and observation models

The nonlinear kinematic observation model of the aircraft is given as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}_m(t), \boldsymbol{b}, t] + \mathbf{G}[\mathbf{x}(t)]\mathbf{w}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{21}$$

$$\mathbf{z}_m(t) = \mathbf{h}[\mathbf{x}(t), \mathbf{u}_m(t), \boldsymbol{b}, t] + \mathbf{v}(t), \quad t = t_i, \quad i = 1,\ 2, \ldots \tag{22}$$

where Eq. (21) is known as the kinematic state equation with input noise vector $\mathbf{w}$ and Eq. (22) is called the observation equation with output noise vector $\mathbf{v}$. The nonlinear vector functions $\mathbf{f}$ and $\mathbf{h}$ may depend both implicitly (via $\mathbf{x}$ and $\mathbf{u}_m$) and explicitly on t and it will be assumed that both $\mathbf{f}$ and $\mathbf{h}$ are continuous and continuously differentiable with respect to all elements of $\mathbf{x}$ and $\mathbf{u}_m$. The system equation variables are defined as follows:

$$\mathbf{x} = [x \quad y \quad z \quad u \quad v \quad w \quad \phi \quad \theta \quad \psi]^T \tag{23}$$

$$\mathbf{u}_m = \mathbf{u} + \boldsymbol{b} + \mathbf{w} = [A_x \quad A_y \quad A_z \quad p \quad q \quad r]^T + [b_x \quad b_y \quad b_z \quad b_p \quad b_q \quad b_r]^T + \mathbf{w} \tag{24}$$

$$\mathbf{z}_m = \begin{bmatrix} R_i & \gamma_i & \lambda_i & \dot{R}_i & \dot{\gamma}_i & \dot{\lambda}_i & \phi_m & \theta_m & \psi_m \end{bmatrix}^T \tag{25}$$

where the aircraft state vector $\mathbf{x}$ in Eq. (23) contains inertial position, body inertial velocity components and aircraft attitude angles. The measured input vector $\mathbf{u}_m$ in Eq. (24) consists of specific forces and angular rates, perturbed with sensor biases and input noise, where the sensor biases are collected in vector $\boldsymbol{b}$, which contributes for the augmented state vector as $\mathbf{x}_{\text{aug}} = [\mathbf{x}, \boldsymbol{b}]$. Finally, there is the measured output vector $\mathbf{z}_m$ in Eq. (**??**), consisting of optical sensor measurement data of range, azimuth, elevation of feature point i, together with the respective time derivatives and INS supplied attitude angles. Also the measured output vector is contaminated with output noise. However, sufficient accurate information about the measurement disturbances (such as the aforementioned downwash) is lacking to construct reasonable observation functions for the air data measurements. This is an interesting topic for future research.

Additionally, the input noise vector $\mathbf{w}(t)$ is a continuous time white noise process and the output noise vector $\mathbf{v}(t_i)$ is a discrete time white noise sequence. Both are mutually uncorrelated as well as between the different input and output channels individually. Moreover, based upon the known on-board measurement equipment characteristics, standard deviations are specified by the equipment manufacturers, taking into account some safety margins. Therefore, the error model can be described as follows:

$$\mathbf{v}(t_i) = \begin{bmatrix} n_R & n_\gamma & n_\lambda & n_{\dot{R}} & n_{\dot{\psi}_i} & n_{\dot{\lambda}} & n_\phi & n_\theta & n_\psi \end{bmatrix}^T \tag{26}$$

$$\mathbf{w}(t) = \begin{bmatrix} n_x & n_y & n_z & n_p & n_q & n_r \end{bmatrix}^T \tag{27}$$

$$E\left\{\mathbf{w}(t)\mathbf{w}^T(\tau)\right\} = \mathbf{Q}\delta(t-\tau); \quad \mathbf{Q} = \text{diag}\left(\sigma_x^2, \sigma_y^2, \sigma_z^2, \sigma_p^2, \sigma_q^2, \sigma_r^2\right) \tag{28}$$

$$E\left\{\mathbf{v}(t_i)\mathbf{v}^T(t_j)\right\} = \mathbf{R}\delta_{ij}; \quad \mathbf{R} = \text{diag}\left(\sigma_R^2, \sigma_\gamma^2, \sigma_\lambda^2, \sigma_{\dot{R}}^2, \sigma_{\dot{\gamma}}^2, \sigma_{\dot{\lambda}}^2, \sigma_\phi^2, \sigma_\theta^2, \sigma_\psi^2\right) \tag{29}$$

$$E\left\{\mathbf{w}(t)\mathbf{v}^T(t_i)\right\} = 0, \quad \text{for} \quad \text{t} = \text{t}_\text{i}, \quad \text{i} = 1,\ 2, \ldots \tag{30}$$

The kinematic equations can be written as follows, where the aircraft is considered as a rigid body above a flat non-rotating earth:

$$\dot{x} = [u\cos\theta + (v\sin\phi + w\cos\phi)\sin\theta]\cos\psi - (v\cos\phi - w\sin\phi)\sin\psi \tag{31}$$

$$\dot{y} = [u\cos\theta + (v\sin\phi + w\cos\phi)\sin\theta]\sin\psi + (v\cos\phi - w\sin\phi)\cos\psi \tag{32}$$

$$\dot{z} = -u\sin\theta + (v\sin\phi + w\cos\phi)\cos\theta \tag{33}$$

$$\dot{u} = A_x - g\sin\theta + rv - qw \tag{34}$$

$$\dot{v} = A_y + g\cos\theta\sin\phi + pw - ru \tag{35}$$

$$\dot{w} = A_z + g\cos\theta\cos\phi + qu - pv \tag{36}$$

$$\dot{\phi} = p + q\sin\phi\tan\theta + r\cos\phi\tan\theta \tag{37}$$

$$\dot{\theta} = q\cos\phi - r\sin\phi \tag{38}$$

$$\dot{\psi} = q\frac{\sin\phi}{\cos\theta} + r\frac{\cos\phi}{\cos\theta} \tag{39}$$

In this system of kinematic equations, all positions and velocities are inertial. The positions are defined in an earth fixed reference frame and the velocity components u, v and w are defined in a body fixed reference frame. Expressions (37) till (39) are the kinematic relations for a sequential rotation along the $Z-$, $Y-$ and $X-$axis respectively and constitute as the conversions of the angular rates from body fixed reference frame towards the earth fixed reference frame (Euler angles). The first three equations, (31) till (33) serve as the conversions of the linear velocity components from body fixed reference frame towards earth fixed reference frame.

The specific forces and angular rates, which are available in the input vector, can be found in expressions (34) till (39). However, since they are supplied as measurements, they are not exact, but contaminated with bias and noise. Consequently, they must be corrected in order to rewrite the system of state equations in the form of expression (21) and the following equalities can be substituded in expressions (34) till (39):

$$A_x = A_{x_m} - b_x - n_x \tag{40}$$
$$A_y = A_{y_m} - b_y - n_y \tag{41}$$
$$A_z = A_{z_m} - b_z - n_z \tag{42}$$

$$p = p_m - b_p - n_p \tag{43}$$
$$q = q_m - b_q - n_q \tag{44}$$
$$r = r_m - b_r - n_r \tag{45}$$

Moreover, the system of state equations can be extended with the following array of equations:

$$\dot{b}_x = 0 \tag{46}$$
$$\dot{b}_y = 0 \tag{47}$$
$$\dot{b}_z = 0 \tag{48}$$
$$\dot{b}_p = 0 \tag{49}$$
$$\dot{b}_q = 0 \tag{50}$$
$$\dot{b}_r = 0 \tag{51}$$

As a result, this set of equations is also extended in order to comply with the augmented state, containing the aircraft states, as well as the inertial equipment biases. Equations (46) till (51) exploit the knowledge that biases are constant. Conventionally, wind speed components in the earth fixed reference frame, which are assumed constant, would be added here in a similar fashion as the biases. However, air data sensors are needed in order to achieve identifiability of wind. Since the presence of reliable air data sensor measurements on a quadrotor vehicle with significant downwash from the rotor blades is not guaranteed, the wind speed components were omitted here, and could be included in a separate state estimation routine where they cannot perturb the other estimates which are considered here.

This rewritten set of state equations can be split up over the nonlinear continuous function $\mathbf{f}$ and the noise contribution function $\mathbf{G}$, which is linear in $\mathbf{w}(t)$. First the nonlinear continuous $\mathbf{f}[\mathbf{x}(t), \mathbf{u}_m(t), \boldsymbol{b}]$ is given:

$$\mathbf{f}[\mathbf{x}(t), \mathbf{u}_m(t), \boldsymbol{b}] = \begin{bmatrix} [u\cos\theta + (v\sin\phi + w\cos\phi)\sin\theta]\cos\psi - (v\cos\phi - w\sin\phi)\sin\psi \\ [u\cos\theta + (v\sin\phi + w\cos\phi)\sin\theta]\sin\psi + (v\cos\phi - w\sin\phi)\cos\psi \\ -u\sin\theta + (v\sin\phi + w\cos\phi)\cos\theta \\ (A_{x_m} - b_x) - g\sin\theta + (r_m - b_r)v - (q_m - b_q)w \\ (A_{y_m} - b_y) + g\cos\theta\sin\phi + (p_m - b_p)w - (r_m - b_r)u \\ (A_{z_m} - b_z) + g\cos\theta\cos\phi + (q_m - b_q)u - (p_m - b_p)v \\ (p_m - b_p) + (q_m - b_q)\sin\phi\tan\theta + (r_m - b_r)\cos\phi\tan\theta \\ (q_m - b_q)\cos\phi - (r_m - b_r)\sin\phi \\ (q_m - b_q)\frac{\sin\phi}{\cos\theta} + (r_m - b_r)\frac{\cos\phi}{\cos\theta} \\ \mathbf{0}_{6\times 1} \end{bmatrix} \tag{52}$$

After which the linear $\mathbf{G}[\mathbf{x}(t)]$ is defined as:

$$\mathbf{G}[\mathbf{x}(t)] = \begin{bmatrix} & & & & \mathbf{0}_{3\times6} & \\ -1 & 0 & 0 & 0 & w & -v \\ 0 & -1 & 0 & -w & 0 & u \\ 0 & 0 & -1 & v & -u & 0 \\ 0 & 0 & 0 & -1 & -\sin\phi\tan\theta & -\cos\phi\tan\theta \\ 0 & 0 & 0 & 0 & -\cos\phi & \sin\phi \\ 0 & 0 & 0 & 0 & -\frac{\sin\phi}{\cos\theta} & -\frac{\cos\phi}{\cos\theta} \\ & & & & \mathbf{0}_{6\times6} & \end{bmatrix} \tag{53}$$

After discussing the kinematic equations in detail, the nonlinear continuous observation equations can be considered, including the output noise contribution $\mathbf{v}(t)$ from (26):

$$R_m = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} + n_R \tag{54}$$

$$\gamma_{mi} = \arctan\frac{\Delta y_{b_i}}{\Delta x_{b_i}} + n_\gamma \tag{55}$$

$$\lambda_{mi} = \arctan\frac{-\Delta z_{b_i}}{R_{\text{hor}_i}} + n_\lambda \tag{56}$$

$$\dot{R}_{mi} = \frac{\Delta x_{b_i}\Delta\dot{x}_{b_i} + \Delta y_{b_i}\Delta\dot{y}_{b_i} + \Delta z_{b_i}\Delta\dot{z}_{b_i}}{R_i} + n_{\dot{R}} \tag{57}$$

$$\dot{\gamma}_{mi} = \frac{\Delta\dot{y}_{b_i}\Delta x_{b_i} - \Delta y_{b_i}\Delta\dot{x}_{b_i}}{R^2_{\text{hor}_i}} + n_{\dot{\gamma}} \tag{58}$$

$$\dot{\lambda}_{mi} = \frac{-\Delta\dot{z}_{b_i}R^2_{\text{hor}_i} + \Delta z_{b_i}\left(\Delta x_{b_i}\Delta\dot{x}_{b_i} + \Delta y_{b_i}\Delta\dot{y}_{b_i}\right)}{R_{\text{hor}_i}R_i^2} + n_{\dot{\lambda}} \tag{59}$$

where Δx_{b_i}, Δy_{b_i}, Δz_{b_i}, $\Delta\dot{x}_{b_i}$, $\Delta\dot{y}_{b_i}$, $\Delta\dot{z}_{b_i}$ and R_{hor_i} are defined in Eq. (12), (16) and (17). Consequently the nonlinear continuous $\mathbf{h}[\mathbf{x}(t), \mathbf{u}_m(t), \boldsymbol{b}]$ in (22) looks as the system of equations (54) to (59) grouped in matrix form, but without the output noise contributions. Since these observation equations make use of the attitude angles ϕ, θ and ψ in Eq. (12) and (17), it is not possible to apply this transformation as a preprocessing step outside the Kalman Filter.

D. State estimator: Iterated Extended Kalman Filter

As already announced and apparent from the structure above in this section, a Kalman Filter can be used in order to estimate the aircraft states $\mathbf{x}$ and inertial sensor biases $\boldsymbol{b}$ based on the available information in $\mathbf{u}_m$ and $\mathbf{z}_m$. The nonlinear behavior of R_i, γ_i, λ_i, $\dot{R}_i$, $\dot{\gamma}_i$, $\dot{\lambda}_i$ and potential large initial state errors (due to the choice of the initial state values) necessitate the use of an Iterated Extended Kalman Filter (IEKF)[51], which contains an additional iteration loop in the correction step, compared to a regular Extended Kalman Filter (EKF):

1) one step ahead prediction (time propagation):

$$\boldsymbol{\eta}_1 = \hat{\mathbf{x}}\,(k+1\,|k\,) = \hat{\mathbf{x}}\,(k\,|k\,) + \int_k^{k+1} \mathbf{f}\,(\mathbf{x}(t), \mathbf{u}_m(t), \boldsymbol{b}, t)\,dt \tag{60}$$

2) prediction of covariance matrix of the state prediction error vector:

$$\mathbf{P}\,(\,k+1|\,k) = \boldsymbol{\Phi}\,(k,\tau)\,\mathbf{P}\,(k\,|k\,)\,\boldsymbol{\Phi}^T\,(k,\tau) + \mathbf{Q}_d\,(k) \tag{61}$$

with:

$$\mathbf{F}(k) = \left.\frac{\partial\mathbf{f}\,(\mathbf{x}(t), \mathbf{u}_m(t), \boldsymbol{b}, t)}{\partial\mathbf{x}(t)}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k}} \qquad \boldsymbol{\Phi}(k,\tau) = e^{\mathbf{F}(k)\Delta t} \tag{62}$$

$$\mathbf{Q}_d\,(k+1|k) = \int_{t_k}^{t_{k+1}} \mathbf{\Phi}\,(t_{k+1},\tau)\,\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^T\mathbf{\Phi}^T\,(t_{k+1},\tau)\,d\tau \tag{63}$$

3) Kalman gain:

$$\mathbf{K}\,(k+1) \;=\; \mathbf{P}\,(k+1\,|k)\,\mathbf{H}^T\,(k+1)\,\times\left[\mathbf{H}\,(k+1)\,\mathbf{P}\,(k+1\,|k)\,\mathbf{H}^T\,(k+1)+\mathbf{R}\,(k+1)\right]^{-1} \tag{64}$$

where $\mathbf{H}(k+1) = \left.\frac{\partial \mathbf{h}(\mathbf{x}(t),\mathbf{u}_m(t),\boldsymbol{b},t)}{\partial \mathbf{x}(t)}\right|_{\mathbf{x}=\boldsymbol{\eta}_1}$, which also holds for the next steps

4) measurement update step:

$$\boldsymbol{\eta}_2 \;=\; \hat{\mathbf{x}}\,(k+1\,|k)+\mathbf{K}\,(k+1)\left[\mathbf{z}_m\,(k+1)-\mathbf{h}(\boldsymbol{\eta}_1)-\mathbf{H}\,(k+1)\left(\hat{\mathbf{x}}\,(k+1\,|k)-\boldsymbol{\eta}_1\right)\right] \tag{65}$$

$$\boldsymbol{\varepsilon} \;=\; \frac{\boldsymbol{\eta}_2-\boldsymbol{\eta}_1}{\boldsymbol{\eta}_2} \tag{66}$$

as long as $|\boldsymbol{\varepsilon}| > |\boldsymbol{\varepsilon}_{\text{crit}}|$, repeat steps 3 and 4 while after each iteration $\boldsymbol{\eta}_1 = \boldsymbol{\eta}_2$.*

5) update covariance matrix of state estimation error vector: as soon as $|\boldsymbol{\varepsilon}| \le |\boldsymbol{\varepsilon}_{crit}|$:

$$\hat{\mathbf{x}}\,(k+1\,|k+1) \;=\; \boldsymbol{\eta}_2 \tag{67}$$

$$\mathbf{P}\,(k+1|\,k+1) \;=\; \tilde{\mathbf{K}}\,(k+1)\,\mathbf{P}\,(k+1\,|k)\,\tilde{\mathbf{K}}^T\,(k+1)+\mathbf{K}\,(k+1)\,\mathbf{R}\,(k+1)\,\mathbf{K}^T\,(k+1) \tag{68}$$

with $\tilde{\mathbf{K}}\,(k+1) = [\mathbf{I}-\mathbf{K}\,(k+1)\,\mathbf{H}\,(k+1)]$

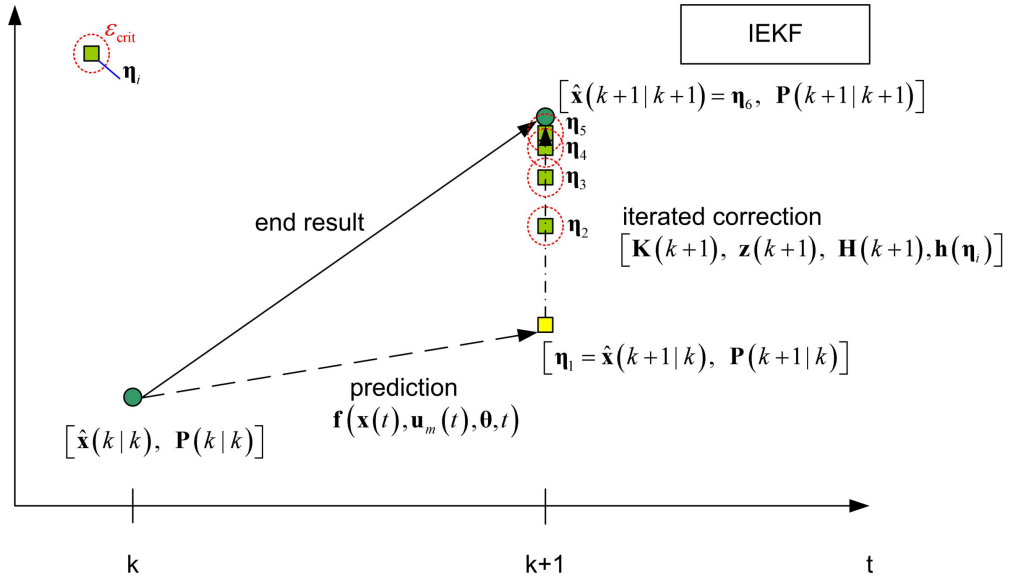


Fig. 4 Working principle of the Iterated Extended Kalman Filter

Fig. 4 illustrates the working principle of the Iterated Extended Kalman Filter, which is an augmentation of the regular Extended Kalman Filter. The correction step is performed iteratively, until the difference between two successive correction updates is below a predefined threshold.

E. Aircraft State Estimation results for the trajectory

The aforementioned IEKF as elaborated in Sec. IV.D and IV.C was applied to the aforementioned sample demonstration mission as introduced in Sec. IV.B. The values for the covariance matrices **Q** and **R** are defined by the corresponding sensor characteristics and are assumed known. These values are defined in Table 2.

The other settings for the IEKF are the maximum number of iterations $n_{\max} = 50$ for the correction step and the precision criterion metric for the iterations $\boldsymbol{\varepsilon}_{\text{crit}} = 10^{-6}$. Furthermore it is assumed that the static position (x_i, y_i, z_i) of

*In theory, this loop continues till the difference condition is satisfied, but in practice often a maximum number of iterations is specified in order to prevent an infinite loop.

input noise		output noise		
accelerometer	rate gyro	optical sensor		integrating gyro
$\sigma_x^2 = 1 \cdot 10^{-4}$	$\sigma_p^2 = 3 \cdot 10^{-8}$	$\sigma_R^2 = 1$	$\sigma_{\dot{R}}^2 = 1$	$\sigma_\phi^2 = 3 \cdot 10^{-8}$
$\sigma_y^2 = 1 \cdot 10^{-4}$	$\sigma_q^2 = 3 \cdot 10^{-8}$	$\sigma_\gamma^2 = 1$	$\sigma_{\dot{\gamma}}^2 = 1$	$\sigma_\theta^2 = 3 \cdot 10^{-8}$
$\sigma_z^2 = 1 \cdot 10^{-4}$	$\sigma_r^2 = 3 \cdot 10^{-8}$	$\sigma_\lambda^2 = 1 \cdot 10^{-8}$	$\sigma_{\dot{\lambda}}^2 = 1$	$\sigma_\psi^2 = 3 \cdot 10^{-8}$

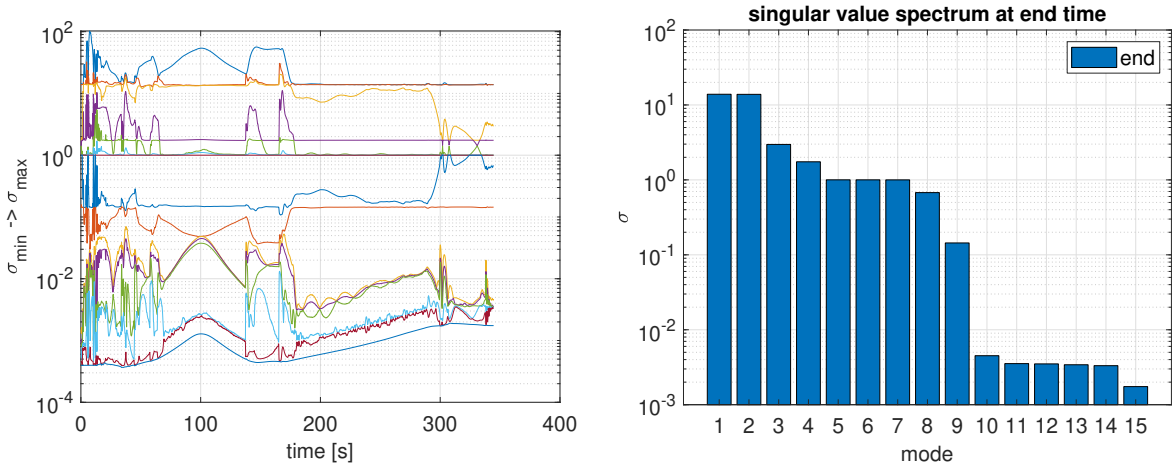
Table 2 Covariance values for Q and R matrices

the feature point i is known, and that all biases for accelerometers and rate gyros are zero. The initial state covariance matrix is defined as: $\mathbf{P}_0 = \text{diag}\,(100)$. For the initial estimated state $\mathbf{x}_0$, the position, velocity and heading are assumed known. Furthermore, the update rate of the correction step takes place at one fifth of the rate of the prediction step.

In order to analyze the reconstructibility of the problem, the observability matrix O was constructed:

$$O = \begin{bmatrix} \mathbf{H} & \mathbf{HF} & \mathbf{HF}^2 & \cdots & \mathbf{HF}^{n-1} \end{bmatrix}^T \tag{69}$$

Full reconstructibility of all the states is analyzed by verifying if the observability matrix O has full rank. Another metric is the singular value based condition number $\gamma = \frac{\sigma_{\max}}{\sigma_{\min}}$ of the observability matrix O. A value of γ of an order less than 10^8 shows that this problem is well defined.



(a) Time histories of all singular values

(b) End values of all singular values

Fig. 5 Singular values of the observability matrix

Numerical analysis has shown that rank $(O) = 15$, which is full rank. Besides, Fig. 5 gives more information on the singular value decomposition of the observability matrix O. Fig. 5(a) shows how the set of singular values σ evolves over time. It is clear that by performing maneuvers, while flying the mission profile, information is flowing into the observability matrix, which forces the singular values to converge towards each other. The problem becomes better defined. Fig. 5(b) shows the spectrum of singular values σ at the end time. The spread between the largest and smallest singular values is not excessive. The condition number here is $\gamma = \frac{\sigma_{\max}}{\sigma_{\min}} = \frac{13.8862}{0.0017} \approx 8 \cdot 10^3$, which indicates that the problem is well defined and thus all states should be relatively well identifiable given the numerical information in the sensor measurements.

Throughout the flown mission profile, the number of iterations per time sample was recorded. The result is shown in Fig. 6. Mostly, no iteration was needed at all. In a few time steps, one extra iteration was necessary, especially in the beginning. This time range corresponds with the time when some estimated states were still converging. Only at one time instant, two iterations were needed, but never more. Of course, the number of iterations depends on the threshold

value of the precision criterion metric $\varepsilon_{\text{crit}}$. In general, for lower values of this metric (higher precision), more iterations will be needed per time sample.

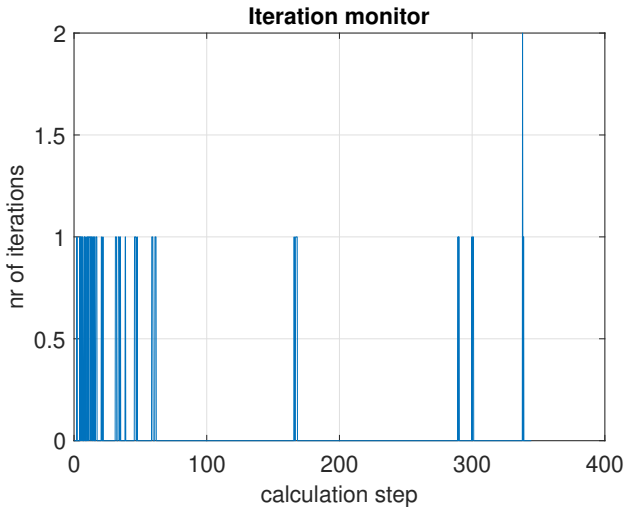


Fig. 6 Number of iterations in the IEKF correction step per time sample

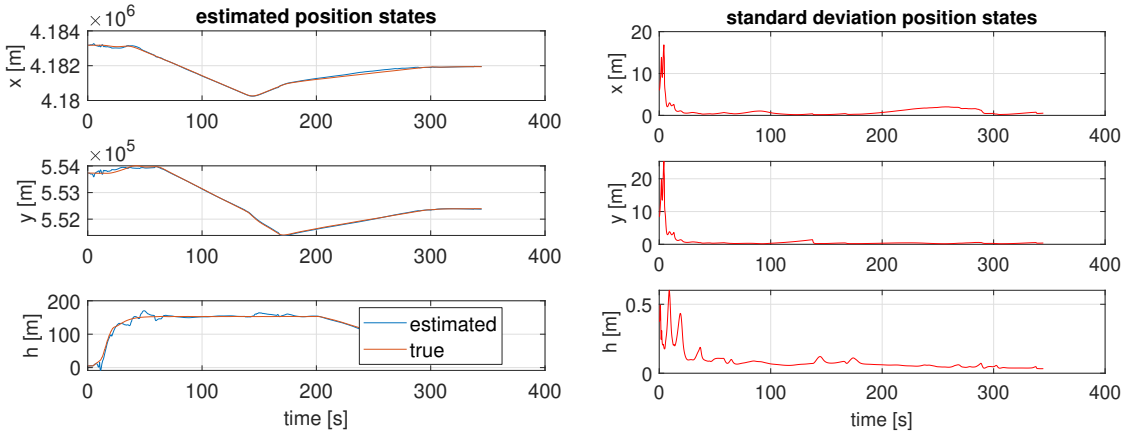
Fig. 7 — 10 show the estimates for the different states and their corresponding standard deviations, namely position (Fig. 7), velocity (Fig. 8), attitude (Fig. 9) and sensor biases (Fig. 10) respectively. Fig. 7 — 8 show that the first 50s are needed for convergence of the position and velocity states. This is most obvious for the lateral speed component v in Fig. 8, since this state has the narrowest magnitude range. Fig. 9 shows that the attitude states converge immediately. These variables are directly available in the measurements, although perturbed with output noise, while the position and velocity states are only indirectly present in the measurements through the optical sensor readings. The decreasing trends of the standard deviations confirm the convergence. Within the first 50s, also the sensor bias estimates converge to their true values of zero, as shown in Fig. 10.

This application example is an initial and very preliminary proof of concept of how optical sensor measurements can fill the gap when GPS data is unreliable due to lost satellite signals in an urban environment. The intention is to continue and scale up this work in future research up to more relevant examples in real life. However, another important aspect in this set up is the use of an adaptive mechanism for switching between GPS-based and feature point based state estimation that relies on a criterion that indicates the 'information richness' of what the optical system can see, as explained in Sec. III. An application example of this adaptive mechanism is given in Sec. IV.F.

F. Adaptive Mechanism Performance

The complete information content measure J as defined by Eq. (11) in Sec. III was applied on the scenario described in Sec. IV. The resulting values of the content measure J are shown in Fig. 11. In order to interpret and evaluate the information that these values represent, a deeper analysis was performed of a series of key point values as marked and numbered in Fig. 11.

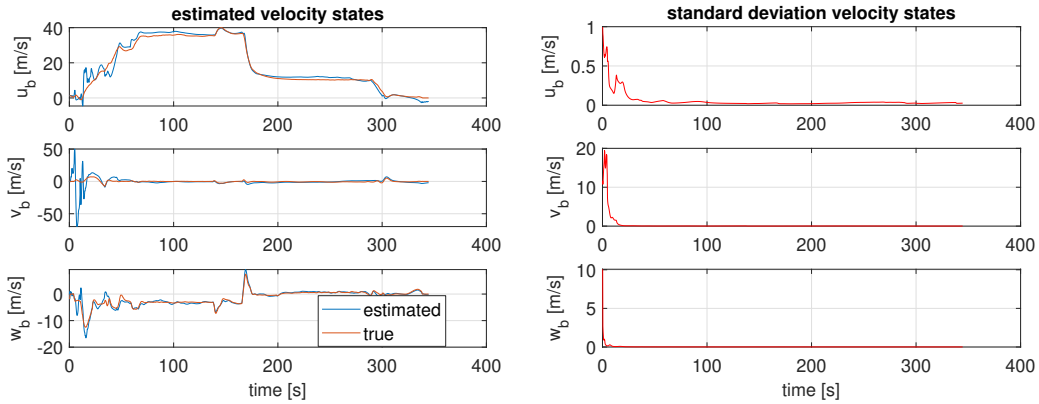
The corresponding locations of these key point values are also marked and numbered in the vehicle's trajectory as shown in Fig. 12, and in addition the relevant images as seen by the optical system at these locations are shown in Fig. 13. Correlating the information from Figs. 11, 12 and 13 shows that the general trends in the values for the content measure J have a physical meaning. The high value of J at key point 1 around the time of takeoff is caused by the rich scenery of the skyline of San Francisco, seen from ground level, as shown in Fig. 13(a). During initial climb, there is a sharp drop in the J value, which corresponds to much of the skyline sinking below the horizon due to the higher elevation vantage point as shown in Fig. 13(b). Between key points 3 and 4, there appears to be an above average noise level in the measure values. Most probably, this is caused by the specific situation where the vehicle is flying above the water surface of the bay, which contains relatively limited information for the optical system, as shown in Figs. 13(c) and 13(d). This can cause specific challenges for the measure calculation routine. Key points 5, 9 and 16 represent very low measure values in Fig. 11, which correspond to straight and level flight segments at or close to cruise altitude as shown in Fig. 12. Figs. 13(e), 13(i) and 13(p) confirm that there is barely any optical gradient information above



(a) Estimated vehicle position states

(b) Standard deviations of position states

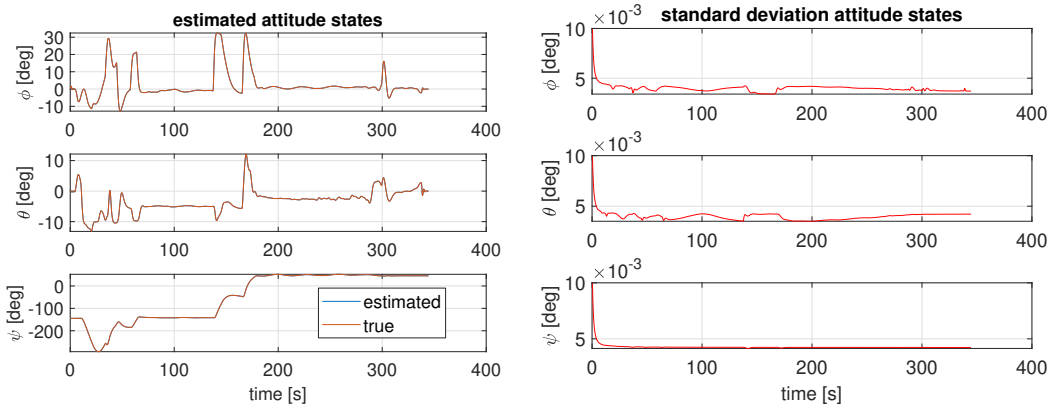
Fig. 7 Estimated values and standard deviations for vehicle position states



(a) Estimated vehicle velocity states

(b) Standard deviations of velocity states

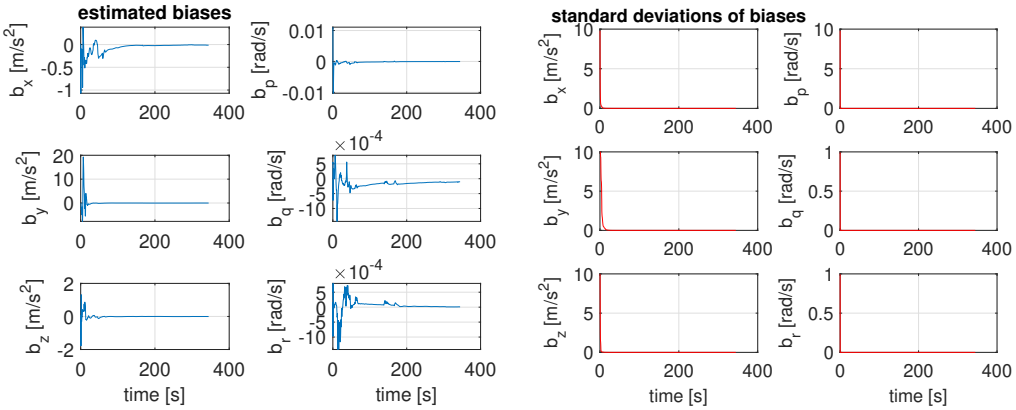
Fig. 8 Estimated values and standard deviations for vehicle velocity states



(a) Estimated vehicle attitude states

(b) Standard deviations of attitude states

Fig. 9 Estimated values and standard deviations for vehicle attitude states



(a) Estimated sensor biases

(b) Standard deviations of sensor biases

Fig. 10 Estimated values and standard deviations for sensor biases

the horizon at these key points and at this altitude. The clusters of key points 6-7-8, 10-11-12 and 13-14-15 exhibit slightly larger measure values in Fig. 11, which correspond to level turn flight segments at cruise altitude as shown in Fig. 12. Furthermore, Figs. 13(f)-13(g)-13(h), 13(j)-13(k)-13(l) and 13(m)-13(n)-13(o) confirm the bank angles in these turns via the tilted horizon. This tilted horizon generates some optical gradient information at these key points and altitude, but not significantly. Note that the information content measure J is compensated for pitch attitude angles θ as explained in Sec. III, but not yet for bank angle ϕ, since the resulting non-rectangular image shape causes difficulties for calculating the J-value. As a consequence, one can also notice the slightly larger J-values at larger bank angles in the middle of the turn, such as at key points 7, 11 and 14, as compared to the slightly lower bank angles earlier or later in the turns, such as at key point pairs 6 and 8, 10 and 12, 13 and 15. These variations in the content measure should not have an impact on the decision to switch to another state estimator. Only the J-values corresponding to key points 17 and 18 show a significant increase in Fig. 11. These key points are in the final descent as shown in Fig. 12, and images 13(q) and 13(r) confirm that the high rise buildings rise above the horizon as seen from these vantage points at lower altitudes. As illustrated in Fig. 11, it is possible to define a threshold that makes a distinction between moderate content measure values, which reflect the irrelevant bank angles in turns, and the significantly larger content measure values, which reflect the lower altitude portions of the flight where relevant optical gradients are caused by the scenery of the skyline. For this specific scenario, a threshold value of 0.4 works to make this distinction.

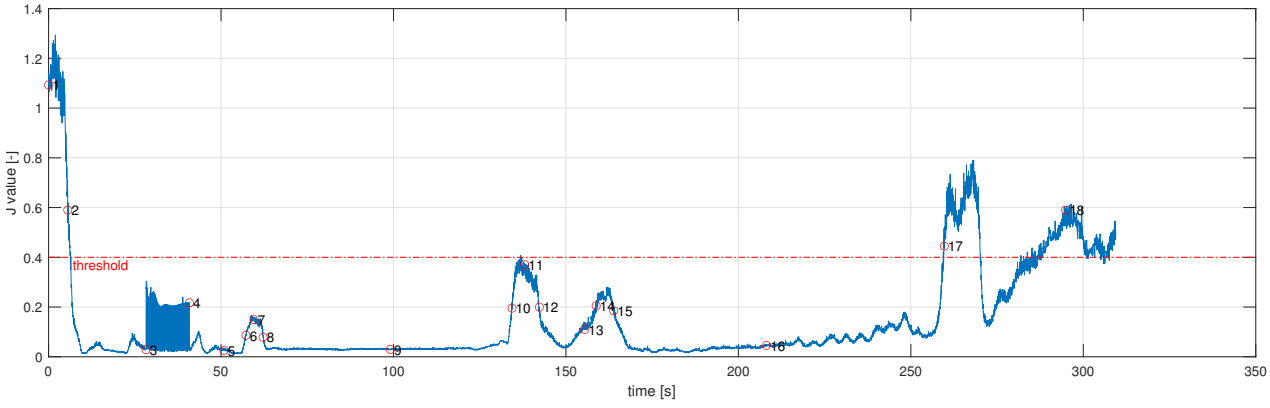
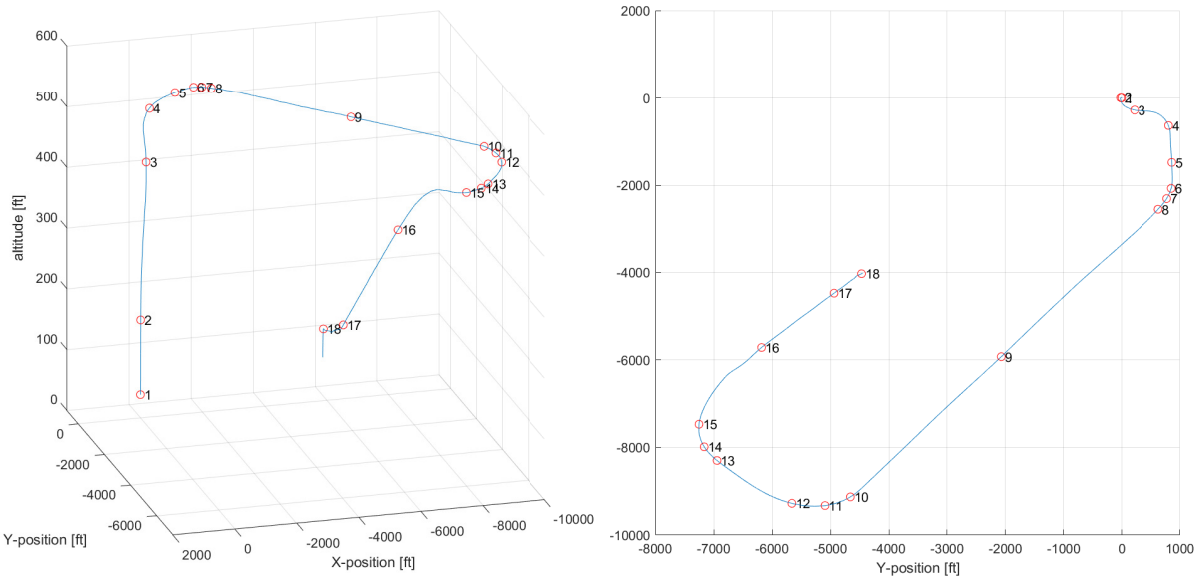


Fig. 11 J-values calculated throughout the sample flight with values of key images shown in Fig. 13

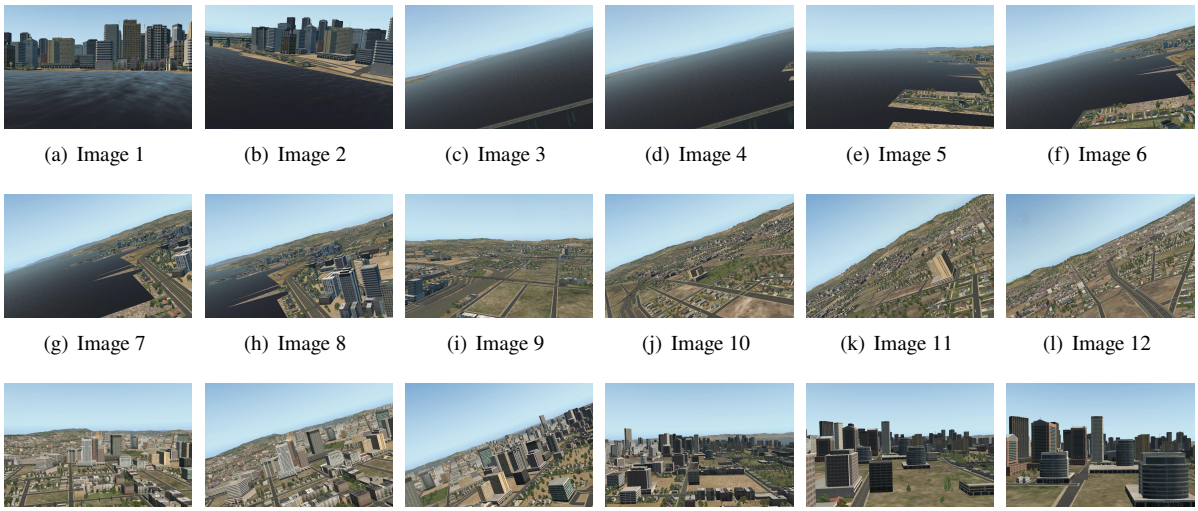
This adaptive mechanism allows to switch between the feature point based state estimator, as defined in Sec. IV, and a conventionally used GPS-based state estimator. The GPS-based state estimator as used in this application example is



(a) 3D trajectory with locations of key images

(b) 2D trajectory with locations of key images

Fig. 12 Trajectory of the sample flight with locations of key images shown in Fig. 13



(a) Image 1 (b) Image 2 (c) Image 3 (d) Image 4 (e) Image 5 (f) Image 6

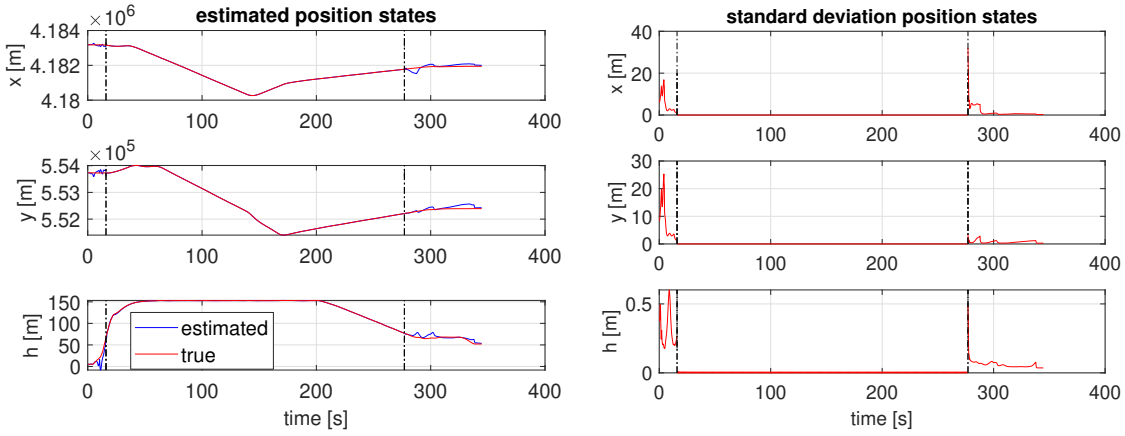
(g) Image 7 (h) Image 8 (i) Image 9 (j) Image 10 (k) Image 11 (l) Image 12

(m) Image 13 (n) Image 14 (o) Image 15 (p) Image 16 (q) Image 17 (r) Image 18

Fig. 13 Key images corresponding to the J-values marked in Fig. 11 and locations marked in Fig. 12

discussed in detail in Ref. [52]. The switching logic works as follows. When the GPS data are valid (direct line of sight with sufficient number of satellites), the GPS-based state estimator is preferred. When GPS data are invalid, e.g. in an urban environment where the field of view of the sky is significantly shielded by nearby high-rise buildings, the adaptive mechanism as described in Sec. III and as illustrated in Fig. 11 indicates if enough optical gradient information is available in the field of view of the optical system to rely on feature point based state estimation. If this information content measure J exceeds a certain predefined threshold J^*, the latter state estimator is preferred above the GPS-based state estimator. In the application example shown here, state estimation starts with the feature point based Kalman Filter,

up to around 16s in the flight, where the field of view of the sky widens and the J-value drops below the threshold. At that point, the GPS-based Kalman Filter takes over for the remainder of the flight, until around 270 s into the flight, during the approach, when the vehicle enters the airspace in between the high rise buildings, blocking a wide field of view of the sky. This is where the feature point based state estimator takes over again. The switch after takeoff involves an initialization of the Kalman filter where the initial state is calculated with the GPS measurements at that time instant (since the feature point based Kalman filter didn't converge yet at that point) and the initial covariance is a diagonal matrix with large values on the diagonal (typically in the order of magnitude 10^2. The switch during approach initializes that Kalman filter with the current state values of the GPS-based Kalman filter (which are converged), and the initial covariance is a diagonal matrix with large values on the diagonal (typically in the order of magnitude 10^3. The results of the state estimation process with filter switching is shown in Figs. 14–15, presented in a similar way as in the previous application example in Sec. IV.E. The vertical black dash-dotted lines represent the time instants when the switching between Kalman Filters takes place. In this specific example, the short time span in the first 16s is not long enough for the feature point based Kalman Filter to converge. As soon as the content measure allows, the GPS-data helps the state estimator to converge quickly. Towards the end of the flight, the feature point based estimator takes over again and converges quickly, however especially the velocity components shown in Fig. 15 experience large initial deviations. Convergence follows relatively quickly in the next few seconds. During this short time span, the IEKF needs a significant higher number of iterations to converge, as is shown in Fig. 16. Where one iteration per time step was sufficient during the initial convergence phase in both previous time spans for both state estimators, the maximum number of iterations is needed here for initial convergence, although relatively briefly. 50 is the maximum number of iterations allowed, in order to avoid infinite loops or frame overruns. Overall, it is clear from Fig. 14–15 that GPS data should be preferred above optical information whenever they are available, because of their accuracy and the faster convergence of its filter.



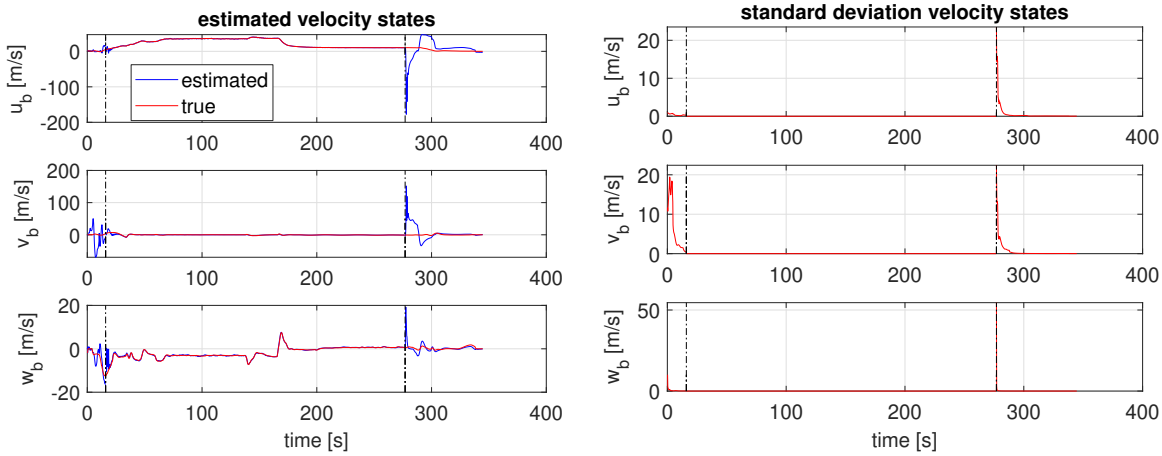
(a) Estimated vehicle position states

(b) Standard deviations of position states

Fig. 14 Estimated values and standard deviations for vehicle position states with combined filters and switching

V. Conclusions and recommendations

An adaptive method was developed to iteratively fuse the information provided by multiple sensors to enable autonomous urban air mobility type operations. One state estimation strategy relies on the combination of inertial and GPS measurements. Alternatively, another state estimator combines measurements from inertial and optical sensors such as camera, LIDAR, etc. which provide feature points with uncertainties. The adaptive method uses an information content measure based on optical gradients in the images provided by the optical sensors. This measure compensates for pitch attitude angle changes of the vehicle. Outside an urban environment and at sufficient altitude, valid GPS data can be used safely. In the airspace within an urban environment and below a critical altitude where the field of view of the sky is limited by high-rise buildings (also called 'urban canyon'), GPS data becomes unreliable. Initial experiments have shown that this measure provides a useful indication that sufficiently rich information content is available in the optical sensor measurements to allow for state estimation based on these optical sensor measurements. These experiments have



(a) Estimated vehicle velocity states

(b) Standard deviations of velocity states

Fig. 15 Estimated values and standard deviations for vehicle velocity states with combined filters and switching

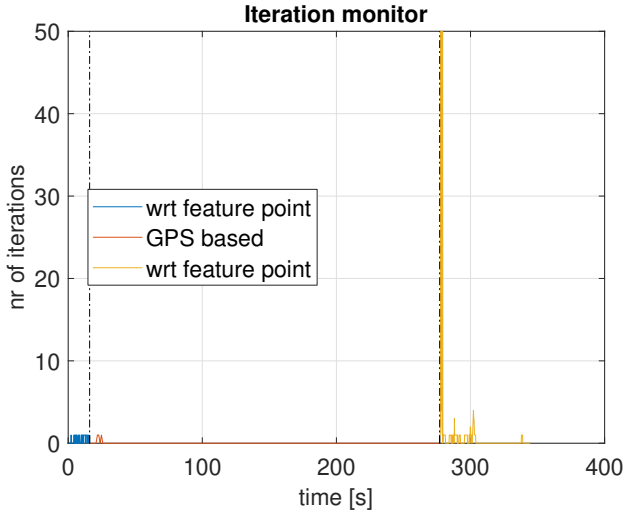


Fig. 16 Number of iterations in the IEKF correction step per time sample with combined filters and switching

also demonstrated that switching between the two different aforementioned state estimators, based on this information content measure, works satisfactorily. It was also found that whenever GPS data as well as sufficiently rich optical sensor information content are available, GPS data should be preferred because of their higher accuracy and faster filter convergence.

Future research will also look at compensation for bank angles in the information content measure calculation, and on the development of a more robust and adaptive threshold to switch between both aforementioned state estimators. In a next step, this position and velocity information, which is estimated independent of the availability of GPS data, will be used by another set of Kalman Filters, making use of the same optical sensors, for the purpose of object tracking. This is expected to be necessary for the purpose of detection and avoidance of uncooperative objects (objects without ADS-B or some other form of position reporting to other traffic), which is one of the important challenges in the context of perception within Urban Air Mobility applications.

Acknowledgment

This work was performed under the NASA Aeronautics Research Mission Directorate (ARMD), Transformative Tools and Technologies (TTT) Project.

References

[1] Amazon. Amazon Prime Air. *Available: https://www.amazon.com/Amazon-Prime-Air/b?ie=UTF8& node=8037720011*, Accessed 2020.

[2] Aurora. Aurora Tests E-VTOL Air Taxi for Uber. *Available: http://aviationweek.com/awincommercial/ aurora-tests-evtol-air-taxi-uber*, Accessed 2020.

[3] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart. Robust visual inertial odometry using a direct EKF-based approach. *In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, page 298?304, 2015.

[4] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser. Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving. *IEEE Transactions on Intelligent Vehicles*, 2(3):194–220, 2017.

[5] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, I. D. Reid J. Neira, and J. J. Leonard. Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.

[6] F. Castanedo. A Review of Data Fusion Techniques, Article ID 704504. *The Scientific World Journal*, pages 1–19, October 2013.

[7] G. Chowdhary, E. N. Johnson, D. Magree, A. Wu, and A. Shein. GPS-denied indoor and outdoor monocular vision-aided navigation and control of unmanned aircraft. *Journal of Field Robotics*, 30(3):415?438, May 2013.

[8] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte Carlo Localization for Mobile Robots. *In Proc. of the IEEE International Conference on Robotics and Automation*, 2:1322–1328, 1999.

[9] T. Dieterle, F Particke, L. Patino-Studencki, and J. Thielecke. Sensor Data Fusion of LIDAR with Stereo RGB-D Camera for Object Tracking. *In Proc. of the IEEE Sensors Conference, doi: 10.1109/ICSENS.2017.8234267*, October 2017.

[10] S. Ebcin and M. Veth. Tightly-Coupled Image-Aided Inertial Navigation Using the Unscented Kalman Filter. *Air Force Institute of Technology, Dayton, OH, Tech. Rep.*, 2007.

[11] K. Eckenhoff, P. Geneva, and G. Huang. Closed-Form Preintegration Methods for Graph-Based Visual-Inertial Navigation. *International Journal of Robotics Research*, 38(5):563–586, 2019.

[12] G. Fink, M. Franke, A. F. Lynch, K. Robenack, and B. Godbolt. Visual Inertial SLAM: Application to Unmanned Aerial Vehicles. *IFAC PapersOnLine*, 50(1):1965–1970, 2017.

[13] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. On-Manifold Preintegration for Real-Time Visual-Inertial Odometry. *IEEE Transactions on Robotics*, 33(1):1–21, 2017.

[14] S. Hening, C. Ippolito, K. Krishnakumar, V. Stepanyan, and M. Teodorescu. 3D LiDAR SLAM Integration with GPS/INS for UAVs in Urban GPS-Degraded Environments. *In Proc. of the AIAA SciTech Forum, Grapevine, Texas, 9 - 13 January*, 2017.

[15] G. Huang. Visual-Inertial Navigation: A Concise Review. *arXiv:1906.02650v1 [cs.RO] 6 Jun 2019*, pages 1–16, June 2019.

[16] V. Indelman, S.Williams, M. Kaess, and F. Dellaert. Information Fusion in Navigation Systems via Factor Graph Based Incremental Smoothing. *Robotics and Autonomous Systems*, 61(8):721–738, 2013.

[17] S. J. Julier and J. K. Uhlmann. New Extension of the Kalman Filter to Nonlinear Systems. *in AeroSense'97*, pages 182–193, 1997.

[18] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[19] J. Kim, D. S. Han, and B. Senouci. Radar and Vision Sensor Fusion for Object Detection in Autonomous Vehicle Surroundings. *In Proc. of the 10th International Conference on Ubiquitous and Future Networks (ICUFN), doi:10.1109/ICUFN.2018.8436959*, July 2018.

[20] E H. Kivelevitch, G. Dionne, T. Roose, P Arora, B. Fanous, R. Salvo, V. Pellissier, R. Losada, and R. Gentile. Sensor Fusion Tools in Support of Autonomous Systems. *In Proc. of the AIAA SciTech Forum, San Diego, California*, 7-11 January 2019.

[21] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. Keyframe-Based Visual-Inertial Odometry Using Nonlinear Optimization. *International Journal of Robotics Research*, 34(3):314–334, 2015.

[22] G. Loianno, M. Watterson, and V. Kumar. Visual Inertial Odometry for Quadrotors on Se(3). *In Proc. of the EEE International Conference on Robotics and Automation (ICRA)*, pages 1544–1551, May 2016.

[23] A. Makni, A. Y. Kibangou, and H. Fourati. Data Fusion-Based Descriptor Approach for Attitude Estimation Under Accelerated Maneuvers. *Asian Journal of Control, doi:10.1002/asjc.2084*.

[24] I. Mourikis, N. Trawny, S. I. Roumeliotis, A. E. Johnson, A. Ansar, and L. Matthies. Vision-aided inertial navigation for spacecraft entry, descent, and landing. *IEEE Transactions on Robotics*, 25(2):264–280, 2009.

[25] NHTSA. Automated Vehicle Safety. *Available: https://www.nhtsa.gov/technology-innovation/automated -vehicles-safety*, Accessed 2020.

[26] T. Qin, P. Li, and S. Shen. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.

[27] F. Rosique, P. J. Navarro, C. Fernandez, and A. Padilla. Survey of Capabilities and Gaps in External Perception Sensors for Autonomous Urban Air Mobility Applications, doi:10.3390/s19030648. *Sensors*, 19(3):1–29, February 2019.

[28] K. Shish, N. Cramer, G. Gorospe, T. Lombaerts, K. Kannan, and V. Stepanyan. A Systematic Review of Perception System and Simulators for Autonomous Vehicles Research. *To be presented in Proc. of the virtual AIAA SciTech Forum*, 11-21 January 2021.

[29] A. Shaghaghian and P. Karimaghaee. Improving GPS/INS Integration Using FIKF-Filtered Innovation Kalman Filter. *Asian Journal of Control; doi:10.1002/asjc.1931*, 21(4):1671–1680, July 2019.

[30] T. Shan and B. Englot. LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. *In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

[31] Joan Sola. Quaternion Kinematics for the Error-state Kalman Filter. *arXiv:1711.02508v1 [cs.RO]*, 3 November 2017.

[32] A. Soloviev. Tight Coupling of GPS, Laser Scanner, and Inertial Measurements for Navigation in Urban Environments. *In Proc. of the EEE/ION Position, Location and Navigation Symposium*.

[33] H. Strasdat, J. M. M. Montiel, and A. J. Davison. Real-time Monocular SLAM: Why Filter. *In Proc. of the International Conference on Robotics and Automation*, pages 2657–2664, 2010.

[34] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar. Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight. *IEEE ROBOTICS AND AUTOMATION LETTERS, arXiv:1712.00036v3 [cs.RO]*, pages 1–8, January 2018.

[35] S. Thrun and J. J. Leonard. *Simultaneous Localization And Mapping, in Springer Handbook of Robotics, pp. 871?889*. Springer, 2008.

[36] Uber. Uber Elevate. *Available: https://www.uber.com/info/elevate*, Accessed 2020.

[37] M. R. Walter, R. M. Eustice, and J. J. Leonard. Exactly Sparse Extended Information Filters for Feature-Based SLAM. *The International Journal of Robotics Research*, 26(4):335–359, May 2007.

[38] S. Weiss and R. Siegwart. Real-time metric state estimation for modular vision-inertial systems. *In Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[39] Google X. Project Wing. *Available: https://x.company/projects/wing*, Accessed 2020.

[40] H. Zhang, Y. Liu, J. Tan, and N. Xiong. RGB-D SLAM Combining Visual Odometry and Extended Information Filter. *Sensors, doi:10.3390/s150818742*, 15:18742–18766, 2015.

[41] J. Zhang and S. Singh. Laser?visual?inertial Odometry and Mapping With High Robustness and Low Drift. *Journal of Field Robotics, doi:10.1002/rob.21809*, 35:1242–1264, 2018.

[42] Z. Zhang and Y. Shan. Incremental Motion Estimation through Modified Bundle Adjustment. *In Proc. of the International Conference on Image Processing*, pages 343–346, 2003.

[43] T. Alldieck, C. H. Bahnsen, and T. b. Moeslund. Context-Aware Fusion of RGB and Thermal Imagery for Traffic Monitoring. *Sensors*, 16(11), 2016, doi:10.3390/s16111947.

[44] K. G. Larkin. Reflections on Shannon Information: In search of a natural information-entropy for images. *arXiv:1609.01117v1 [cs.IT]*, 2016.

[45] J. P. Pluim, J. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2000, Springer*, pages 452–461, 2000.

[46] F. M. Ramos, R. R. Rosa, C. R. Neto, and A. Zanandrea. Generalized complex entropic form for gradient pattern analysis of spatio-temporal dynamics. *Phys. Stat. Mech. Its Appl. 283*, pages 171–174, 2000.

[47] C. E. Shannon and W. Weaver. The Mathematical Theory of Communication. *University of Illinois Press, Urbana, IL*, 1949.

[48] R. Shams, P. Sadeghi, and R. A. Kennedy. "Gradient intensity: A new mutual information based registration method. *In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[49] C. Silva, W. R. Johnson, E. Solis, M. D. Patterson, and K. R. Antcliff VTOL Urban Air Mobility Concept Vehicles for Technology Development. *2018 Aviation Technology, Integration, and Operations Conference, American Institute of Aeronautics and Astronautics*, 2018. doi:10.2514/6.2018-3847.

[50] T. Lombaerts, J. Kaneshige, S. Schuet, B. L. Aponso, K. H. Shish, and G. Hardy. Dynamic Inversion based Full Envelope Flight Control for an eVTOL Vehicle using a Unified Framework. *In Proc. of the AIAA SciTech Forum, Orlando, Florida, 6 - 10 January*, 2020, doi:10.2514/6.2020-1619.

[51] A.H. Jazwinski. Stochastic Processes and Filtering Theory. Academic Press, New York, 1970.

[52] T. Lombaerts. Fault Tolerant Flight Control based on a Physical Approach. *PhD thesis, Delft University of Technology*, 2010, `http://resolver.tudelft.nl/uuid:538b0174-fe84-43af-954d-02f256b2ec50`