

# Metric Visual-Inertial Navigation System Using Single Optical Flow Feature

Sammy Omari<sup>1</sup> and Guillaume Ducard<sup>2</sup>

**Abstract**—This paper introduces a state estimation framework that allows estimating the attitude, full metric speed and the orthogonal metric distance of an IMU-camera system with respect to a plane. The filter relies only on a single optical flow feature as well as gyroscope and accelerometer measurements. The underlying assumption is that the observed visual feature lies on a static plane. The orientation of the observed plane is not required to be known a priori and is also estimated at runtime. The estimation framework fuses visual and inertial measurements in an Unscented Kalman Filter (UKF). The theoretical limitations of the UKF are investigated using a nonlinear observability analysis based on Lie-derivatives.

Simulations using realistic sensor noise values successfully demonstrate the performance of the filter as well as validate the findings of the observability analysis. It is shown that the state estimate is converging correctly, even in presence of substantial initial state errors. To the authors' knowledge, this paper documents for the first time the estimation of the heading and metric distance to a wall with no range sensor, relying solely on optical flow as the only exteroceptive sensing modality.

## I. INTRODUCTION

In recent years, a considerable body of work was presented for position or velocity estimation of robots, in particular unmanned aerial vehicles (UAV), in GPS-denied environments using vision sensors [1],[2],[3],[4]. Especially optical flow-based approaches have attracted a lot of attention since their implementation requires only little computational power and can be run at high update rates on light-weight, low-power microcontrollers [5]. Additionally, optical flow-based approaches are also able to handle scenes with only little contrast [5] - especially compared to feature-based SLAM frameworks that require high-contrast, salient features that can be tracked over longer periods of time [1].

In [6], the authors present a framework to fuse optical flow and inertial measurements to obtain the full metric speed in an unstructured environment by assuming constant scene depth. In [7], the authors use optical flow from a down-looking camera to control a UAV in hovering and vertical landing. They assume that the optical flow features lie on a horizontal plane whose normal can be estimated with the UAVs on-board IMU. Using this approach, they can compute the ratio  $\frac{v}{d}$  between the translational velocity  $v$  in the camera frame and orthogonal distance to the plane  $d$ . In [5], they resolve this ambiguity by using an ultrasound range sensor to measure the distance  $d$  to the plane. In [8], the authors estimate not only the ratio  $\frac{v}{d}$ , but also the normal vector of

the plane using optical flow by incorporating the continuous homography constraint.

In contrast to these approaches ([5]-[8]), we present a novel UKF-based observer to estimate both the full metric velocity and distance to the plane as well as its normal vector. The estimator makes no assumption on the orientation of the plane and does not rely on any distance or bearing sensors. Instead, only optical flow and inertial sensors are used as measurement input to the UKF. Using this setup, we can not only estimate the normal vector of the plane, but the full attitude of the IMU-camera system with respect to the plane.

The previously mentioned approaches require multiple optical flow features to estimate the scaled velocity through a continuous homography- or epipolar-based approach. In our scheme, we do not require this intermediate step and directly use each optical flow feature as a measurement input to the UKF. This allows us to properly propagate the uncertainty of the optical flow into the covariance of the state estimate. This also enables the design of a stochastic outlier detection mechanism since the predicted optical flow vector (with its covariance) can be directly compared to its corresponding measurement.

The structure of the paper is as follows. In section II, the setup of the system is described. Section III discusses the required sensors and the corresponding stochastic models. In section IV, the estimation scheme, based on the Unscented Kalman Filter is discussed. Simulations of the estimation scheme and the discussion of the algorithm and its limitation are found in section V.

Concerning the mathematical notation, all vectors and matrices are written in bold text. We use a skew symmetric matrix notation for the vector cross product  $\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_b$ . The canonical basis of  $\mathbb{R}^3$  is denoted  $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ .

## II. SYSTEM SETUP

In this section, we describe the setup of the system as depicted in Fig. 1.

The system consists of an IMU (gyroscope and accelerometer) rigidly fixed to a spherical camera. The body frame is denoted  $\{B\}$  and is attached to the IMU-camera system. Its origin coincides with the accelerometer of the IMU. The orientation of the body frame with respect to the inertial frame  $\{I\}$  is defined by the unit quaternion  $\mathbf{q}_i^b$ . Equivalently, the same orientation mapping can also be expressed using the rotation matrix  $\mathbf{C}$  defined by the quaternion  $\mathbf{q}_i^b$ . The position and translational velocity of the body frame, expressed in the inertial frame, are denoted  $\mathbf{r}_i \in \mathbb{R}^3$  and  $\mathbf{v}_i \in \mathbb{R}^3$  respectively. The body angular rates are denoted  $\boldsymbol{\omega} \in \mathbb{R}^3$ .

<sup>1</sup>S. Omari is with the Autonomous Systems Lab, ETH Zurich, Switzerland. [omaris@ethz.ch](mailto:omaris@ethz.ch)

<sup>2</sup>G. Ducard is with I3S UNS-CNRS, France. [ducard@i3s.unice.fr](mailto:ducard@i3s.unice.fr)

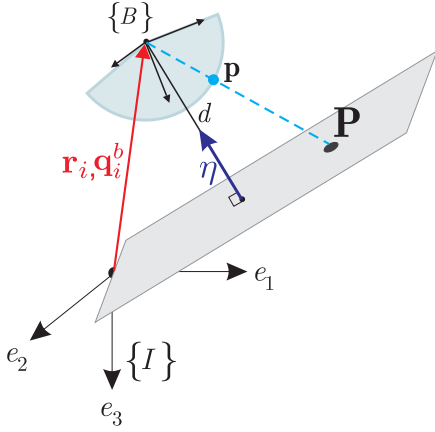


Fig. 1. Schematic representation of the system setup. The position and orientation (red) of the body frame  $\{B\}$  are described in the inertial coordinate frame  $\{I\}$ . The target plane passes through the origin of  $\{I\}$  and is described by its normal vector  $\eta$ . The projection of  $\mathbf{P}$  onto the spherical image plane of the camera is  $\mathbf{p}$ .

The camera is observing a target plane which contains visual features. We define  $\mathbf{P} \in \mathbb{R}^3$  to be the coordinate of a visual feature on the target plane expressed in the camera frame. Its corresponding coordinates projected on the spherical image plane of the camera are  $\mathbf{p} \in \mathbb{S}^2$ . The target plane is defined by its normal  $\eta \in \mathbb{S}^2$  expressed in the inertial frame.

For the derivation of the system equations, we make the following assumptions:

*Assumption 2.1:* The spherical camera has unit focal length and is rigidly fixed to the vehicle.

*Assumption 2.2:* The intrinsic camera calibration parameters, such as the focal length and the principal point are known. As a consequence, the camera that is used in this scheme does not necessarily have to be spherical. Using the calibration parameters, we can map the pixel coordinates of a regular pinhole-model-like camera to the spherical image plane, as shown in section III-C.

*Assumption 2.3:* For the sake of simplicity, we assume that the camera origin coincides with the body frame origin. In reality, we can extract the position and orientation of the camera with respect to the body frame origin using CAD or an offline estimation procedure. Using this position and orientation offset we would then have to adapt the measurement equations, as described in section III-D.

*Assumption 2.4:* We define the target plane to contain the origin of the inertial frame. This is a valid assumption since we do not have any measurements of a globally fixed inertial position (e.g. provided by GPS). In fact, we will show that we can only observe the relative orthogonal distance of the camera w.r.t. the plane. Therefore, we can set the origin of the plane arbitrarily.

*Assumption 2.5:* The target plane comprises at least one visual feature  $\mathbf{P}$  visible from the camera that can be tracked over two consecutive frames.

*Assumption 2.6:* The visual features and the target plane are considered to be stationary in the inertial frame. Therefore, the dynamics of the features in the camera frame only

depend on the vehicle dynamics.

### III. MEASUREMENT MODELS

This section discusses the required sensors and the corresponding stochastic measurement models. Throughout the paper, external disturbances and sensor noises are modeled as continuous Gaussian noise or as discrete Gaussian noise processes. This assumption makes it possible to use the Kalman Filter as an optimal least-square estimator (LSE).

#### A. Accelerometer Model

The accelerometer output  $\mathbf{a}_m \in \mathbb{R}^3$  is composed of the gravity vector  $\mathbf{g} = g \mathbf{e}_3 \in \mathbb{R}^3$  in the inertial frame, the true linear accelerations  $\mathbf{a} \in \mathbb{R}^3$  in the inertial frame, some slowly time-varying bias  $\mathbf{b}_a \in \mathbb{R}^3$  in the body frame and Gaussian white noise  $\mathbf{n}_a \in \mathbb{R}^3$ . The bias  $\mathbf{b}_a$  is modeled as a random walk process governed by the Gaussian white noise  $\mathbf{n}_{b_a} \in \mathbb{R}^3$ . We can write:

$$\mathbf{a} = \mathbf{C}^T (\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) + \mathbf{g} \quad (1)$$

$$\dot{\mathbf{b}}_a = \mathbf{n}_{b_a} \quad (2)$$

The noises  $\mathbf{n}_a$  and  $\mathbf{n}_{b_a}$  are specified by the corresponding covariance matrices  $\mathbf{Q}_a$  and  $\mathbf{Q}_{b_a}$  respectively. As discussed in [9], the covariance matrices can be evaluated by examining the measured Allan variances. For the sake of simplicity, we assume that both covariance matrices are diagonal, each with identical entries.

#### B. Gyroscope Model

The gyroscopes on the IMU measures the angular velocity  $\omega_m \in \mathbb{R}^3$  in the body frame which is composed of the true angular speed  $\omega \in \mathbb{R}^3$ , some slowly time-varying bias  $\mathbf{b}_\omega \in \mathbb{R}^3$  in the body frame and some Gaussian white noise  $\mathbf{n}_\omega \in \mathbb{R}^3$ . As in the case of accelerometers, each bias is modeled as a random walk process. We can write:

$$\omega = \omega_m - \mathbf{b}_\omega - \mathbf{n}_\omega \quad (3)$$

$$\dot{\mathbf{b}}_\omega = \mathbf{n}_{b_\omega} \quad (4)$$

As in the case of accelerometers, gyro noises are specified by their corresponding diagonal covariance matrices  $\mathbf{Q}_\omega$  and  $\mathbf{Q}_{b_\omega}$ .

#### C. Camera Model

Since we assume that the camera is intrinsically calibrated, we can map camera pixel coordinates  $\phi \in \mathbb{R}^2$  to their corresponding position  $\mathbf{P} \in \mathbb{R}^3$  in the camera frame up to a scaling factor  $\lambda$

$$\lambda \mathbf{P} = \text{cam}(\phi, \theta) \quad (5)$$

using some generic camera model  $\text{cam}$ , and the intrinsic camera calibration parameter vector  $\theta$ . For more information on intrinsic camera calibration, the user is referred to [10].

Since we only know  $\mathbf{P}$  up to scale, we simply normalize it to obtain the feature vector  $\mathbf{p} \in \mathbb{S}^2$  on the spherical image plane with focal length  $f = 1$

$$\mathbf{p} = \frac{\mathbf{P}}{|\mathbf{P}|}. \quad (6)$$

From now on, we consider the camera to be a bearing vector sensor, i.e. it directly outputs feature vectors  $\mathbf{p}$ .

#### D. Optical Flow

Since the camera is fixed to the vehicle and the observed target point is considered to be stationary, the feature vector  $\mathbf{p}$  inherits the dynamics of the vehicle. It can be shown [11] that the dynamics of the feature vector  $\mathbf{p}$  can be expressed as

$$\mathbf{u} = -[\boldsymbol{\omega}]\mathbf{p} - \pi_{\mathbf{p}} \frac{\mathbf{v}_b}{|\mathbf{P}|} \quad (7)$$

where  $\mathbf{v}_b$  is the translational velocity of the camera expressed in the body frame<sup>1</sup> and the projection operator is defined as  $\pi_{\mathbf{p}} = (\mathbf{I}_3 - \mathbf{p}\mathbf{p}^T)$ . The vector  $\mathbf{u} \in \mathbb{R}^3$  is the optical flow corresponding to the feature vector  $\mathbf{p}$  and can be extracted from two consecutive images using a wide range of algorithms, such as Lucas-Kanade [12] or Horn-Schunck [13].

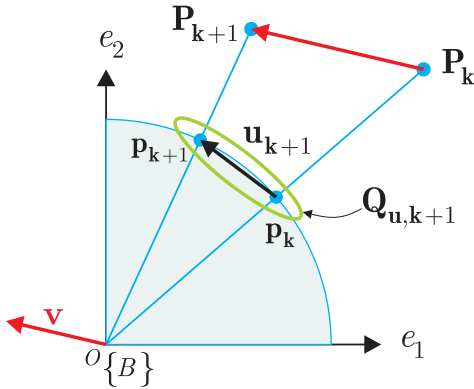


Fig. 2. Schematic representation of the optical flow computation as seen from the body frame  $\{B\}$ . The camera, moving with velocity  $\mathbf{v}$ , observes a visual feature  $\mathbf{P}$  whose projection on the spherical image plane is  $\mathbf{p}$ . The optical flow  $\mathbf{u}_{k+1}$  is calculated from the projected features at  $t_k$  and  $t_{k+1}$ . Its covariance ellipse is  $\mathbf{Q}_{\mathbf{u},k+1}$ .

The measured optical flow  $\mathbf{u}_m$  is perturbed by Gaussian noise  $\mathbf{n}_u \in \mathbb{R}^3$  specified by the covariance matrix  $\mathbf{Q}_u$

$$\mathbf{u} = \mathbf{u}_m - \mathbf{n}_u \quad (8)$$

Since the camera output is not continuous, the optical flow vector  $\mathbf{u}_{k+1}$  at time  $t_{k+1}$  is approximated by  $\mathbf{u}_{k+1} \approx \frac{\mathbf{p}_{k+1} - \mathbf{p}_k}{t_{k+1} - t_k}$ , as depicted in Fig. 2. Since the feature vectors during both time steps are on the unit sphere, the noise of the optical flow vector is constrained to the surface of the sphere. Therefore, we approximate the covariance matrix  $\mathbf{Q}_{\mathbf{u},k+1}$  as a flat ellipsoid tangent to the unit sphere in direction of the average feature vector  $\frac{(\mathbf{p}_k + \mathbf{p}_{k+1})}{2}$ , as depicted in Fig. 2.

#### E. Target Plane Model

Since we assume that the target plane contains the origin of the inertial frame, it is sufficient to describe the plane by its normal vector  $\boldsymbol{\eta}$  expressed in the inertial frame. Since the normal vector is of unit length, it has two degrees of freedom. We can therefore parameterize the normal vector using two angles  $\alpha$  and  $\beta$ . We can write

<sup>1</sup>If the camera origin were not coinciding with the body frame origin, we would have to replace  $\mathbf{v}_b$  by the camera velocity  $\mathbf{v}_c$  using the relation  $\mathbf{v}_c = \mathbf{v}_b + [\boldsymbol{\omega}]\mathbf{r}_b^c$  where  $\mathbf{r}_b^c$  is the baseline from the body frame- to the camera frame origin.

$$\boldsymbol{\eta}(\alpha, \beta) = \begin{bmatrix} \sin(\alpha) \cos(\beta) \\ -\sin(\beta) \\ \cos(\alpha) \cos(\beta) \end{bmatrix} \quad (9)$$

For the sake of readability, we replace  $\boldsymbol{\eta}(\alpha, \beta)$  by  $\boldsymbol{\eta}$  from now on. We consider the angles to be slowly time-varying and model them as a random walk processes

$$\dot{\alpha} = n_\alpha \quad (10)$$

$$\dot{\beta} = n_\beta \quad (11)$$

with noises specified by their corresponding covariance terms  $Q_\alpha$  and  $Q_\beta$ ;

#### IV. UNSCENTED KALMAN FILTER

For space reasons, we omit the discussion of the actual UKF equations and refer the reader to [14] for a detailed introduction to UKF for inertial navigation. Instead, we discuss the required building blocks of the UKF in detail: the minimal state vector representation, the prediction and measurement equations and the derivation of the process- and measurement noise covariance matrices.

##### A. State Dynamics and State Vector Representation

Using the sensor models defined in (1) - (4) and the dynamics of the plane (10) - (11), the state dynamics can be written as:

$$\dot{\mathbf{r}}_i = \mathbf{v}_i \quad (12)$$

$$\dot{\mathbf{v}}_i = \mathbf{C}^T(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) + \mathbf{g} \quad (13)$$

$$\dot{\mathbf{q}}_i^b = \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega}_m - \mathbf{b}_\omega - \mathbf{n}_\omega)\mathbf{q}_i^b \quad (14)$$

$$\dot{\mathbf{b}}_\omega = \mathbf{n}_{b_\omega} \quad (15)$$

$$\dot{\mathbf{b}}_a = \mathbf{n}_{b_a} \quad (16)$$

$$\dot{\alpha} = n_\alpha \quad (17)$$

$$\dot{\beta} = n_\beta \quad (18)$$

The matrix  $\boldsymbol{\Omega}(\boldsymbol{\omega}) \in \mathbb{R}^{4 \times 4}$  relates the body angular rates to the corresponding quaternion rate by

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} 0 & -\boldsymbol{\omega}^T \\ \boldsymbol{\omega} & -[\boldsymbol{\omega}] \end{bmatrix}. \quad (19)$$

As a minimal state vector representation, we define the 18-element state vector

$$\mathbf{x} = [\mathbf{r}_i \ \mathbf{v}_i \ \mathbf{q}_i^b \ \mathbf{b}_\omega \ \mathbf{b}_a \ \alpha \ \beta]^T. \quad (20)$$

##### B. Error Dynamics and Process Noise Covariance Matrix

To derive the process noise covariance matrix, the error dynamics of the system (12) - (18) have to be derived first. Special care has to be taken for the error representation of the quaternion since it is a four-element vector but only has three degrees of freedom. Therefore its error should also be represented as a 3-dimensional rotation vector  $\delta\phi$ . We follow the convention in [14] and define the relation between the true quaternion  $\mathbf{q}$  with its estimate  $\hat{\mathbf{q}}$  and the error rotation vector  $\delta\phi$  by

$$\mathbf{q} = \delta\mathbf{q} \otimes \hat{\mathbf{q}} \quad (21)$$

$$\delta\mathbf{q} \approx [1 \ \frac{1}{2}\delta\phi]^T \quad (22)$$

where  $\otimes$  is the quaternion multiplication operator.

Following the approach in [9] and [15] the error dynamics are now derived using the approximation (21, 22) and neglecting all higher-order terms:

$$\dot{\delta \mathbf{r}}_i = \delta \mathbf{v}_i \quad (23)$$

$$\dot{\delta \mathbf{v}}_i = -\mathbf{C}^T [\mathbf{a}_m - \mathbf{b}_a] \delta \phi - \mathbf{C}^T \delta \mathbf{b}_a - \mathbf{C}^T \mathbf{n}_a \quad (24)$$

$$\dot{\delta \phi} = -[\boldsymbol{\omega}_m - \mathbf{b}_\omega] \delta \phi - \delta \mathbf{b}_\omega - \mathbf{n}_\omega \quad (25)$$

$$\delta \dot{\mathbf{b}}_\omega = \mathbf{n}_\omega \quad (26)$$

$$\delta \dot{\mathbf{b}}_a = \mathbf{n}_a \quad (27)$$

$$\dot{\delta \alpha} = n_\alpha \quad (28)$$

$$\dot{\delta \beta} = n_\beta \quad (29)$$

Using the error dynamics (23) - (29), the discrete process covariance matrix  $\mathbf{Q}_k$  at time  $t_k$  is obtained after tedious calculations [9]:

$$\mathbf{Q}_k = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{C} & 0 & 0 \\ \mathbf{B} & \mathbf{E} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{F} & 0 & 0 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{G} & \mathbf{H} & \mathbf{0}_3 & 0 & 0 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{J} & \Delta t \mathbf{Q}_{b_\omega} & \mathbf{0}_3 & 0 & 0 \\ \mathbf{K} & \mathbf{L} & \mathbf{0}_3 & \mathbf{0}_3 & \Delta t \mathbf{Q}_{b_a} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta t Q_\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \Delta t Q_\beta \end{bmatrix} \quad (30)$$

with

$$\begin{aligned} \mathbf{A} &= \frac{\Delta t^3}{3} \mathbf{Q}_a + \frac{\Delta t^5}{20} \mathbf{Q}_{b_a}, & \mathbf{G} &= \Delta t \mathbf{Q}_\omega + (\hat{\Gamma}_{3,k} + \hat{\Gamma}_{3,k}^T) \mathbf{Q}_{b_\omega} \\ \mathbf{B} &= \frac{\Delta t^2}{2} \mathbf{Q}_a + \frac{\Delta t^4}{8} \mathbf{Q}_{b_a}, & \mathbf{H} &= -\hat{\Gamma}_{2,k}^T \mathbf{Q}_{b_\omega} \\ \mathbf{C} &= -\frac{\Delta t^3}{6} \hat{\mathbf{C}}_k^T \mathbf{Q}_{b_a}, & \mathbf{J} &= -\mathbf{Q}_{b_\omega} \hat{\Gamma}_{2,k} \\ \mathbf{E} &= \Delta t \mathbf{Q}_a + \frac{\Delta t^3}{3} \mathbf{Q}_{b_a}, & \mathbf{K} &= -\frac{\Delta t^3}{6} \mathbf{Q}_{b_a} \hat{\mathbf{C}}_k^T \\ \mathbf{F} &= -\frac{\Delta t^2}{2} \hat{\mathbf{C}}_k^T \mathbf{Q}_{b_a}, & \mathbf{L} &= -\frac{\Delta t^2}{2} \mathbf{Q}_{b_a} \hat{\mathbf{C}}_k^T \end{aligned}$$

using

$$\hat{\Gamma}_{n,k} = \mathbf{I}_3 \Delta t^n + \frac{\Delta t^{n+1}}{(n+1)!} [\boldsymbol{\omega}_{m,k} - \hat{\mathbf{b}}_{\omega,k}] + \frac{\Delta t^{n+2}}{(n+2)!} ([\boldsymbol{\omega}_{m,k} - \hat{\mathbf{b}}_{\omega,k}])^2 \quad (31)$$

### C. Measurement Equation

We now have to express the optical flow equation (7) using the estimated state vector  $\hat{\mathbf{x}}$ . The body angular rates  $\hat{\boldsymbol{\omega}}$  can be written as

$$\hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_m - \hat{\mathbf{b}}_\omega. \quad (32)$$

The velocity in the body frame  $\hat{\mathbf{v}}_b$  can be written as

$$\hat{\mathbf{v}}_b = \hat{\mathbf{C}} \hat{\mathbf{v}}_i. \quad (33)$$

The computation of  $|\hat{\mathbf{P}}|$  is a bit more involved. By inspection of Fig. 1, we express the orthogonal distance of the camera to the target plane  $\hat{d}$  using the position of the camera in the inertial frame  $\hat{\mathbf{r}}_i$  and the normal vector of the plane  $\hat{\boldsymbol{\eta}}$  as

$$\hat{d} = |\hat{\mathbf{r}}_i^T \hat{\boldsymbol{\eta}}| \quad (34)$$

Equivalently, since we assume that the observed visual feature is lying on the plane, we can express  $\hat{d}$  as a function of the feature coordinate  $\hat{\mathbf{P}}$  in the body frame and the plane normal vector  $\hat{\boldsymbol{\eta}}$  in the world frame

$$\hat{d} = |\hat{\mathbf{P}}^T (\hat{\mathbf{C}} \hat{\boldsymbol{\eta}})| \quad (35)$$

Combining equation (6), (34) and (35), we can eliminate  $\hat{d}$  and solve for  $|\hat{\mathbf{P}}|$ :

$$|\hat{\mathbf{P}}| = \left| \frac{\hat{\mathbf{r}}_i^T \hat{\boldsymbol{\eta}}}{\mathbf{p}^T (\hat{\mathbf{C}} \hat{\boldsymbol{\eta}})} \right|. \quad (36)$$

Now, by plugging equations (32), (33) and (36) in the optical flow equation (7), we obtain

$$\hat{\mathbf{u}} = -[(\boldsymbol{\omega}_m - \hat{\mathbf{b}}_\omega)] \mathbf{p} - \left| \frac{\mathbf{p}^T (\hat{\mathbf{C}} \hat{\boldsymbol{\eta}})}{\hat{\mathbf{r}}_i^T \hat{\boldsymbol{\eta}}} \right| \pi_{\mathbf{p}} (\hat{\mathbf{C}} \hat{\mathbf{v}}_i) \quad (37)$$

Since the body angular rates  $\boldsymbol{\omega}$  are not part of the state vector  $\hat{\mathbf{x}}$ , we cannot not directly use the optical flow as measurement input to the UKF. Instead, by exploiting the distributive property of the vector product, we define the derotated optical flow  $\bar{\mathbf{u}}$  as a new measurement to the UKF

$$\hat{\mathbf{u}} = \bar{\mathbf{u}} - [\mathbf{p}] \boldsymbol{\omega}_m = -[\mathbf{p}] \hat{\mathbf{b}}_\omega - \left| \frac{\mathbf{p}^T (\hat{\mathbf{C}} \hat{\boldsymbol{\eta}})}{\hat{\mathbf{r}}_i^T \hat{\boldsymbol{\eta}}} \right| \pi_{\mathbf{p}} (\hat{\mathbf{C}} \hat{\mathbf{v}}_i) \quad (38)$$

The derotated optical flow can now be predicted using only the state vector  $\hat{\mathbf{x}}$ . In case multiple optical flow vectors are extracted at one timestep, the innovation term  $\bar{\mathbf{u}}$  of the UKF can be written as

$$\bar{\mathbf{u}} = \begin{bmatrix} (\mathbf{u}_{m,1} - [\mathbf{p}_1] \boldsymbol{\omega}_m) - \hat{\mathbf{u}}_1 \\ \vdots \\ (\mathbf{u}_{m,N} - [\mathbf{p}_N] \boldsymbol{\omega}_m) - \hat{\mathbf{u}}_N \end{bmatrix} \quad (39)$$

One potential issue with the measurement equation (38) is the division by  $|\hat{\mathbf{r}}_i^T \hat{\boldsymbol{\eta}}|$  since the position state can evolve unconstrained and this term could become 0. The corresponding physical representation would be that the projection center of the camera is touching the plane, which is not possible. However, due to transients in the Kalman filter, this term can approach 0. Therefore, in the implementation, we replace this term with  $\max(\epsilon, |\hat{\mathbf{r}}_i^T \hat{\boldsymbol{\eta}}|)$  with  $\epsilon \ll 1$ .

### D. Measurement Covariance Matrix

The derotated optical flow measurement  $\bar{\mathbf{u}}_m$  is perturbed by the pixel noise of the optical flow algorithm  $\mathbf{n}_u$  as well as the noise introduced through the derotation term  $[\boldsymbol{\omega}] \mathbf{p}$

Since we assume that  $\boldsymbol{\omega}$  is perturbed by additive noise and we take the cross product of  $\boldsymbol{\omega}$  with  $\mathbf{p}$  which is unit norm, the noise that is introduced by the derotation term is bounded by  $\mathbf{n}_\omega$ . Therefore, following Eq. 8, we can write

$$\bar{\mathbf{u}} = \bar{\mathbf{u}}_m - \mathbf{n}_u + \mathbf{n}_\omega. \quad (40)$$

The covariance matrix of the derotated optical flow measurement is the sum of the gyroscope covariance  $\mathbf{Q}_\omega$  and covariance of the optical flow algorithm  $\mathbf{Q}_u$ . If multiple optical flow vectors are observed at one timestep, the measurement covariance matrices can be stacked together diagonally as

$$\mathbf{R}_k = \begin{bmatrix} \mathbf{Q}_{u,1} + \mathbf{Q}_\omega & \dots & \mathbf{0}_3 \\ \vdots & \ddots & \vdots \\ \mathbf{0}_3 & \dots & \mathbf{Q}_{u,N} + \mathbf{Q}_\omega \end{bmatrix} \quad (41)$$

### E. Outlier Rejection

Before using the optical flow measurements in the update step of the UKF, we want to robustly detect and reject outlier measurements. While a continuous homography based rejection approach, as proposed in [21], could be used to this end, we use a statistical approach that takes into account the covariance of the predicted optical flow measurement. We follow the approach in [16] and employ a Mahalanobis-distance test to detect outlier measurements. We compute the Mahalanobis distance as

$$\chi^2 = (\bar{\mathbf{u}}_m - \hat{\mathbf{u}})^T \mathbf{S}^{-1} (\bar{\mathbf{u}}_m - \hat{\mathbf{u}}). \quad (42)$$

The matrix  $\mathbf{S}$  is the innovation covariance of the UKF as defined by eq. (77) in [15] and can be seen as a weighting term depending on the confidence in the current prediction. Measurements that exceed a certain threshold  $\chi_{\text{th}}^2$  are rejected and are not employed in the update step of the estimation scheme.

### F. Observability Analysis

The analysis of the observability characteristics of the nonlinear UKF reveals the theoretical limitations of the observer.

A tool for the analysis of the observability properties of a nonlinear Kalman filter is the observability rank condition based on Lie derivatives [17]. Due to space limitations, we refer the reader to [18], [19], and [20] for a in-detail discussion of nonlinear observability analysis for Kalman filters. Thus, we only discuss the most important outcomes of the analysis.

To facilitate the analysis, a coordinate transform from the inertial frame into a new, plane-centric coordinate frame is introduced. The coordinate transform is described by a rotation in which the vector  $\mathbf{e}_1$  of the plane coordinate frame coincides with normal vector of the plane  $\boldsymbol{\eta}$ ,  $\mathbf{e}_2$  is in direction  $\mathbf{g} \times \boldsymbol{\eta}$  and  $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$ .

Computing the observability matrix  $\mathcal{O}$  using this new coordinate transform reveals that the position in direction  $\mathbf{e}_1$  is fully observable while the position in direction of  $\mathbf{e}_2$  and  $\mathbf{e}_3$  is not observable. The remaining states, namely the velocity and attitude with respect to the plane as well as the accelerometer and gyroscope biases in the body frame are also observable. However, these states are only observable under the condition that (1) the system is excited by accelerations on at least two independent axis and (2) there is at least one optical flow measurement with its feature vector  $\mathbf{p}$  not parallel to the velocity vector in the body frame. The requirement for the acceleration can be explained using an intuitive argument: If the system observes a constant optical flow, the observer cannot distinguish the ratio between velocity and distance.

However, while the attitude is observable with respect to the plane frame, this does not imply that it is fully observable in the inertial frame. In fact, the rotation around the gravity vector in the inertial frame is not observable. This is due to the fact that the system has no measurement of an absolute, inertial yaw angle. Therefore, the yaw angle of the system in

the inertial frame is not observable. The roll- and pitch angles of the attitude in the inertial frame are observable since the gravity vector acts as reference. If a virtual measurement of the yaw-angle is introduced as an additional measurement to the analysis, as suggested in [20], the unobservable manifold of the attitude in the inertial frame became fully observable.

The case where the normal vector coincides with gravity requires special consideration. In this case, the yaw component of the attitude of the body frame w.r.t. the plane frame is not observable anymore. This can be explained by the well known property from the TRIAD method [22], which states that two independent reference vectors are required to fully describe the attitude of a system. In the case where the gravity vector coincides with the normal vector, we only have one reference vector and as a consequence, the yaw angle of the body frame w.r.t. the plane frame and the yaw bias in the body frame become unobservable.

## V. RESULTS

To evaluate the performance of the proposed estimation scheme and to validate the findings of the observability analysis, the scheme is implemented in simulation. To generate the ground truth as well as the sensor readings, we simulate a UAV flight in our in-house UAV flight simulator, as depicted in Fig. 3. The simulated flight consists of a

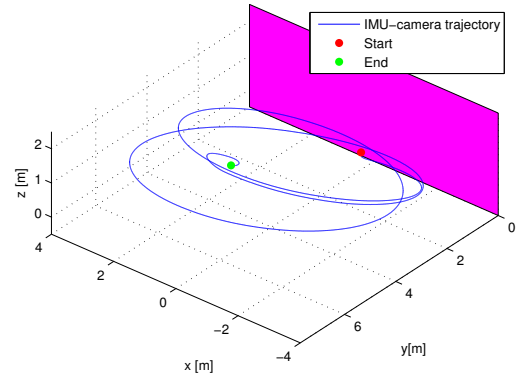


Fig. 3. Trajectory of IMU-camera system.

high-dynamic phase where the UAV is tracking a ellipse-like trajectory at a velocity of up to 2.5 m/s with a maximal acceleration of  $1.5\text{m/s}^2$  at a distance between 0.5 – 7.5m from the observed plane. After 30 seconds of circling, the UAV position controller is set to hover-mode to examine the observer performance in low-dynamics flight. In simulation, the UAV is equipped with an accelerometer and a gyroscope as well as a front-looking camera that is pointing towards the plane with a normal vector  $\boldsymbol{\eta} = [0 \ 1 \ 0]^T$ . The inertial sensor readings are generated in this simulation at 100 Hz and are perturbed with realistic noise values present during UAV flight ( $\sigma_\omega = 3 \text{ deg/s}$ ,  $\sigma_a = 0.5 \text{ N}$ ). We use a simulated global shutter camera with  $150^\circ$  FOV and a resolution of  $752 \times 480$  running at 30 Hz. We use the camera model from [10] with coefficients of a real calibrated PointGrey Firefly camera. While a single optical flow feature ensures the observability of the state, in practice, due to measurement noise, at least 50 features have to be extracted at every timestep to ensure proper convergence of the observer. For the simulation, we

use 75 optical flow features perturbed by Gaussian noise with a standard deviation of 1.5 pixels. Additionally, we add 20 outlier optical flow features by generating correct optical flow vectors and then simply reversing their signs.

The initial conditions of the filter state are perturbed with offsets described in Table I.

TABLE I  
INITIAL CONDITIONS OF FILTER

State	Initial Value	True Value
Distance to Plane	2.5 m	0.5 m
Velocity	[0.2 0.2 0.2] m/s	[0 0 0] m/s
Roll, Pitch, Yaw of $\mathbf{q}_v^b$	[5 -5 20] deg	[0 0 0] deg
Gyro Bias	[0 0 0] rad/s	[0.03 0.03 -0.03] rad/s
Accelerometer Bias	[0 0 0] N	[0.1 0.1 0.1] N
Plane Normal Vector	[ 0.42 0.89 0.13]	[0 1 0]

### A. Estimation of the UAV's Metric Velocity and Metric Distance to the Wall

In Fig. 4, the orthogonal distance of the UAV to the plane with its  $1\sigma$  bounds is depicted. The filter is able to correct for the initial state offset and converges to the true distance within 12 seconds. At  $t = 30$ s, when the UAV starts hovering, the distance starts diverging slowly. The observability analysis explains that the state is only fully observable if the system is excited with accelerations in at least two independent axis. Additionally, during hovering, the derotated optical flow is almost zero. This yields only the information that the ratio between the velocity and the distance to the plane is zero. Therefore, no information on the distance to the plane can be extracted from optical flow during hovering. The growing covariance bounds correctly account for this. The covariance bounds also nicely show the sensitivity of the filter w.r.t. the distance to the plane. As the distance becomes larger, the signal-to-noise ratio of the optical flow measurement becomes worse and therefore, the covariance of the distance estimate grows. The RMS error of the distance after convergence at  $t = 12$ s until the end of the high-dynamic flight phase at  $t = 30$ s is 0.060m and 0.095m if the preceding low-dynamics flight phase is included as well. In Fig. 5, the velocity of the system in the body

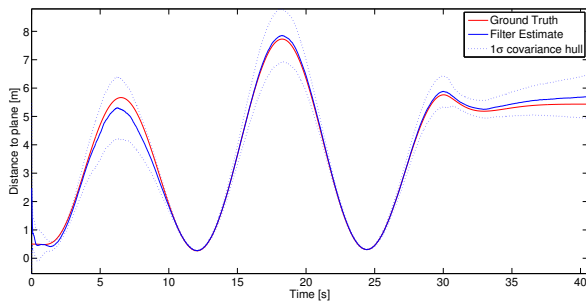


Fig. 4. Evolution of the estimate of the orthogonal distance to the plane with its corresponding  $1\sigma$  bounds.

frame is depicted. The velocity converges to the true value and remains converged even during hover flight. The RMS velocity error after convergence at  $t = 12$ s is [0.031 0.024 0.017] m/s.

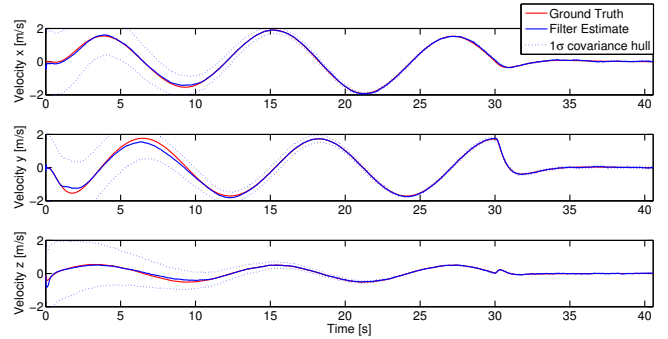


Fig. 5. Evolution of the estimated velocities expressed in the body frame with their corresponding  $1\sigma$  bounds.

### B. Estimation of the Gyroscope Bias

After  $t = 12$ s, the biases have and remain converged even during hover flight. The covariance bounds of the yaw bias is slightly larger than the roll- and pitch covariance bounds. This is due to the fact that the yaw bias uncertainty is correlated to the estimation of the normal vector of the plane. The evolution of the accelerometer bias is equivalent to the gyroscope bias and is omitted for space reasons. In Fig. 6, the estimates of the gyroscope biases are shown with their corresponding  $1\sigma$  bounds.

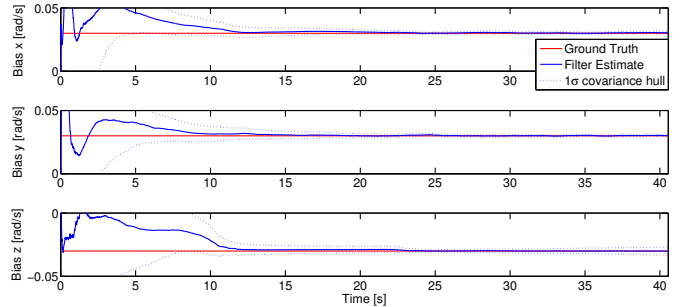


Fig. 6. Evolution of gyroscope biases with  $1\sigma$  covariance hull.

### C. Estimation of the Body Frame Attitude with Respect to Inertial- and Wall Frame

In Fig. 7, the estimation of the roll-, pitch- and yaw-angle of the UAV with respect to the inertial frame is depicted. The roll- and pitch angle converge to the ground truth as the gyroscope and accelerometer biases converge. The RMS error of the roll- and pitch angle after convergence at  $t = 12$ s is 0.8 and 0.9 degrees respectively. As explained in Section IV-F, the yaw-angle does not converge since no measurement of a heading vector in the inertial frame is available. However, following the observability analysis, the observer is able to estimate the attitude of the body frame with respect to the wall frame  $\mathbf{C}_b^w$ . The wall frame is defined in section IV-F. By defining the error rotation matrix  $\tilde{\mathbf{C}}_b^w = \mathbf{C}_b^w \tilde{\mathbf{C}}_b^{wT}$  it can be shown that  $\tilde{\mathbf{C}}_b^w \rightarrow \mathbf{C}_b^w$  if  $\text{tr}(\mathbf{I}_3 - \tilde{\mathbf{C}}_b^w) \rightarrow 0$  [23]. In Fig. 8, the evolution of the rotational error term  $\text{tr}(\mathbf{I}_3 - \tilde{\mathbf{C}}_b^w)$  is depicted. As the error

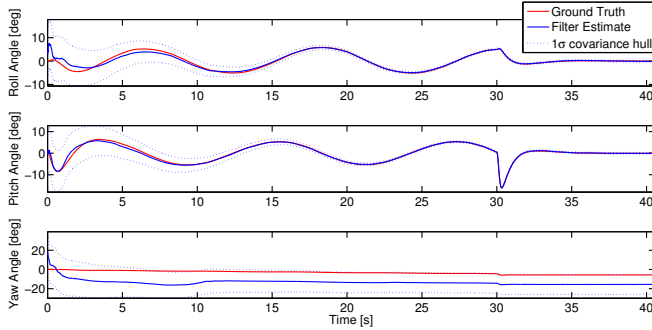


Fig. 7. Evolution of the roll- pitch and yaw angles of the attitude of the UAV in the inertial frame with the corresponding  $1\sigma$  bounds.

converges to zero, the estimate of the attitude of the body frame w.r.t. the wall frame converges as well. Since the plane

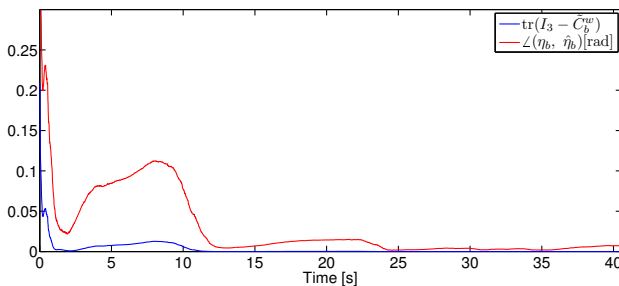


Fig. 8. Evolution of the error of the rotation matrix between wall- and body frame (blue) and angular error of the estimated plane normal vector  $\hat{\eta}_b$  expressed in the body frame.

normal vector  $\eta$  is only observable in the body frame (see section IV-F), we evaluate the angular error of the estimated normal vector in the body frame. As expected from the convergence of  $\hat{C}_b^w$ , the estimated normal vector converges too, as depicted in Fig. 8. The RMS angular error of the planes' normal vector in the body frame after convergence at  $t = 12s$  is 1.3 deg.

## VI. CONCLUSIONS

In this paper, a novel state estimation framework has been presented that allows estimating the attitude, full metric velocity and the orthogonal metric distance of an IMU-camera system w.r.t. a plane. The filter relies on optical flow as well as gyroscope and accelerometer measurements and does not require any range- or bearing sensors. The theoretical limitations of the filter have been investigated using a nonlinear observability analysis. Experiments in simulation successfully demonstrate the performance of the filter as well as validate the findings of the observability analysis. The simulation results successfully show convergence of the filter in presence of substantial initial state offsets and sensor noises present during the flight of a UAV.

This generic estimation framework can be used as a vision-aided navigation system in a wide range of applications, on a variety of robotic platforms. Due to the minimal sensor configuration, which is both light-weight and low-cost, this framework is an appealing choice to be used on weight-sensitive robotic platforms such as UAVs.

## REFERENCES

- [1] M. Bloesch, S. Weiss, D. Scaramuzza, R. Siegwart, "Vision Based MAV Navigation in Unknown and Unstructured Environments," Proc. of the IEEE/RSJ Int. Conf. on Robotics and Automation (ICRA), 2010, Anchorage, AK.
- [2] M. Achtelik, A. Bachrach, R. He, S. Prentice, N. Roy, "Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments," Unmanned Systems Technology XI., 2009, Orlando, FL.
- [3] M. Achtelik, S. Lynen, S. Weiss, L. Kneip, M. Chli, R. Siegwart, "Visual-Inertial SLAM for a Small Helicopter In Large Outdoor Environments," Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2012, Vilamoura, Portugal.
- [4] S. Weiss, D. Scaramuzza, R. Siegwart, "Monocular-SLAM-based navigation for autonomous micro helicopters in GPS-denied environments," Journal of Field Robotics, Vol. 28, No. 6, 2011, 854-874
- [5] P.-J. Bristeau, F. Callou, D. Vissiere, N. Petit, "The Navigation and Control technology inside the AR.Drone micro UAV," 18th IFAC World Congress, pp. 1477-1484, 2011, Milano, Italy.
- [6] S. Weiss, M.W. Achtelik, S. Lynen, M. Chli, R. Siegwart, "Real-time Visual-Inertial State Estimation and Self-Calibration of MAVs in Unknown Environments," Proc. of the IEEE/RSJ Int. Conf. on Robotics and Automation (ICRA), 2012, Saint-Paul, MN.
- [7] B. Herisse, F.-X. Russotto, T. Hamel, R. Mahony, "Hovering flight and vertical landing control of a VTOL Unmanned Aerial Vehicle using Optical Flow," Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2008, Nice, France.
- [8] V. Grabe, H.H. Buelthoff, P.R. Giordano, "On-board Velocity Estimation and Closed-loop Control of a Quadrotor UAV based on Optical Flow," Proc. of the IEEE/RSJ Int. Conf. on Robotics and Automation (ICRA), 2012, Saint-Paul, MN.
- [9] M. Bloesch, M. Hutter, M. Hoepflinger, S. Leutenegger, C. Gehring, C. Remy, R. Siegwart, "State Estimation for Legged Robots - Consistent Fusion of Leg Kinematics and IMU," Proc. of the 8th Robotics: Science and Systems Conference (RSS), 2012, Sydney, Australia.
- [10] D. Scaramuzza, A. Martinelli and R. Siegwart, "A Toolbox for Easily Calibrating Omnidirectional Cameras," Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2006, Beijing, China.
- [11] R. Mahony, P. Corke, T. Hamel, "Dynamic Image-Based Visual Servo Control Using Centroid and Optic Flow Features," J. Dyn. Sys., Meas., Control, Volume 130, Issue 1, 2008.
- [12] B.D Lucas, T. Kanade. "An iterative image registration technique with an application to stereo vision," Proc. of the 7th international joint conference on Artificial intelligence. 1981.
- [13] B. K. Horn, B.G. Schunck, "Determining optical flow," Artificial intelligence 17, No. 1, pp. 185-203, 1981.
- [14] J.L. Crassidis, F.L. Markley, "Unscented filtering for spacecraft attitude estimation," Journal of Guidance Control and Dynamics 26.4 pp. 536-542, 2003.
- [15] J.L. Crassidis, "Sigma-point Kalman filtering for integrated GPS and inertial navigation," IEEE Transactions on Aerospace and Electronic Systems, 42.2, pp. 750-756, 2006.
- [16] F.M. Mirazei, S.I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration," Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2007, San Diego, CA.
- [17] R. Hermann, A. Krener, "Nonlinear controllability and observability," IEEE Trans. Autom. Control, vol. AC-22, No. 5, pp 728-740, 1977.
- [18] A. Martinelli, "Vision and IMU Data Fusion: Closed-Form Solutions for Attitude, Speed, Absolute Scale, and Bias Determination," IEEE Transaction on Robotics, Vol. 28, No. 1, pp 40-60, 2012.
- [19] F.M. Mirazei, S.I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation," IEEE Transactions on Robotics, Vol. 24, No.5, 2008.
- [20] S. M. Weiss, "Vision Based Navigation for Micro Helicopters," Ph.D. dissertation, ETH Zurich, 2012.
- [21] V. Grabe, H.H. Buelthoff, P.R. Giordano, "Robust Optical-Flow Based Self-Motion Estimation for a Quadrotor UAV," Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2012, Vilamoura, Portugal.
- [22] H. Black, "A Passive System for Determining the Attitude of a Satellite," AIAA Journal 2 (7), pp. 1350-1351, 1963.
- [23] M.-D. Hua, "Contributions to the Automatic Control of Aerial Vehicles". Ph.D. thesis, University of Nice-Sophia Antipolis, 2009.