4DoF Drift Free Navigation Using Inertial Cues and Optical Flow

Stephan Weiss, Roland Brockers, Larry Matthies
Jet Propulsion Laboratory, California Institute of Technology
{stephan.weiss@ieee.org, (roland.brockers, lhm)@jpl.nasa.gov}

***Abstract*— In this paper, we describe a novel approach in fusing optical flow with inertial cues (3D acceleration and 3D angular velocities) in order to navigate a Micro Aerial Vehicle (MAV) drift free in 4DoF and metric velocity. Our approach only requires two consecutive images with a minimum of three feature matches. It does not require any (point) map nor any type of feature history. Thus it is an inherently failsafe approach that is immune to map and feature-track failures. With these minimal requirements we show in real experiments that the system is able to navigate drift free in all angles including yaw, in one metric position axis, and in 3D metric velocity. Furthermore, it is a *power-on-and-go* system able to online self-calibrate the inertial biases, the visual scale and the full 6DoF extrinsic transformation parameters between camera and IMU.**

I. Introduction and Related Work

Micro Aerial Vehicles (MAVs) have seen an increased popularity in recent years due to a wide range of new applications in reconnaissance, surveillance, search and rescue, and environmental monitoring. In particular, multicopter systems (e.g. quadrotors) have a distinct advantage in their hovering capabilities and agility to counteract strong winds. This agility, however, comes at a cost: quadrotors are inherently unstable in flight and require good state estimation and control to maintain a position or to perform defined maneuvers. Navigating such a system is particularly challenging since there is no "hold" function as compared to ground vehicles that can be entered as a safety regime (e.g. holding all actuators will not result in zero velocity but in a crash). Therefor, multi-rotor MAV require constant and accurate state estimation and control. Furthermore, a misalignment to gravity not only results in a velocity vector but in an acceleration. Similarly, simply integrating IMU accelerations for position hold will result in a crash due to noise and bias terms on the accelerometers. A minimal requirement to keep a quadrotor airborne is to have a continuous metric velocity[1] and a precise gravity aligned attitude estimate. The MAV may still drift in position and yaw, however, given an obstacle free area, the vehicle remains airborne.

There is a large body of work for indoor MAV control using motion tracking systems [1], [2], [3] and for outdoor operations using GPS signals [4], [5]. In contrast to these approaches that depend on external positioning information, this paper focuses on MAV control in environments where an external tracking setup is infeasible (e.g. large outdoor areas) and GPS signals my be corrupted or unavailable (e.g. in cities, caves etc.).

A. Control and Navigation Based on Maps

A popular approach is to control and navigate MAVs based on maps without the need of a motion capturing system or GPS. This is often done using known markers, pre-built maps or maps built on the fly using SLAM or keyframe based visual odometry (VO) techniques. Such approaches usually control the vehicle in its 6DoF pose (position and attitude). The drift in position and yaw is either very low or even eliminated by using known, world-fix structures. Sensors that are commonly used for map generation are laser scanners [6], [7] or cameras incorporating known markers [8], [9] or SLAM/VO techniques [10], [11]. Since laser scanners are still too heavy and power hungry for small quadrotor systems, our work focuses on vision based approaches. Common to all approaches using a map for motion estimation is, that the map can get corrupted or lost. In such a case recovery is difficult if not impossible and the vehicle is prone to crash.

B. Control and Navigation Without Maps and Feature History

In order to avoid the issue of a map loss, we follow the approach of not having any type of feature history except the feature matches between two consecutive images: i.e. optical flow (OF). A history free approach augments the algorithms robustness due to the independence on past readings. OF approaches without including 6D inertial measurements are presented in [12], [13]. While already showing the capabilities of OF for MAV navigation, these approaches act on a reactive manner to keep the vehicle away from ground or from obstacles. Reactive control to position-keep or trajectory-navigate a micro air vehicle is not sufficient since the unavailability of metric information can result in instability of these inherently unstable systems.

More recent work includes 6D inertial reading and successfully estimates not only the metric velocity [14], [15] but also inertial biases inter-sensor transformations and a gravity aligned attitude of the MAV [16] - the minimal requirements in order to keep a MAV airborne.

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

[1] if the velocity estimate is not metric, the controller cannot tell if the vehicle drifts with $1\frac{m}{s}$ or $\lambda * 1\frac{m}{s}$. This results in oscillation and instability.

C. Full Information Acquisition for Complete State Estimation

Our work is a continuation of previous work in [16] with the novelty that we use the full information provided by OF and inertial readings to achieve a complete vehicle state estimation not discarding any information. The main contribution of this work is to show in real-world tests that:

- we can estimate the metric distance to the over-flown terrain using only OF (i.e. only two consecutive images) and inertial readings. This directly leads to robust, metric terrain-following capabilities.
- with OF and inertial readings only, it is additionally possible to estimate the terrain inclination towards gravity. This leads to the elimination of the global yaw drift and enables the MAV to navigate in 3D terrain.
- a concise analysis for the visual scale propagation in the Extended Kalman Filter (EKF) framework leads to improved results compared to our previous.

Compared to commercial products, our approach does not require an active sensor (e.g. ultra sonic altimeter). This reduces the weight and power consumption of the overall sensor suite. More important, our approach does not require a gravity aligned ground plane for theoretically correct functioning. In fact, our approach gains on performance in inclined terrain as explained later.

The remainder of the paper is organized as follows. In Section II we briefly describe the computation of the optical measurements used later in the EKF framework. These measurements are the arbitrarily scaled 3D camera velocity vector and the terrain plane-parameters in the camera frame. Section III describes our EKF framework showing the capability of additionally estimating the terrain inclination and metric terrain distance to the MAV. We also discuss here the novel approach for the visual scale propagation within the EKF. Section IV presents the real world results to show the functioning of the proposed approach. We conclude the paper in Section V.

II. Computation of Optical Measurements

This section is dedicated to the theoretical background on how to compute the scaled visual 3D velocity vector, the terrain inclination as perceived in the camera frame and the scaled distance from the camera to the terrain. These measurements will be used in the next section as an update step of the EKF framework estimating all states of the system and the terrain. As in our previous work [16] we make use of the continuous epipolar constraint and the additional information of the angular velocities measured by the Inertial Measurement Unit (IMU) to compute the optical measurements.

A. 3D Velocity Computation up to Unified Scale

As described in [17], the continuous motion of a feature with respect to the camera is

$$\dot{\mathbf{X}}(t) = \lfloor \vec{\omega}(t) \rfloor \mathbf{X}(t) + \vec{V}(t) \tag{1}$$

With $\mathbf{X}$ being the 3D feature position, $\vec{V}$ the camera velocity, and $\lfloor \vec{\omega}(t) \rfloor$ the skew symmetric matrix of the camera angular velocities. The feature scale factor λ represents the optical flow of a feature point with $\mathbf{X} = \lambda \vec{x}$, where $\vec{x}$ is the unit length feature direction vector. Together with its derivative, we get the continuous epipolar constraint [17]:

$$\dot{\vec{x}}^T \lfloor \vec{v}(t) \rfloor \vec{x} + \vec{x}^T \lfloor \vec{\omega}(t) \rfloor \lfloor \vec{v}(t) \rfloor \vec{x} = 0 \tag{2}$$

Similar to the feature scale factor λ we apply a velocity scale factor η with $\vec{V} = \eta \vec{v}$ in the above equation. Unrotating the optical flow with the angular velocities from the IMU eliminates the second term in (2) and reduces the problem to

$$(\lfloor \dot{\vec{x}}(t) \rfloor \vec{x})^T \vec{v} \quad = \quad 0 \tag{3}$$

This equation can be solved for $\vec{v}$ using N features and SVD. Note, that the complexity of the SVD is only of dimension three (for the 3D velovity vector $\vec{v}$). Since any scaled version of the vector $\vec{v}$ solves (3), we can chose it to be of unit length without loss of generality.

As suggested in [17] and from (1) with $\vec{\omega} = 0$ using the IMU any (un-rotated) feature i used in (3) needs to fulfill its motion equation

$$\dot{\lambda}_i(t)\vec{x}_i(t) + \lambda_i(t)\dot{\vec{x}}_i(t) = \eta \vec{v}(t) \tag{4}$$

Whereas the scale factor λ_i is different for every feature (reflecting the 3D structure of the terrain), the velocity scale factor η is the same for all features.

When stacking all λ_i, $\dot{\lambda}_i$, and η into the vector

$$\vec{\lambda} = [\lambda_1, \ldots, \lambda_n, \dot{\lambda}_1, \ldots, \dot{\lambda}_n, \eta] \tag{5}$$

(4) can be rephrased as the SVD problem

$$M(\vec{x}, \dot{\vec{x}}, \vec{v})\vec{\lambda} = 0 \tag{6}$$

The solution $\vec{\lambda}$ unifies all scale factors in a consistent way up to one common scale factor Λ. This is essentially a continuous triangulation of feature positions since we can reconstruct the 3D scene up to common scale Λ with

$$\mathbf{X}_{\mathbf{i}_{\text{metric}}} \Lambda = \lambda_i \vec{x}_i \tag{7}$$

We note at this point, that the scene can be of any 3D structure and that the algorithm is not bound to planar terrain.

Similar to (3) the solution of the SVD in (6) can be arbitrarily scaled. We first introduce the notion of the *terrain-plane* and the computation of its parameters. Then we propose a specific normalization of the above velocity scale factor such that we are able to define analytically the dynamics of the common scale Λ later in the EKF framework. Such an analytic description improves the scale state estimate in the EKF prediction step.

B. Terrain-plane Parameter-Computation

The described scene reconstruction with $\mathbf{X_i} = \lambda_i \vec{x}_i$ yields a 3D point cloud representing the 3D structure of the scene in the camera frame.

We define the term *terrain-plane* as the plane fitted to this point cloud by the regression

$$[\vec{n}_{tp}^T,\ d_{tp}][\lambda_i \vec{x}_i^{\,T},\ 1]^T = 0 \tag{8}$$

with the normal vector $\vec{n}_{tp}$ and the distance to the origin (i.e. camera center) d_{tp}.

Furthermore, we denote a terrain as *locally reasonably flat* if the regression model of the terrain-plane is locally constant in a fixed world frame (Fig. 1). Single outlier objects (e.g. a single tree) can easily be accounted for using robust outlier rejection methods during the plane fitting process. Features on such objects, however, are still used for the above velocity computation. Note, that the terrain is assumed to be *locally flat* with respect to the vehicle dynamics. For agile MAVs, this assumption holds for a large variety of outdoor terrains.

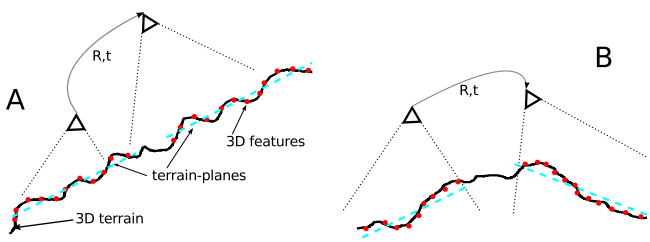


Fig. 1. A shows a locally reasonable flat terrain. That is, when the camera (black triangle) moves and observes a different part of the terrain (black line), the plane parameters in the world frame based on the regression of the triangulated features (red dots) still remain constant. The terrain-plane is depicted as dashed cyan lines. In B, the plane parameters change when the camera moves, which violates the assumption of a constant terrain-plane in the world frame.

Fig. 2 visualizes the different values computed using only the continuous epipolar constraint and optical flow of two consecutive images.

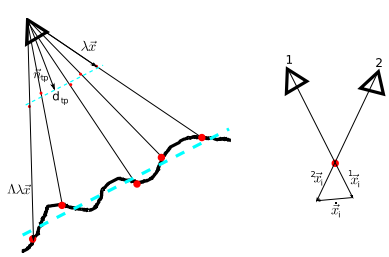


Fig. 2. We summarize that the features on the terrain can be reconstructed in 3D using their specific scale factor and their direction vector $\lambda\vec{x}$ up to a metric scale factor Λ. A planar regression of these features yield the terrain plane which is characterized with its normal vector $\vec{n}_{tp}$ and its (scaled) distance to the camera d_{tp}. On the right, the computation of the optical flow $\dot{\vec{x}}$ using two consecutive camera frames is illustrated.

C. Scene Depth Normalization

From (6) we know that the body-velocity and terrain plane distance computation are only correct up to a common scale factor Λ. For the later estimation of Λ in an EKF framework with added inertial cues, it is highly desirable to have Λ constant or to know its continuous motion model. In [16] we assumed a hovering MAV with a camera always pointing strictly down and normalized the solution vector $\vec{\lambda}$ of (6) with respect to all feature scale factors λ_i. In practice, the MAV rolls and pitches with respect to the terrain-plane and never is exactly hovering.

Using optical flow cues for computing the terrain-plane parameters, has the advantage of providing the distance d_{tp} between the camera and the terrain-plane. which can be used as a normalization factor to keep the common scale factor Λ constant even if the camera attitude changes with respect to the terrain-plane. With this normalization, we can calculate the exact motion model for the common scale Λ in the EKF propagation step.

The normalization

$$\vec{\lambda}_n \quad = \quad \frac{\vec{\lambda}}{d_{tp}} \tag{9}$$

renders the change of the common scale Λ inverse proportional to the change of the distance between the camera and the terrain-plane. As we will see in the next section, this distance can be computed by only using state variables of the EKF state vector. This makes it possible to have an exact propagation model of Λ in the EKF propagation phase.

III. Extended Kalman Filter

In the previous section, we showed how to compute the following visual cues which have an arbitrary but common scale factor Λ to metric values:

- 3D camera velocity vector (same as in [16])
- terrain plane parameters: normal vector and distance of the camera to the plane (novel contribution)

We used a novel normalization approach such that we will be able to express the dynamics of Λ in the propagation step of an EKF framework using the state vector and inertial system inputs. The computed visual cues are then used as measurements in the EKF update step. We focus on our contributions of the terrain plane representation and the propagation of Λ and summarize the remaining EKF parts as standard approaches.

We assume that the IMU inputs have the following model with bias b and zero mean white Gaussian noise n. We denote the accelerometer model with subscript a and the gyroscope model with subscript ω:

$$\omega = \omega_m - b_\omega - n_\omega \ , \qquad a = a_m - b_a - n_a \tag{10}$$

$$\dot{b_\omega} = n_{b_\omega} \ , \qquad \dot{b_a} = n_{b_a} \tag{11}$$

The state vector

$$\chi = \{p_w^i \ v_w^i \ q_w^i \ b_\omega \ b_a \ \Lambda \ p_i^c \ q_i^c \ \alpha\} \tag{12}$$

contains the IMU-centered MAV position p_w^i, velocity v_w^i and attitude q_w^i with respect to the world frame. It also contains the IMU biases on gyroscopes b_ω and accelerometers b_a, the common visual scale factor Λ and the 6D transformation between the IMU and the camera

in translation p_i^c and rotation q_i^c. Thus, we provide a self-calibrating and so-called power-on-and-go system.

A plane – the terrain plane in this case – is represented using three parameters: two for the unit normal vector (elevation α and azimuth β) and one for the distance of the plane to the origin. Intuition and a thorough non-linear observability analysis using differential geometry as proposed in [18] reveal that the azimuth β and the distance of the plane to the origin are unobservable states. Intuitively, only having a visual body velocity and the plane parameters in the floating camera frame as measurements does not anchor the system to a fix world frame in position and yaw (gravity included in the IMU readings anchor the system in global roll and pitch). Hence, the distance of the terrain plane to the world origin and the system's position p_w^i are only *jointly* observable. So are the systems yaw in q_w^i and the terrain plane's azimuth. If two states are jointly observable it is sufficient to provide a measurement for one of them to render the other observable. In our case, we assume that we operate in *locally reasonably flat terrain* which keeps the terrain plane locally constant in the world frame. Hence, without loss of generality, we can anchor the world frame in the terrain plane such that its distance to the origin and azimuth vanishes. This directly renders the system's global yaw and the system's position-dimension perpendicular to the terrain plane observable. Fig. 3 depicts the setup and frame alignment.

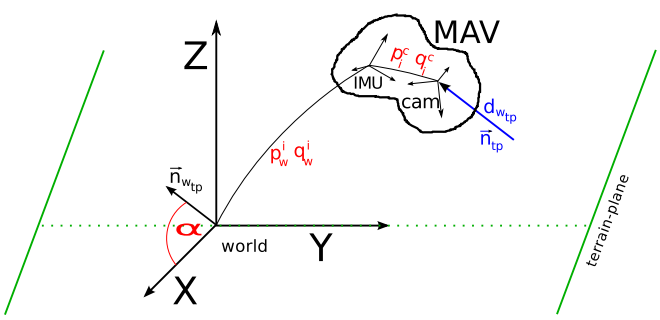


Fig. 3. Frame setup and state definition for the EKF framework. Without loss of generality, we can lock the gravity aligned world y-axis along the terrain plane. The terrain plane normal vector can then be described as $n_{w_{tp}} = [\cos(\alpha)\ 0\ \sin(\alpha)]^T$ in the world frame. The red values are states estimated in the EKF framework, whereas the blue values are the scaled visual measurements normalized with our proposed approach aid of the terrain-plane.

As a result, the MAV can not only keep its metric distance to the terrain but also keep its full attitude aligned in roll and pitch with gravity and in yaw with respect to the terrain plane. This allows advanced terrain following missions in large environments. We note at this point that we still use optical flow measurements (i.e. two consecutive images) and the corresponding IMU readings only. We do not use any temporal history of features or states. The MAV, of course, will drift in the position-dimensions parallel to the terrain plane.

A. System Dynamics

The following differential equations govern the state χ:

$$\dot{p}_w^i = v_w^i \tag{13}$$

$$\dot{v}_w^i = C_{(q_w^i)}^T(a_m - b_a - n_a) - g \tag{14}$$

$$\dot{q}_w^i = \frac{1}{2}\Omega(\omega_m - b_\omega - n_\omega)q_w^i \tag{15}$$

$$\dot{b}_\omega = n_{b_\omega},\quad \dot{b}_a = n_{b_a},\quad \dot{p}_i^c = 0,\quad \dot{p}_i^c = 0,\quad \dot{\alpha} = 0, \tag{16}$$

where $C_{(q)}$ is the rotational matrix corresponding to the quaternion q, g is the gravity vector in the world frame, and $\Omega(\omega)$ is the quaternion multiplication matrix of ω. The transformation between camera and IMU (p_w^i, q_w^i) and the terrain-plane-vector elevation α are assumed not to change. We design the filter in its error state for minimal representation and better handling of the quaternions [19].

Rather than detailing the standard procedure for the error representation, we focus on the dynamics of the common scale factor Λ. We could assume a random walk of Λ since it changes on every optical flow reading. However, due to our specific normalization in the previous section, we can express the dynamics of Λ analytically, yielding improved results compared to the random walk assumption.

We recall that the normalization factor used in the previous section was the distance d_{tp} from the camera to the terrain plane (9). We can express the metric representation of d_{tp} using the states of the system by first expressing the camera position in the world frame and then projecting this position along the plane normal vector $n_{w_{tp}}$ in the world frame in order to get the metric distance $d_{w_{tp}}$

$$d_{w_{tp}} = (p_w^i + C_{(q_w^i)}^T p_i^c)^T n_{w_{tp}} \tag{17}$$

The camera position in the world frame is the IMU position in the world frame p_w^i plus the IMU-attitude (q_w^i) dependent translation between camera and IMU p_i^c expressed in the IMU frame. Since $n_{w_{tp}}$ is a unit vector and the world frame is anchored in the terrain plane up to an elevation angle and arbitrary azimuth, we can write $n_{w_{tp}} = [\cos(\alpha)\ 0\ \sin(\alpha)]^T$.

d_{tp} being the normalization vector, the visual distance measurement between camera and terrain plane will always be $dn_{tp} = \frac{d_{tp}}{d_{tp}} = 1$. Furthermore, we defined in (7) the common scale as $p_{\text{vision}} = \Lambda p_{\text{metric}}$ Thus, the further away the camera moves from the plane the smaller becomes the common scale factor Λ. More precisely, if the MAV doubles its distance, Λ will be cut in half. In other words, the change in percent of the common scale is inversely proportional to the change in percent of the distance of the camera to the terrain plane.

The change of the metric distance $d_{w_{tp}}$ is

$$\begin{aligned} \dot{d}_{w_{tp}} &= (\dot{p}_w^i)^T n_{w_{tp}} + (p_w^i)^T \dot{n}(\alpha)_{w_{tp}}\dot{\alpha} \\ &\quad + (\dot{p}_i^c)^T C_{(q_w^i)} n_{w_{tp}} + (p_i^c)^T \dot{C}_{(q_w^i)} n_{w_{tp}} \\ &\quad + (p_i^c)^T C_{(q_w^i)} \dot{n}_{w_{tp}}\dot{\alpha} \\ &= (v_w^i)^T n_{w_{tp}} + (p_i^c)^T \dot{C}_{(q_w^i)} n_{w_{tp}} \end{aligned} \tag{18}$$

The change in percent is then

$$d^p_{w_{tp}} = \frac{\dot{d}_{w_{tp}}}{d_{w_{tp}}} \tag{19}$$

Inverting this change leads to the change of the common scale factor per time-step

$$\Lambda_{t+dt} = \frac{\Lambda_t}{1 + \int_t^{t+dt} d^p_{w_{tp}}} \tag{20}$$

Fig. 4 shows the correlation between the scale change in percent and the inverted camera-to-terrain-plane distance change in percent.

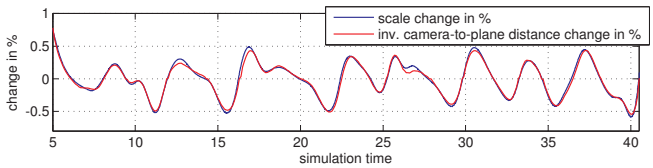


Fig. 4. Correlation between the scale change in percent and the inverted camera-to-terrain-plane distance change in percent. Thanks to our novel approach in computing the terrain plane parameters aid of optical flow, we can normalize the optical flow based body velocity readings with respect to the plane distance. This allows to have a visual scale change inverse proportional to the plane distance - and in turn this distance can be represented by the systems state allowing a precise scale prediction step. The graph is computed based on noisy simulation data.

We compare the ability to correctly propagate the visual scale factor versus the assumption of the scale factor being a random walk as done in [16]. Naturally, the scale-change depends on the camera position with respect to the scene and never reflects an arbitrary random walk. Fig. 5(a) shows the EKF-estimated versus true scale factor Λ by applying a random walk as dynamic model of the scale. We note that for this result, we hand-tuned the noise parameter of the random walk to obtain the best result possible. In a different run this noise parameter would change due to the false assumption of the scale being a random walk. Fig. 5(b) shows the same simulation run but using our dynamic model for the scale propagation in the EKF. Even though the performance only marginally improves (below 1%) we note that in this case, there are no hand-tunable parameters for the scale propagation and the model is valid for any different motion. It is interesting that in both cases we have constantly an under-estimation of about 5% with respect to the true scale. This is subject to further investigation.

B. EKF Measurements

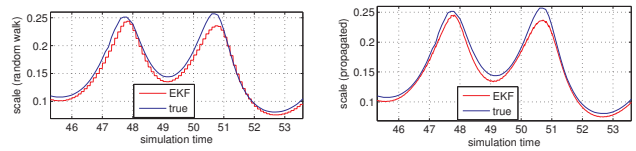
The EKF system has three different measurements: the scaled camera body velocity $z_v = \eta\vec{v}$, the terrain-plane normal vector in the camera frame $z_n = \vec{n}_{tp}$ and the distance from the camera center to the terrain plane $z_d = d_{n_{tp}} = 1$.

The measurement equations are:

$$\hat{z}_v = (C_{(q^c_i)} C_{(q^i_w)} v^i_w + C_{(q^c_i)}(\lfloor\omega\rfloor p^c_i))\Lambda + n_{z_v} \tag{21}$$

$$\hat{z}_n = C_{(q^c_i)} C_{(q^i_w)} n_{w_{tp}}(\alpha) + n_{z_n} \tag{22}$$

$$\hat{z}_d = (p^i_w + C^T_{(q^i_w)} * p^c_i)^T n_{w_{tp}}(\alpha)\Lambda + n_{z_d} \tag{23}$$



(a) (b)

Fig. 5. a) Performance of the EKF estimate for the visual scale factor Λ when assuming a random walk as a motion model and b) when using our proposed motion model. The noise of the random walk is hand-tuned to obtain the best possible result. The value of this parameter is not valid anymore for a different simulation run since the motion will be different. Conversely, our proposed model is valid for all motions. Note the step-wise behavior in a) versus the smooth, more accurate propagation in b).

where $C_{(q^i_w)}$ and $C_{(q^c_i)}$ is the attitude of the IMU and the rotation between the IMU and camera respectively. All these measurements are results of the optical measurements based on solutions of SVD problems. We can use the approach of the mean squared error (MSE) in statistics to determine the accuracy of the SVD solution. Thus, the covariance of the noise vector $\vec{n}_z = [n_{z_v}\ n_{z_n}\ n_{z_d}]$ can be computed along with the optical measurements in Section II.

The above measurement equations can be linearized with respect to the state vector χ and stacked to the measurement vector $z = [z_v\ z_n\ z_d]^T$ which yields $\hat{z} = H\chi$. Then, the standard EKF procedure can be applied

1) compute the residual $r = z - \hat{z}$
2) compute the innovation $S = HPH^T + R$
3) compute the Kalman gain $K = PH^T S^{-1}$
4) compute the correction $\hat{\chi} = Kr$
5) update the covariance matrix
$P_{k+1|k+1} = (\mathbf{I_d} - KH)P_{k+1|k}(\mathbf{I_d} - KH)^T + KRK^T$.

With this EKF framework we achieved the following:

- fully continuous self-calibrating platform including IMU biases, visual scale and inter-sensor transformation between IMU and camera
- terrain plane aligned observable global yaw and metric distance to this plane effectively eliminating drift in the position-dimension vertical to the plane
- precise visual scale propagation such that the system can track even fast scale-changing motions of the MAV

While the first point was already fulfilled in related work, the second and third points only became possible with our contribution of computing the terrain plane parameters using the optical flow cues. First, locking the world frame onto this terrain plane rendered global yaw and metric distance with respect to the terrain observable. Second, the scaled visual distance measurement to the plane allowed to normalize the visual readings such that the change of the common scale factor Λ can be recovered using system states.

IV. Performance Evaluation

This section discusses the performance of the proposed system in a real-world scenario. We implemented our approach on-board an AscTec Pelican quadrotor that is equipped with an Intel Atom 1.6GHz processing board and a global shutter WVGA camera (Fig. 6) using ROS for inter

process communication. Even though our approach would run at 50Hz on this platform, for the following experiments, we ran the camera at 30Hz and the IMU at 1kHz to provide a safety margin. The excitation of the system has an RMS value of 0.5m/s^2 in acceleration and 0.25rad/s in angular velocities.

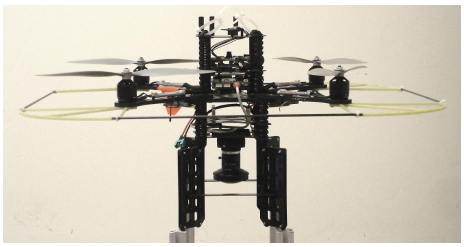


Fig. 6. AscTec pelican equipped with a 1.6GHz Intel Atom single core processor-board and a WVGA global shutter camera.

While the evaluation of the self-calibration and the metric velocity estimation was provided in [16], we focus on the novel contribution of estimating the terrain parameters and absolute yaw. Fig. 7 shows schematically the setup. For our experiments we used textured styrofoam plates mounted rigidly at hand measured angles of 40 and 60 degrees respectively. The MAV moves at a distance of 1m to 1.5m of these planes.

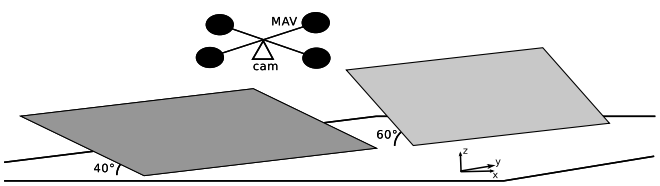


Fig. 7. Test setup to evaluate our approach: We use two differently inclined terrain planes (textured styrofoam plates) at hand measured 40 and 60 degrees respectively. We moved the MAV across the two planes to show the different behaviors for different plane inclinations.

We first show the capability of the real system to maintain global yaw. The normal vector of an inclined terrain-plane computed in the camera frame (Section II) contains a component perpendicular to the gravity vector, which is in fact the information that renders global yaw observable in our proposed system. In Fig. 8 we show the evolution of the estimate in global yaw during 80 seconds where the MAV's camera is observing a terrain with an inclination of 60 degrees. Fig. 9 shows the estimated elevation angle of the plane normal vector. The ground truth for yaw was obtained from the MAV magnetometer with initial alignment.

In Fig. 10 we compare the quality of this estimate for two different setups, one in which the terrain has an inclination of about 60 degrees and one with an inclination of about 40 degrees. In both graphs we see that the system converges to the correct yaw, however, the terrain which is inclined by 40 degrees provides much less directional information perpendicular to gravity. Hence, global yaw is less constrained. A front looking camera looking at vertical

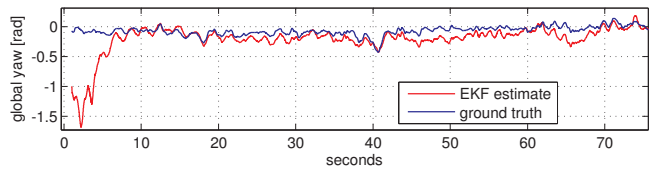


Fig. 8. Global yaw estimated by our proposed approach. The observed terrain has an inclination of 60 degrees. After a wrong initialization, the filter quickly converges to the correct yaw and remains there without drift.

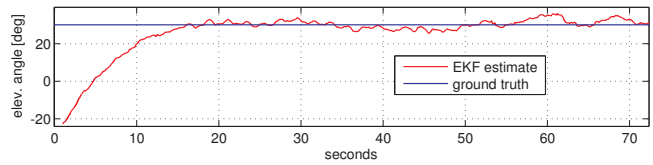


Fig. 9. After a wrong initialization, our proposed approach converges to the true elevation angle of the normal vector of the observed terrain. Note that the framework estimates the elevation angle of the plane normal vector, not the plane inclination. Hence the correct 30deg for a 60deg inclined plane.

walls would yield best results. In addition to the yaw, roll and pitch are also observable (c.f. our previous work in [16]). Hence our approach is able to estimate the full 3DoF attitude of the MAV without drift.

Next, we show that we can maintain the distance to the terrain. By rotating the estimated position from our proposed EKF framework to the terrain frame using the estimated inclination angle α, we expect drift in both directions x and y, whereas z remains constant. In the terrain frame, the z axis represents the direction of distance vector between between the MAV and the terrain. The estimated z-position can directly be used for terrain-following controllers.

Fig. 11 shows the expected results over a terrain with an inclination of 60 degrees. Note, that thanks to the visual scale estimation in the EKF framework, the vehicle keeps a constant *metric* distance to the terrain.

In Section II we stated that our approach for estimating the terrain parameters requires *locally reasonably flat* terrain. Fig. 12 shows that our approach can quickly adapt to

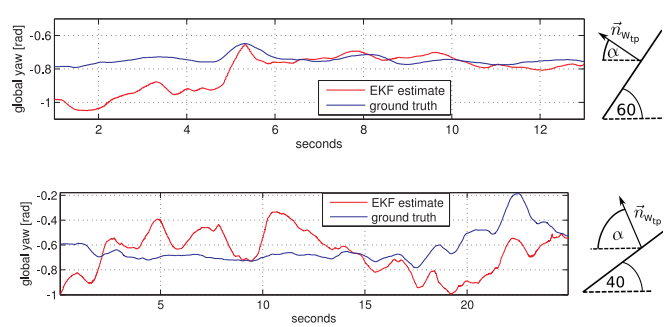


Fig. 10. The quality of the yaw estimation depends on the inclination of the terrain. The steeper the terrain is, the more is its normal vector perpendicular to gravity and the better can global yaw be estimated. The graphs show the yaw estimation over terrains with an inclination of 60 and 40 degrees respectively. The intial large error in the top graph is due to a an offset to the true value in the initialization. The yaw remains unobservable on flat terrain and is best observable with a front looking camera looking at walls. This is a well desired aspect for example in inspection tasks.

changing scenes, even when this requirement is not met. Quadrotors usually have sufficient motion to make the terrain parameters quickly converge to the new situation. In our experiment, the vehicle first observes a terrain with an inclination of 60 degrees (this corresponds to an elevation angle of 30 degrees), then the terrain changes to an inclination of 40 degrees (elevation angle is 50 degrees).

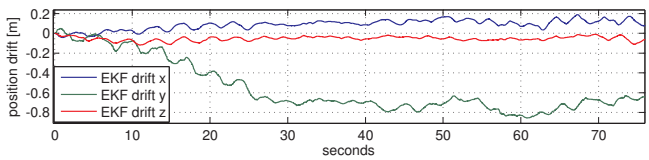


Fig. 11. We rotate the estimated global position into the terrain frame with z being in the direction of the terrain normal vector. Since we compute this scaled value and estimate the scale in the EKF framework the MAV is able to maintain this *metric* distance constant. This is crucial for terrain following tasks. x and y direction drift, as expected.

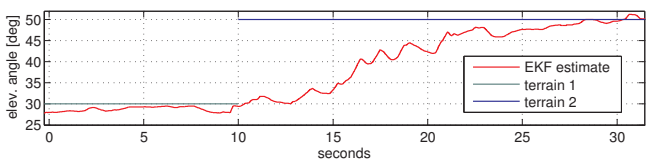


Fig. 12. Even though our approach requires *locally reasonably flat* terrain, this plot shows that our approach can quickly adapt to terrains with different inclinations. In this case, the terrain switched from a 60 degree to a 40 degree inclination after 10 seconds of flight. The plane normal vectors have an elevation of 30 and 50 degrees respectively.

V. Conclusion

In this work, we present an inertial-optical flow approach which significantly advances the state-of-the-art with respect to drift free MAV navigation and terrain following.

Based on the ability to estimate metric body velocities with optical flow and inertial readings, our approach uses the definition of the continuous epipolar constraint to compute the terrain parameters and the metric distance between the vehicle and the terrain. We motivate to use the latter as visual normalization factor in order to express the dynamics of the visual scale analytically during the EKF prediction step.

Estimating the metric distance to the terrain plane also provides the key-information for a controller performing terrain following. Even though we make the assumption of navigating in locally reasonable flat areas, we show in real experiments that the system can quickly adapt to new terrain parameters and is thus capable of robustly follow different terrain.

Locking the world frame to the terrain and estimating the terrain inclination to gravity made global yaw of the system observable. Also, the metric distance to the plane eliminates the system's position drift in one dimension. In addition, the results show that the drift in the other two position axes is minimal. We demonstrate in real experiments that the quality of the yaw is dependent on the terrain inclination. This reflects the influence of the perpendicular component to gravity on the estimation quality. Our proposed approach yields a state estimation that is drift free in 4DoF (position drifts only parallel to the terrain plane) while only using inertial readings and two consecutive images (i.e. optical flow) without any map or feature history.

References

[1] N. Michael, J. Fink, and V. Kumar, "Cooperative manipulation and transportation with aerial robots," *Autonomous Robots*, 2010.

[2] N. Michael, D. Mellinger, Q. Lindsey, and V. Kumar, "The grasp multiple micro UAV testbed," *IEEE Robotics and Automation Magazine*, May 2010.

[3] S. Lupashin, A. Schoellig, M. Sherback, and R. D'Andrea, "A simple learning strategy for high-speed quadrocopter multi-flips," in *International Conference on Robotics and Automation*, 2010.

[4] N. Abdelkrim, N. Aouf, A. Tsourdos, and B. White, "Robust nonlinear filtering for INS/GPS UAV localization," in *Proceedings of the 16th IEEE Mediterranean Conference on Control and Automation*, Ajaccio, Corsica, France, 2008, pp. 695–702.

[5] B. Yun, K. Peng, and B. Chen, "Enhancement of GPS signals for automatic control of a UAV helicopter system," in *Proceedings of the IEEE International Conference on Control and Automation*, Hong Kong, China, 2007, pp. 1185–1189.

[6] A. Bachrach, R. He, and N. Roy, "Autonomous flight in unstructured and unknown indoor environments," in *European Conference on Micro Aerial Vehicles (EMAV)*, 2009.

[7] S. Shen, N. Michael, and V. Kumar, "Autonomous multi-floor indoor navigation with a computationally constrained MAV," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.

[8] T. Cheviron, T. Hamel, R. Mahony, and G. Baldwin, "Robust nonlinear fusion of inertial and visual data for position, velocity and attitude estimation of UAV," in *International Conference on Robotics and Automation*, Apr. 2007, pp. 2010–2016.

[9] K. E. Wenzel, A. Masselli, , and A. Zell, "Automatic take off, tracking and landing of a miniature UAV on a moving carrier vehicle," in *UAV'10 3rd International Symposium on Unmanned Aerial Vehicles*, 2010.

[10] S. Ahrens, D. Levine, G. Andrews, and J. How, "Vision-based guidance and control of a hovering vehicle in unknown, GPS-denied environments," in *International Conference on Robotics and Automation*, May 2009.

[11] S. Weiss, D. Scaramuzza, and R. Siegwart, "Monocular-SLAM based navigation for autonomous micro helicopters in GPS-denied environments," *Journal of Field Robotics*, vol. 28, no. 6, pp. 854–874, 2011. [Online]. Available: http://dx.doi.org/10.1002/rob.20412

[12] B. Hérissé, T. Hamel, R. Mahony, and F.-X. Russotto, "A terrain-following control approach for a VTOL unmanned aerial vehicle using average optical flow," *Autonomous Robots*, vol. 29, pp. 381–399, 2010.

[13] J. Zufferey and D. Floreano, "Toward 30-gram autonomous indoor aircraft: Vision-based obstacle avoidance and altitude control," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2005.

[14] V. Lippiello, G. Loianno, and B. Siciliano, "Mav indoor navigation based on a closed-form solution for absolute scale velocity estimation using optical flow and inertial data," in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, dec. 2011, pp. 3566 –3571.

[15] V. Grabe, H. Bulthoff, and P. Giordano, "Robust optical-flow based self-motion estimation for a quadrotor uav," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, oct. 2012, pp. 2153 –2159.

[16] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.

[17] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An invitation to 3D vision : from images to geometric models*, Springer, Ed. Springer, 2000.

[18] A. Martinelli, "State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration," *Robotics, IEEE Transactions on*, vol. 27, no. 2, pp. 239 –255, april 2011.

[19] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *International Journal of Robotics Research (IJRR)*, vol. 30, no. 1, pp. 56–79, 2011.