Unscented Kalman Filter on Lie Groups for Visual Inertial Odometry
Martin Brossard, Silvere Bonnabel, Axel Barrau

To cite this version:
Abstract—Fusing visual information with inertial measurements for state estimation has aroused major interests in recent years. However, combining a robust estimation with computational efficiency remains challenging, specifically for low-cost aerial vehicles in which the quality of the sensors and the processor power are constrained by size, weight and cost. In this paper, we present an innovative filter for stereo visual inertial processor power are constrained by size, weight and cost. In this paper, we present an innovative filter for stereo visual inertial processor power are constrained by size, weight and cost. In this paper, we present an innovative filter for stereo visual inertial processor power are constrained by size, weight and cost. In this paper, we present an innovative filter for stereo visual inertial processor power are constrained by size, weight and cost. In this paper, we present an innovative filter for stereo visual inertial

1) We benefit from the Unscented Kalman Filter (UKF) methodology as compared to the standard Extended Kalman Filter (EKF). Since the unscented transform spares the computation of Jacobians, the algorithm is versatile and allows fast prototyping in the presence of variations in the model (e.g., the camera model).

2) We build upon the theory of the Unscented Kalman Filter on Lie Groups (UKF-LG) [3], and leverage the Lie group structure of the SLAM problem introduced in [4].

We demonstrate that the proposed stereo VIO filter is able to achieve similar or even higher accuracy than state-of-the-art solutions on two distinct datasets with high efficiency.

A. Links with Previous Literature and Contributions

Over the last decades, tremendous progresses have been achieved in visual localization frameworks, whose estimation and robustness can be improved by tightly coupling visual and inertial informations, which is the major focus of this paper. Most approaches combine data using filtering based solutions [2,5]–[10], or optimization/bundle adjustment techniques, e.g., [11]–[13]. Optimization based methods are more efficient but generally come with higher computational demands, and filtering approaches are well suited to real time applications. One drawback of conventional VIO filter-based algorithms are their inconsistency [5,14], which is resolved by using the observability constrained approach [8,10,15,16]. Among popular solutions, the MSCKF and its state-of-the-art variants [2,8,9] offer an efficient compromise between accuracy and computational complexity.

Recently, manifold and matrix Lie group representations of the state variables have drawn increasing attention for computer vision and robotics applications [11,17,18]. With the natural capability for associating uncertainty to rigid body motion [19], matrix Lie group representations have been successfully utilized to design optimization or filter-based methods. Additionally, it has been shown that, by expressing the EKF estimation error directly on the Lie group and leveraging an Invariant-EKF, consistency guarantees can be obtained without ad hoc remedies both for wheel odometry SLAM [4,20] and VIO [6] filtering algorithms. Specifically for VIO purpose, in [7], the authors devise an UKF that takes advantage of the Lie group structure of the robot’s (quadrotor) pose $SE(3)$, and uses a probability distribution directly defined on the group (the distributions in [21]) to generate the sigma points, which is akin to the general unscented
Kalman filtering on manifolds of [22], that contrasts with the generation of the sigma points directly in the Lie algebra we proposed in [3].

In this paper, we propose an UKF-based stereo VIO solution that leverages the Lie group structure of the state space $SE(3)$. Our main contributions are:

- an embedding of the state and the uncertainties into a matrix Lie group which additionally considers the unknown IMU to camera transformation;
- the derivation of a Kalman filter that combines an (Invariant-)EKF propagation for computational efficiency and an UKF-LG update, in which our choice is motivated by the UKF superiority performance compared with the EKF for many non-linear problems, and its ease of implementation for the practitioner, allowing him to readily handle additional measurements (such as GPS measurements) or variations in the output model (the camera model) since the update is derivative free;
- since the computational demands of a standard UKF update is generally greater than those of the EKF, we provide: i) a computationally efficient strategy for computing our UKF-LG update in the formalism of an (Invariant-)EKF update, inspired by [16,23]; and ii) a closed-form expression for alternatively performing the update in a full (Invariant-)EKF manner;
- the publicly available C++/ROS source code used for this paper, available at https://github.com/mbrossar/msckf_vio.git. It uses building blocks from the code of [2].

Finally, the accuracy and computational complexity of the proposed filter is validated and compared with state-of-the-art VIO solutions on two challenging real-world MAV datasets [2,24].

B. Paper’s Organization

Section II formulates the filtering problem. Section III contains mathematical preliminaries on matrix Lie groups and unscented Kalman filtering on Lie groups. Section IV describes the proposed filter for stereo VIO. Section V illustrates the performances of the proposed filter based on two publicly available datasets, and Section VI concludes the paper.

II. PROBLEM MODELING

We define in this section the kynodynamic model for flying devices equipped with an IMU where $N$ past cloned camera poses make up the state [9]. We then detail the stereo visual measurement model, and we finally pose the filtering problem we seek to address.

A. Variables of Interest and Dynamical Model

Let us consider an aerial body navigating on flat earth equipped with an IMU. The dynamics of the system read

\begin{equation}
\begin{aligned}
\dot{\mathbf{R}} &= \mathbf{R} (\omega - \mathbf{b}_\omega + \mathbf{n}_\omega) x \\
\dot{\mathbf{v}} &= \mathbf{R} (\mathbf{a} - \mathbf{b}_a + \mathbf{n}_a) + \mathbf{g} \\
\dot{\mathbf{b}}_\omega &= \mathbf{n}_b_\omega \\
\dot{\mathbf{b}}_a &= \mathbf{n}_b_a \\
\mathbf{T}^L &= 0 \\
\mathbf{T}^R_i &= 0, \quad i = 1, \ldots, N
\end{aligned}
\end{equation}

where the state we want to estimate consists of the current orientation $\mathbf{R} \in SO(3)$ of the body frame (referred to as the IMU frame), that is, the rotation matrix that maps the IMU frame to the global frame, velocity $\mathbf{v} \in \mathbb{R}^3$, position $\mathbf{x} \in \mathbb{R}^3$, IMU biases $\mathbf{b}_\omega \in \mathbb{R}^3$ and $\mathbf{b}_a \in \mathbb{R}^3$, as well as the relative transformation between the IMU frame and the left camera frame $\mathbf{T}^L \in SE(3)$ and an arbitrary number of $N$ recorded left camera poses $\mathbf{T}^C_i \in SE(3)$, $i = 1, \ldots, N$. Finally, $\omega_x$ denotes the skew symmetric matrix associated with the cross product with vector $\omega \in \mathbb{R}^3$, and the various white Gaussian noises can be stacked as

$$\mathbf{n} = [\mathbf{n}_\omega^T \mathbf{n}_a^T \mathbf{n}_b_\omega^T \mathbf{n}_b_a^T]^T \sim N(0, \mathbf{Q}).$$

These equations model the dynamics of small MAVs such as quadrotors where the IMU measurements $\omega$ and $a$ in (1) are considered as noisy and biased inputs of the system.

B. Measurement Model

In addition to the IMU measurements used as inputs for the dynamics, the vehicle observes and tracks static landmarks in the global frame from a calibrated stereo camera. A landmark $\mathbf{p}^j \in \mathbb{R}^3$ is observed through both the left and right cameras corresponding to the recorded $i$-th pose as

$$\mathbf{y}^j = h_i (\mathbf{T}^C_i, \mathbf{p}^j) + \mathbf{n}_y \in \mathbb{R}^4,$$

where the non-linear stereo measurement model $h(\cdot, \cdot)$ is given as [2]

$$h_i (\mathbf{T}, \mathbf{p}) = \begin{bmatrix} \frac{1}{sz} & 0 & x_l \\ 0 & \frac{1}{sz} & y_l \\ \frac{1}{st} & 0 & x_r \\ 0 & \frac{1}{st} & y_r \end{bmatrix},$$

Fig. 1. The coordinate systems that are used in the paper. The IMU pose $(\mathbf{R}, \mathbf{x})$ maps vectors expressed in the IMU frame to the global frame $(\mathbf{G})$. The transformations $\mathbf{T}^L$ and $\mathbf{T}^R$ pass, respectively, from the IMU frame to the left camera $(\mathbf{C}_L)$ frame, and from the left camera frame to the right camera $(\mathbf{C}_R)$ frame. The pose $\mathbf{T}^C_i$ consists of the position of $\mathbf{C}_L$ in the global frame and the rotation mapping vectors expressed in the $\mathbf{C}_L$ frame to vectors expressed in the global frame. Unknown 3D features $\mathbf{p}^j$ (expressed in the global frame) are tracked across a stereo camera system.
where $\mathbb{R}^n$ is the covariance of the small, noisy perturbation representing a large, noise-free and deterministic value, whereas $\chi$ is a white Gaussian noise. Essentially two different UKFs follow from the above uncertainty representation.

1) **Right-UKF-LG**: the state is modeled as $x_{n+1} \sim N_R(x_n, p_n)$, that is, using the representation (9) of the uncertainties. The mean state is thus encoded in $x_n$ and dispersion in $\mathcal{N}(0, Q_n)$ is a white Gaussian noise, associated with generic discrete measurements of the form

$$y_n = h(x_n, v_n),$$

where $v_n \sim \mathcal{N}(0, R_n)$ is a white Gaussian noise. Essentially two different UKFs follow from the above uncertainty representation.

2) **Left-UKF-LG**: the state is alternatively modeled as $x_{n+1} \sim N_L(x_n, p_n)$, that is, using the left-equivariant formulation (10) of the uncertainties.

**D. Unscented Based Inferred Jacobian for UKF-LG Update**

In [23], the authors interpret the conventional UKF as a linear regression Kalman filter and show how the propagation and update steps in UKF can be performed in a similar fashion as an EKF, which can save execution time [16]. The method is here straightforwardly adapted for the case of the Right-UKF-LG update. Within this interpretation, the
filter seeks to find the optimal linear approximation to the nonlinear function
\[ y = h(\exp_G(\xi) \tilde{x}) = g(\xi) \] (13)
\[ \simeq \mathbf{H}_\xi + \dot{y} \] (14)
given a weighted discrete representation (the so-called sigma-points) of the distribution \( \xi \sim \mathcal{N}(0, \mathbf{P}) \). The objective is thus to find the regression matrix \( \mathbf{H} \) and vector \( \dot{y} \) that minimize the linearization error \( e = y - (\mathbf{H}_\xi + \dot{y}) \). The optimal linear regression matrix is given as [23]
\[ \mathbf{H} = \mathbf{P}_y \xi \mathbf{P}^{-1}, \] (15)
where \( \mathbf{P}_y \xi \) is the cross-correlation between \( y \) and \( \xi \), and \( \dot{y} \) is the estimated mean of \( y \), both computed from the unscented transform of the UKF-LG. The numerically inferred Jacobian \( \mathbf{H} \) serves as a linear approximation to \( y \) and \( \xi \). The optimal linear approximation to \( y \) and \( \xi \) can then be used for the Right-EKF-LG update.

A. State and Error Embedding on Lie Groups

Based on Section III-E, we embed the state into a high dimensional Lie group, by letting \( \chi \in G \) be the matrix that represents:
- the IMU variables \( \mathbf{R}, \mathbf{v} \) and \( \mathbf{x} \) through \( \chi^T \in SE_2(3) \), a group obtained by letting \( p = 0 \) in (16);
- the IMU bias \( \mathbf{b}_R^{\mathbf{b}} = [b_{\mathbf{b}}^R]^T \in \mathbb{R}^6 \);
- the IMU to left camera transformation \( \mathbf{T}_{IC} \in SE(3) \);
- the \( N \) left camera poses \( \mathbf{T}_i^C \in SE(3), i = 1, \ldots, N \).

The dispersion on the state
\[ \chi^T, \chi^b, \mathbf{T}_{IC}, \mathbf{T}_1^C, \ldots, \mathbf{T}_N^C \equiv \chi \sim \mathcal{N}_R(\tilde{\chi}, \xi), \] (19)
where \( \equiv \) stands for “identifiable to” and \( \sim \) for estimated mean value, is partitioned into
\[ \xi = [\xi_R^T \, \xi_v^T \, \xi_x^T \, \xi_{\mathbf{a}}^T \, \xi_{\mathbf{a}}^{T,IC} \, \xi_{\mathbf{a}}^{IC} \, \cdots \, \xi_{\mathbf{a}}^{IC,N}]^T, \] (20)
and encoded using the right uncertainty (9), i.e., the uncertainty representation is defined as
\[ \exp(\xi) \tilde{\chi} = \begin{pmatrix} \exp_{SE_2(3)}(\xi_R^T) & \xi_v^T & \xi_x^T & \xi_{\mathbf{a}}^T & \xi_{\mathbf{a}}^{T,IC} & \xi_{\mathbf{a}}^{IC} & \cdots & \xi_{\mathbf{a}}^{IC,N} \end{pmatrix}^T. \] (21)
and we define for convenience the IMU error as \( \xi_i = [\xi_R^i, \xi_v^i, \xi_x^i, \xi_{\mathbf{a}}^i, \xi_{\mathbf{a}}^{T,IC}, \xi_{\mathbf{a}}^{IC} ]^T \). Any unknown feature \( p^j \), albeit not explicitly considered in the state, appear in the measurement (4) and consequently we have to define an error on this feature. In the following and inspired from Section III-E, we propose to identify each \( (\mathbf{T}_i^C, \mathbf{p}^j) \) as a element of the Lie group \( SE_2(3) \) [4]. Note that, error \( \xi_{\mathbf{p}} \) on landmark \( j \) then differs from the standard Euclidean error.

Remark 3: using another camera pose than \( \mathbf{T}_i^C \) does not influence the performances of the filter in our experiments.

B. Propagation Step

Let us now present the proposed filter’s mechanics. To deal with discrete time measurement from the IMU, we essentially proceed along the lines of [9]. We apply a 4-th order Runge-Kutta numerical integration of the model dynamic (1) to propagate the estimated state \( \tilde{\chi} \). To propagate the uncertainty of the state, let us consider the dynamic of the IMU linearized error as
\[ \dot{\xi}_i = \mathbf{F}_i \xi_i + \mathbf{G}_i \mathbf{n}, \] (22)
where
\[ \mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 0 & -\mathbf{R} \\ \mathbf{g}^T & 0 & 0 & -\mathbf{R} & -\mathbf{v}^T \mathbf{R} \\ 0 & 1 & 0 & 0 & -\mathbf{x}^T \mathbf{R} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \] (23)
\[ G = \begin{bmatrix} 0 & R & 0 & 0 \\ R & (v) & R & 0 \\ 0 & (x) & R & 0 \\ 0 & 0 & I & 0 \end{bmatrix} , \]  
\[ (24) \]

and first compute the discrete time state transition matrix
\[ \Phi_n = \Phi(t_{n+1}, t_n) = \exp_m \left( \int_{t_n}^{t_{n+1}} F(\tau) d\tau \right) \]  
\[ (25) \]

and discrete time noise covariance matrix
\[ Q_n = \int_{t_n}^{t_{n+1}} \Phi(t_{n+1}, \tau) GQG^T \Phi(t_{n+1}, \tau)^T d\tau. \]  
\[ (26) \]

The covariance matrix from \( t_n \) to \( t_{n+1} \) is propagated as the combination of partitioned covariance matrix as follows. The propagated covariance of the IMU state becomes
\[ P_{n+1}^{II} = \Phi_n P_n^{II} \Phi_n + Q_n, \]  
\[ (27) \]

and the full uncertainty propagation can be computed as
\[ P_{n+1} = \begin{bmatrix} P_{n+1}^{II} & \Phi_n P_n^{IC} \\ P_n^{CI} \Phi_n & P_n^{CC} \end{bmatrix}. \]  
\[ (28) \]

When new images are received, the state should be augmented with the new camera state. The new augmented covariance matrix becomes
\[ P_{n+1} = \begin{bmatrix} I & J \\ J & P_{n+1} \end{bmatrix} I^T. \]  
\[ (29) \]

To obtain the expression of \( J \) in (29), let us denote \( T \in SE(3) \) as the sub-matrix of \( \chi^i \) that contain only the IMU pose \((R, x)\) with \( \xi = [\xi_R^T \xi_x^T]^T \) its corresponding uncertainties and thus write the current camera pose as
\[ T^C = TT^{IC} = \exp_{SE(3)}(\xi_T) T \exp_{SE(3)}(\xi_{IC}) T^{IC} = \exp_{SE(3)}(\xi_T) \exp_{SE(3)}(\Ad_T \xi_{IC}) T^{IC} \simeq \exp_{SE(3)}(\xi_T) T^{IC}, \]  
\[ (30) \]

in which \( T^C = TT^{IC} \) and \( \xi_C \simeq \xi_T + \Ad_T \xi_{IC} \) after using a first order BCH approximation, and where \( \Ad_T \) is the adjoint notation of \( SE(3) \), which finally leads to
\[ J = \begin{bmatrix} I & 0 & 0 & 0 & 0_{3x6} \\ 0 & 0 & I & 0 & 0_{3x6} \end{bmatrix} R_0 \times R_0 \times 0_{3x6}. \]  
\[ (31) \]

C. Update Step in the MSCKF Methodology

Let us first consider the observation of a single feature \( p^i \), in which the estimated unbiased feature position \( L^{\text{LS}} p^i \) is computed using least squares estimate based on the current estimated camera poses \( \bar{x} \). Linearizing the measurement model at the obtained estimates \( \bar{x} \), \( L^{\text{LS}} p^i \), the residual of the measurement is approximated as
\[ r^i_j = y^i_j - L^{\text{UKF}} y^i_j = L^{\text{UKF}} H^i_j \xi + H^i_{p^i} \xi_{p^i} + n^i_j, \]  
\[ (32) \]

where \( L^{\text{UKF}} H^i_j \), \( H^i_{p^i} \), and \( L^{\text{UKF}} y^i_j \) are computed in Section IV-D and are not the usual Jacobians appearing in [9] (beyond the fact they are computed using the unscented transform) since we use here alternative state errors \( \xi \), \( \xi_{p^i} \) related to the Lie group structure we have endowed the state space with. By stacking multiple observations of the same feature \( p^i \), we then dispose of
\[ r^i_j = y^i_j - L^{\text{UKF}} y^i_j = L^{\text{UKF}} H^i_j \xi + H^i_{p^i} \xi_{p^i} + n^i_j. \]  
\[ (33) \]

Then, along the lines of [9], to eliminate the landmarks from the residual, measurements are projected onto the null space \( V \) of \( H^i_{p^i} \), i.e.,
\[ r_{o}^i_j = V^T r^i_j = V^T L^{\text{UKF}} H^i_j \xi + V^T n_j = H^i_{o} \xi + n_{o}^i. \]  
\[ (34) \]

Based on (34) and after stacking residual and Jacobians for multiple landmarks in, respectively, \( r_o \) and \( H_o \), the update step is carried out by first computing \( S = R + H_o P_o H_o^T \) and the gain matrix \( K = P_o H_o^T / S \). We then compute the innovation \( \xi \) to update the mean state as [3,17]
\[ \xi = K r_o, \]  
\[ (35) \]
\[ \bar{x}^+ = \exp(\xi) \bar{x}, \]  
\[ (36) \]

which is an update that is consistent with our uncertainties (21) defined using the right multiplication based representation (9). The associated covariance matrix writes
\[ P^+_n = P_n (I - KH_o). \]  
\[ (37) \]

The filter concludes the update step with a possible marginalization of camera poses.

D. Proposed Lie Group Based Update Using the Unscented Transform of UKF-LG

In this section, we describe the computation of \( L^{\text{UKF}} H^i_j \), \( H^i_{p^i} \), and \( L^{\text{UKF}} y^i_j \) in (32) following Section III-D.

1) Computation of \( H^i_{p^i} \): since the covariance of \( \xi^i \) is unknown, we can not apply UKF-LG update and consequently we compute \( H^i_{p^i} \) in closed-form, as \( H^i_{p^i} = 0 \), and for \( i > 1 \), as
\[ H^i_{p^i} = J^i_j \begin{bmatrix} \bar{R}_i^T \\ \bar{R}_i^R \end{bmatrix} T^C, \]  
\[ (38) \]

where \( \bar{R}_i^L \) and \( \bar{R}_i^R \) are the orientations of the \( i \)-th left and right cameras, and where
\[ J^i_j = \begin{bmatrix} 1/\bar{z}_i & 0 & -\bar{x}_i/\bar{z}_i^2 \\ 0 & 1/\bar{z}_i & -\bar{y}_i/\bar{z}_i^2 \\ 1/\bar{z}_r & 0 & -\bar{x}_r/\bar{z}_r^2 \\ 0 & 1/\bar{z}_r & -\bar{y}_r/\bar{z}_r^2 \end{bmatrix} \]  
\[ (39) \]

is the intermediate Jacobian after applying the chain rule in (4). We stress that even if (38) is the same expression as the Jacobians appearing in [9], they are associated to the alternative state errors \( \xi^p \).

2) Computation of \( L^{\text{UKF}} H^i_j \) and \( L^{\text{UKF}} y^i_j \): we apply the UKF-LG [3] to numerically infer the Jacobian \( L^{\text{UKF}} H^i_j \) and estimated measurement \( L^{\text{UKF}} y^i_j \) (see Section III-D). This computation allows us to then project the residual in (34) and can be computed efficiently as follow. First, the filter stacks multiple observations w.r.t. the same camera pose in a vector \( y_i \) that can be written in the form \( y_i = g_i(\xi_{C_i}, \xi_{C_i}) \), when noise and landmark errors are marginalized, as follow. Since
where \( y_i \) is the concatenation of (4) for multiple landmarks, it is sufficient to write \( y_i^j \) as a function of \( \xi_r, \xi_c \), and \( \xi_1 \), only. Let us first write our alternative state error as

\[
\begin{aligned}
\exp_{SO(3)}(\xi_r) &= R_{C_i} R_{C_i}^T, \\
\xi_r &= R_{C_i} R_{C_i}^T x_{C_i} - x_{C_i}, \\
\xi_p &= R_{C_i} R_{C_i}^T p - L^S p_j,
\end{aligned}
\]

(40)

(41)

(42)

and consider the measurement function of the stereo camera such that

\[
\begin{aligned}
y_j^l &= l(R_{C_i}^T (p_j^i - x_{C_i})),
\end{aligned}
\]

(43)

where \( l(\cdot) \) is the projection function. After inserting the uncertainties (40)-(42) in (43), we obtain

\[
\begin{aligned}
y_j^l &= l(R_{C_i}^T (\exp_{SO(3)}(\xi_r)_{C_i}) \exp_{SO(3)}(-\xi_r)_{C_i}) L^S p_j \\
&- R_{C_i}^T (x_{C_i} + \xi_x_{C_i})),
\end{aligned}
\]

(44)

which depends on \( \xi_r, \xi_c \), and \( \xi_1 \), only, such that we can write \( y_j^l \) and by extension \( y_j \) as function of \( \xi_r, \xi_c \), and \( \xi_1 \). Following then [16], we sample only sigma-points from variables the observations depend on, i.e., from the rotational part of \( \xi_r, \xi_c, \xi_1 \), such that the complexity remains dominated by (37) and comparable to the S-MSCKF, which is illustrated in Section V.

E. Alternative Invariant-EKF Update

Albeit the UKF-LG update is computed efficiently, an Invariant-EKF update remains slightly more computationally efficient and can be adopted for computationally restricted platforms. Thus, as an alternative, we provide the closed-form expression for an Invariant-EKF update, yielding a stereo extension of [6], and where moreover the IMU to camera transformation is also estimated. The mean \( y_j = (T^C_{i}, p^i) \) is alternatively computed with estimated state, \( H_{p}^j \) in (38) and \( H_{i}^j \) as

\[
H_{p}^j = J_j^i \begin{bmatrix} 0 & H_{r}^j & 0 & H_{c}^j & 0 \end{bmatrix},
\]

(45)

where \( J_j^i \) is computed in (39), for \( i > 1 \),

\[
H_{r}^j = \begin{bmatrix} - (R_{i}^T (L^S p_j) \times 0) \\
- (R_{i}^T (L^S p_j) \times 0) \end{bmatrix},
\]

(46)

\[
H_{c}^j = \begin{bmatrix} (R_{i}^T (L^S p_j) \times (R_{i}^T)) \\
(R_{i}^T (L^S p_j) \times (R_{i}^T)) \end{bmatrix},
\]

(47)

and, for \( i = 1 \),

\[
H_{p}^j = H_{r}^j = \begin{bmatrix} 0 & (R_{i}^T) \\
0 & (R_{i}^T) \end{bmatrix}.
\]

(48)

The rest of the update follows Section IV-C.

F. Filter Update Mechanism and Image Processing Frontend

Our implementation builds upon [2], and preserves its original methodology both for the filter update mechanism, marginalization of camera poses, outlier removal and the image processing frontend.

V. Experimental Results

In this section, we compare the performances of the proposed S-UKF-LG and its full (Invariant-)EKF variant, that we call S-IEKF, with state-of-the-art VIO algorithms including S-MSCKF [2], OKVIS (stereo-optimization) [12] and VINS-MONO (monocular-optimization) [13], i.e., with different combinations of monocular, stereo, filter-based and optimization-based solutions. We first evaluate the algorithms on the EuRoC dataset [24], and then on a runway environment [2] with high speed flights. In both of the experiments, VINS-MONO considers only the images from the left camera and has its loop closure functionality disabled. The three filters (S-MSCKF, S-UKF-LG and S-IEKF) use the same frontend and the same parameters provided from the S-MSCKF github repository, with \( N = 20 \) camera poses. Finally, we provide to each algorithms the off-line estimating extrinsic parameters between the IMU and camera frames.

Summary of the results can be found in Figure 2 and Figure 4. They reveal good performances of our proposed S-UKF-LG, which favorably compares to its conventional Stereo-MSCKF counterpart in terms of RMSE both on position and orientation. Note that, optimization based OKVIS achieves best estimation accuracy, but at the cost of extended CPU load.

A. EuRoC Dataset

The EuRoC [24] dataset includes synchronized 20 Hz stereo images and 200 Hz IMU messages collected on a MAV. The dataset contains sequences of flights with different level of flight dynamics. Figure 2 shows the Root Mean Square Error (RMSE) and Figure 3 the average CPU load of the different algorithms. As in [2], the filter-based methods do not work properly on V2_03_difficult due to their same KLT optical flow algorithm. In terms of accuracy, the filters compete with the stereo-optimization approach OKVIS, whereas the results of VINS-MONO are affected by its monocular camera frontend. The S-UKF-LG and S-IEKF compare favorably to the recent S-MSCKF. Since filters use the same frontend, differences come from the filters’ backends. In terms of computational complexity, filter-
based solutions reclaims clearly less computational resources than optimization-based methods, in which 80% of the computation is caused by the frontend including feature detection, tracking and matching, on Precision Tower 7910 armed with CPU E5-2630 v4 2.20 Hz. The filters themselves take about 10% of one core when the camera rate is set at 20 Hz.

B. Fast Flight Dataset

To further test the accuracy and the robustness of the proposed S-UKF-LG, the algorithms are evaluated on four outdoor flight datasets with different top speeds of 5 m/s, 10 m/s, 15 m/s, and 17.5 m/s [2]. During each sequence, the quadrotor goes 300 m straight and returns to the starting point. The configuration includes two cameras running at 40 Hz and one IMU running at 200 Hz. Figure 4 compares the accuracy of the different VIO solutions on the fast flight datasets. The accuracy is evaluated by computing the RMSE of estimates w.r.t. GPS positions only in the x y directions after making corrections on both time and yaw offsets. In this experiment, OKVIS obtains the best results in terms of accuracy, while VINS-MONO generally obtains an accurate estimation with higher variance than the other methods which increase its RMSE. The S-UKF-LG appears to be slightly more robust than S-MSCKF and S-IEKF. From both experiments, it can be observed that the three filters achieve the lowest CPU usage while maintaining comparable accuracy regarding optimization solutions. However, compared to the experiments with the EuRoC dataset, the image processing frontend spends more computational effort, since the image frequency and resolution are higher, and the fast flight requires more detections of new features.

VI. Conclusion

In this paper, we introduced a novel filter-based stereo visual inertial state estimation algorithm that is a stereo multi-state constraint Kalman filter variant. It has the merit of using the UKF approach while achieving execution times that are akin to the standard EKF-based solution, namely S-MSCKF. The UKF approach is generally more robust to non-linearities, and allows fast prototyping since Jacobian explicit computation is not required (and thus readily adapts to model modifications, estimation of additional parameters, and fusion with other sensors). We exploited an efficient inference of the Jacobian that leads to similar computational complexity between our S-UKF-LG solution and the S-MSCKF. We also provided the closed-forms for the measurement Jacobian that lead to an alternative S-IEKF, that can be considered as a stereo version of [6] extending it also with an on-line estimation of the extrinsic parameters, resulting in a consistent filter with high level of accuracy and robustness. Accuracy, efficiency, and robustness of our filters are demonstrated using two challenging datasets.

REFERENCES


