Unscented Kalman Filter on Lie Groups for Visual Inertial Odometry

Martin Brossard, Silvere Bonnabel, Axel Barrau

Unscented Kalman Filter on Lie Groups for Visual Inertial Odometry

Martin BROSSARD*, Silvère BONNABEL* and Axel BARRAU†
*MINES ParisTech, PSL Research University, Centre for Robotics, 60 Boulevard Saint-Michel, 75006 Paris, France
†Safran Tech, Groupe Safran, Rue des Jeunes Bois-Châteaufort, 78772, Magny Les Hameaux Cedex, France

***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 odometry building on: i) the recently introduced stereo multi-state constraint Kalman filter; ii) the invariant filtering theory; and iii) the unscented Kalman filter (UKF) on Lie groups. Our solution combines accuracy, robustness and versatility of the UKF. We then compare our approach to state-of-art solutions in terms of accuracy, robustness and computational complexity on the EuRoC dataset and a challenging MAV outdoor dataset.**

***Index Terms*—Lie groups, unscented Kalman filter, visual inertial odometry, aerial vehicle, localization**

I. INTRODUCTION

Fusion of visual and inertial measurements, albeit a well established field of research, is receiving increasing attention owing to the development of highly maneuvering autonomous robots able to solve tasks such as mapping or search and rescue [1]. In such scenarios, Micro Aerial Vehicles (MAVs) have to navigate in cluttered and GPS-denied environments that pose challenges to Visual Inertial Odometry (VIO) algorithms both for the frontend image processor and the estimation itself, with, e.g., highly varying lighting conditions and vehicle attitude. The visual inertial system, which consists of an Inertial Measurement Unit (IMU) associated with a camera and equips most of MAVs, constitutes an attractive sensor suite both for localization and environment awareness, due to its low-cost and reduced lightweight. Indeed, the IMU measures noisy biased accelerations and rotational velocities at high sampling time ($100-200\,\mathrm{Hz}$) whereas camera provides rich information for visual tracking at lower rate ($20\,\mathrm{Hz}$).

In this paper, we tackle the problem of fusing IMU signals with stereo vision, since adopting a stereo configuration provides higher robustness compared to the popular monocular configuration. We propose a novel algorithm whose implementation mainly builds on the very recent Stereo Multi-State Constraint Kalman Filter (S-MSCKF) [2]. The difference with the latter paper is twofold.

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 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-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 accurate 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 $SE(3)_{2+p}$. 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.

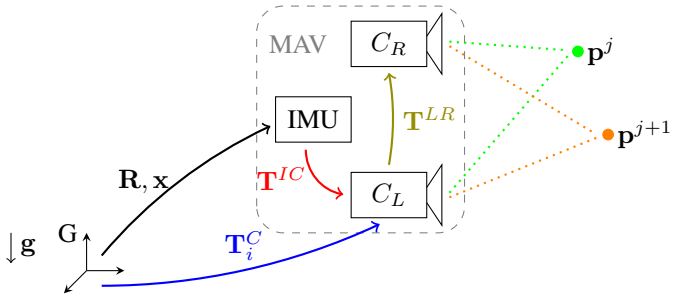

Fig. 1. The coordinate systems that are used in the paper. The IMU pose $(\mathbf{R}, \mathbf{x})$ maps vectors expressed in the IMU frame to the global frame (G). The transformations $\mathbf{T}^{IC}$ and $\mathbf{T}^{LR}$ pass, respectively, from the IMU frame to the left camera (C_L) frame, and from the left camera frame to the right camera (C_R) frame. The pose $\mathbf{T}_i^C$ consists of the position of C_L in the global frame and the rotation mapping vectors expressed in the 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.

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

$$\text{IMU-related state} \begin{cases} \dot{\mathbf{R}} = \mathbf{R}\left(\boldsymbol{\omega} - \mathbf{b}_{\boldsymbol{\omega}} + \mathbf{n}_{\boldsymbol{\omega}}\right)_\times \\ \dot{\mathbf{v}} = \mathbf{R}\left(\mathbf{a} - \mathbf{b}_{\mathbf{a}} + \mathbf{n}_{\mathbf{a}}\right) + \mathbf{g} \\ \dot{\mathbf{x}} = \mathbf{v} \\ \dot{\mathbf{b}}_{\boldsymbol{\omega}} = \mathbf{n}_{\mathbf{b}_{\boldsymbol{\omega}}} \\ \dot{\mathbf{b}}_{\mathbf{a}} = \mathbf{n}_{\mathbf{b}_{\mathbf{a}}} \end{cases} \tag{1}$$

$$\text{camera-related state} \begin{cases} \dot{\mathbf{T}}^{IC} = \mathbf{0} \\ \dot{\mathbf{T}}_i^C = \mathbf{0}, \ i = 1, \dots, N \end{cases} \tag{2}$$

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 $\mathbf{b}_{\boldsymbol{\omega}} \in \mathbb{R}^3$ and $\mathbf{b}_{\mathbf{a}} \in \mathbb{R}^3$, as well as the relative transformation between the IMU frame and the left camera frame $\mathbf{T}^{IC} \in SE(3)$ and an arbitrary number of N recorded left camera poses $\mathbf{T}_i^C \in SE(3)$, $i = 1, \dots, N$, see Figure 1. Finally, $(\boldsymbol{\omega})_\times$ denotes the skew symmetric matrix associated with the cross product with vector $\boldsymbol{\omega} \in \mathbb{R}^3$, and the various white Gaussian noises can be stacked as

$$\mathbf{n} = \left[\mathbf{n}_{\boldsymbol{\omega}}^T \ \mathbf{n}_{\mathbf{a}}^T \ \mathbf{n}_{\mathbf{b}_{\boldsymbol{\omega}}}^T \ \mathbf{n}_{\mathbf{b}_{\mathbf{a}}}^T\right]^T \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right). \tag{3}$$

These equations model the dynamics of small MAVs such as quadrotors where the IMU measurements $\boldsymbol{\omega}$ and $\mathbf{a}$ 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

$$\mathbf{y}_i^j = h\left(\mathbf{T}_i^C, \mathbf{p}^j\right) + \mathbf{n}_{\mathbf{y}} \in \mathbb{R}^4, \tag{4}$$

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

$$h\left(\mathbf{T}, \mathbf{p}\right) = \begin{bmatrix} \frac{1}{z_l}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \frac{1}{z_r}\mathbf{I} \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ x_r \\ y_r \end{bmatrix}, \tag{5}$$

in which $\mathbf{p}_l = \begin{bmatrix} x_l & y_l & z_l \end{bmatrix}^T$ and $\mathbf{p}_r = \begin{bmatrix} x_r & y_r & z_r \end{bmatrix}^T$ refer to the landmark coordinates expressed in the left and in the right camera frames (i.e., $\begin{bmatrix} \mathbf{p}_l^T & 1 \end{bmatrix}^T = \mathbf{T}^{-1} \begin{bmatrix} \mathbf{p}^T & 1 \end{bmatrix}^T$ and $\begin{bmatrix} \mathbf{p}_r^T & 1 \end{bmatrix}^T = \left(\mathbf{T}\mathbf{T}^{LR}\right)^{-1} \begin{bmatrix} \mathbf{p}^T & 1 \end{bmatrix}^T$). Note that the stereo cameras have different poses at the same time instance, represented as $\mathbf{T}_i^C$ for the left camera and $\mathbf{T}_i^C \mathbf{T}^{LR}$ for the right camera, although the state only contains the pose of the left camera, since using the assumed known extrinsic parameters $\mathbf{T}^{LR} \in SE(3)$ leads to an expression for the pose of the right camera, see Figure 1.

C. Estimation/Fusion Problem

We would like to compute the probability distribution of the system's state $(\mathbf{R}, \mathbf{x}, \mathbf{v}, \mathbf{b}_{\boldsymbol{\omega}}, \mathbf{b}_{\mathbf{a}}, \mathbf{T}^{IC}, \mathbf{T}_1^C, \ldots, \mathbf{T}_N^C)$ defined through an initial Gaussian prior and the probabilistic evolution model (1)-(2), *conditionally* on the measurements of the form (4). This is the standard probabilistic formulation of the (Stereo-)MSCKF [9].

III. Mathematical Preliminaries

In this section we provide the reader with the bare minimum about matrix Lie groups and the UKF on Lie Groups (UKF-LG) introduced in [3].

A. Matrix Lie Groups

A matrix Lie group $G \subset \mathbb{R}^{N \times N}$ is a subset of square invertible matrices such that the following properties hold:

$$\mathbf{I} \in G;\ \forall \boldsymbol{\chi} \in G, \boldsymbol{\chi}^{-1} \in G;\ \forall \boldsymbol{\chi}_1, \boldsymbol{\chi}_2 \in G, \boldsymbol{\chi}_1 \boldsymbol{\chi}_2 \in G. \quad (6)$$

Locally about the identity matrix $\mathbf{I}$, the group G can be identified with an Euclidean space $\mathbb{R}^q$ using the matrix exponential map $\exp_m(\cdot)$, where $q = \dim G$. Indeed, to any $\boldsymbol{\xi} \in \mathbb{R}^q$ one can associate a matrix $\boldsymbol{\xi}^\wedge$ of the tangent space of G at $\mathbf{I}$, called the Lie algebra $\mathfrak{g}$. We then define the exponential map $\exp_G : \mathbb{R}^q \to G$ for Lie groups as

$$\exp_G(\boldsymbol{\xi}) = \exp_m(\boldsymbol{\xi}^\wedge), \quad (7)$$

Locally, it is a bijection, and one can define the Lie logarithm map $\log_G : G \to \mathbb{R}^q$ as the exponential inverse, leading to

$$\log_G(\exp_G(\boldsymbol{\xi})) = \boldsymbol{\xi}. \quad (8)$$

B. Uncertainty on Lie Groups

To define random variables on Lie groups, we cannot apply the usual approach of additive noise for $\boldsymbol{\chi} \in G$ as G is not a vector space. In contrast, we define the probability distribution $\boldsymbol{\chi} \sim \mathcal{N}_R(\bar{\boldsymbol{\chi}}, \mathbf{P})$ for the random variable $\boldsymbol{\chi} \in G$ as [17,19]

$$\boldsymbol{\chi} = \exp_G(\boldsymbol{\xi}) \bar{\boldsymbol{\chi}},\ \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}), \quad (9)$$

where $\mathcal{N}(\cdot, \cdot)$ is the classical Gaussian distribution in Euclidean space $\mathbb{R}^q$ and $\mathbf{P} \in \mathbb{R}^{q \times q}$ is a covariance matrix. In (9), the original Gaussian $\boldsymbol{\xi}$ of the Lie algebra is moved over by right multiplication to be centered at $\bar{\boldsymbol{\chi}} \in G$, hence the letter R which stands for "right", this type of uncertainty being also referred to as right-equivariant [3]. In (9), $\bar{\boldsymbol{\chi}}$ may represent a large, noise-free and deterministic value, whereas $\mathbf{P}$ is the covariance of the small, noisy perturbation $\boldsymbol{\xi}$. We stress that we have defined this probability density function directly in the vector space $\mathbb{R}^q$ such that $\mathcal{N}_R(\cdot, \cdot)$ is not Gaussian distribution.

Remark 1: one can similarly define the distribution $\boldsymbol{\chi} \sim \mathcal{N}_L(\bar{\boldsymbol{\chi}}, \mathbf{P})$ for left multiplication of $\bar{\boldsymbol{\chi}}$, as

$$\boldsymbol{\chi} = \bar{\boldsymbol{\chi}} \exp_G(\boldsymbol{\xi}),\ \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}). \quad (10)$$

The advantages of using (9) for SLAM, instead of (10) or a standard Euclidean error can be found in [4,6].

Remark 2: defining random Gaussians on Lie groups through (9) [or (10)] is advocated notably in [17,19], and the corresponding distribution is sometimes referred to as concentrated Gaussian on Lie groups, see [25]. An alternative approach, introduced to our best knowledge in [21], and used in [26], consists in defining a (Gaussian) density directly on the group using the Haar measure. In the latter case, the group needs be unimodular, but such a requirement is in fact unnecessary to define the random variable (9).

C. Unscented Kalman Filtering on Lie Groups

By representing the state error as a variable $\boldsymbol{\xi}$ of the Lie algebra, we can build two alternative unscented filters for any state living in Lie groups following the methodology recently introduced in [3]. Let us consider a discrete time dynamical system of the form

$$\boldsymbol{\chi}_{n+1} = f(\boldsymbol{\chi}_n, \mathbf{u}_n, \mathbf{w}_n), \quad (11)$$

where the state $\boldsymbol{\chi}_n$ lives in G, $\mathbf{u}_n$ is a known input variable and $\mathbf{w}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_n)$ is a white Gaussian noise, associated with generic discrete measurements of the form

$$\mathbf{y}_n = h(\boldsymbol{\chi}_n, \mathbf{v}_n), \quad (12)$$

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

1) Right-UKF-LG: the state is modeled as $\boldsymbol{\chi}_n \sim \mathcal{N}_R(\bar{\boldsymbol{\chi}}_n, \mathbf{P}_n)$, that is, using the representation (9) of the uncertainties. The mean state is thus encoded in $\bar{\boldsymbol{\chi}}_n$ and dispersion in $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_n)$. The sigma points are generated based on the $\boldsymbol{\xi}$ variables, and mapped to the group through the model (9). Note that, this is in slight contrast with [7,26], which generate sigma points through a distribution defined directly on the group. The filter consists of two steps along the lines of the conventional UKF: propagation and update, and compute estimates $\bar{\boldsymbol{\chi}}_n$ and $\mathbf{P}_n$ at each n.

2) Left-UKF-LG: the state is alternatively modeled as $\boldsymbol{\chi}_n \sim \mathcal{N}_L(\bar{\boldsymbol{\chi}}_n, \mathbf{P}_n)$, 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

$$\mathbf{y} = h\left(\exp_G(\boldsymbol{\xi})\bar{\boldsymbol{\chi}}\right) = g(\boldsymbol{\xi}) \tag{13}$$

$$\simeq \mathbf{H}\boldsymbol{\xi} + \bar{\mathbf{y}} \tag{14}$$

given a weighted discrete representation (the so-called sigma-points) of the distribution $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{P})$. The objective is thus to find the regression matrix $\mathbf{H}$ and vector $\bar{\mathbf{y}}$ that minimize the linearization error $\mathbf{e} = \mathbf{y} - (\mathbf{H}\boldsymbol{\xi} + \bar{\mathbf{y}})$. The optimal linear regression matrix is given as [23]

$$\mathbf{H} = \mathbf{P}_{\mathbf{y}\boldsymbol{\xi}}\mathbf{P}^{-1}, \tag{15}$$

where $\mathbf{P}_{\mathbf{y}\boldsymbol{\xi}}$ is the cross-correlation between $\mathbf{y}$ and $\boldsymbol{\xi}$, and $\bar{\mathbf{y}}$ is the estimated mean of $\mathbf{y}$, both computed from the unscented transform of the UKF-LG. The numerically inferred Jacobian $\mathbf{H}$ serves as a linear approximation to $\mathbf{y} - \bar{\mathbf{y}} = g(\boldsymbol{\xi}) - \bar{\mathbf{y}}$ and can then be used for the Right-EKF-LG update.

E. The Special Euclidean Group $SE_{2+p}(3)$

As early noticed in [27], the SLAM problem bears a natural Lie group structure, through the group $SE_{1+p}(3)$ for wheel odometry SLAM, that is properly introduced and leveraged in [4], to resolve some well-known consistency issues of EKF based SLAM. Some other properties have also recently been proved in [20]. Any matrix $\boldsymbol{\chi} \in SE_{2+p}(3)$ is defined as

$$\boldsymbol{\chi} = \begin{bmatrix} \mathbf{R} & \mathbf{v} & \mathbf{x} & \mathbf{p}_1 & \cdots & \mathbf{p}_p \\ \mathbf{0}_{2+p\times 3} & & & \mathbf{I}_{p+2\times p+2} & & \end{bmatrix}, \tag{16}$$

The uncertainties, defined as $\boldsymbol{\xi} = \left[\boldsymbol{\xi}_{\mathbf{R}}^T \boldsymbol{\xi}_{\mathbf{v}}^T \boldsymbol{\xi}_{\mathbf{x}}^T \boldsymbol{\xi}_{\mathbf{p}_1}^T \cdots \boldsymbol{\xi}_{\mathbf{p}_p}^T\right]^T \in \mathbb{R}^{9+3p}$, are mapped to the Lie algebra through the transformation $\boldsymbol{\xi} \mapsto \boldsymbol{\xi}^\wedge$ defined as

$$\boldsymbol{\xi}^\wedge = \begin{bmatrix} (\boldsymbol{\xi}_{\mathbf{R}})_\times & \boldsymbol{\xi}_{\mathbf{v}} & \boldsymbol{\xi}_{\mathbf{x}} & \boldsymbol{\xi}_{\mathbf{p}_1} & \cdots & \boldsymbol{\xi}_{\mathbf{p}_p} \\ & & \mathbf{0}_{2+p\times 5+p} & & & \end{bmatrix}. \tag{17}$$

The closed-form expression for the exponential map is given as

$$\exp_{SE_{2+p}(3)}(\boldsymbol{\xi}) = \mathbf{I} + \boldsymbol{\xi}^\wedge + \frac{1-\cos\left(\|\boldsymbol{\xi}_{\mathbf{R}}\|\right)}{\|\boldsymbol{\xi}_{\mathbf{R}}\|}\boldsymbol{\xi}^{\wedge 2} + \frac{\|\boldsymbol{\xi}_{\mathbf{R}}\| - \sin\left(\|\boldsymbol{\xi}_{\mathbf{R}}\|\right)}{\|\boldsymbol{\xi}_{\mathbf{R}}\|^3}\boldsymbol{\xi}^{\wedge 3}. \tag{18}$$

IV. Proposed Filters

In this section, we propose the Stereo-UKF-LG (S-UKF-LG), a VIO filtering solution which embeds the state in a specially defined and high dimensional Lie group. Our solution operates in two steps, as for any Kalman filter-based algorithm:

- a propagation step that propagates both the mean state and the error covariance, where the matrix covariance is computed with (Invariant-)EKF linearization [17] for computational efficiency.
- an update step that considers the visual information obtained from the feature tracking, in which we used as a basis the UKF-LG [3]. We additionally provide Jacobian expressions to alternatively perform (Invariant-)EKF 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 $\boldsymbol{\chi} \in G$ be the matrix that represents:

- the IMU variables $\mathbf{R}, \mathbf{v}$ and $\mathbf{x}$ through $\boldsymbol{\chi}^I \in SE_2(3)$, a group obtained by letting $p = 0$ in (16);
- the IMU bias $\boldsymbol{\chi}^{\mathbf{b}} = \left[\mathfrak{b}_{\boldsymbol{\omega}}^T \ \mathfrak{b}_{\mathbf{a}}^T\right]^T \in \mathbb{R}^6$;
- 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

$$\boldsymbol{\chi}^I, \boldsymbol{\chi}^{\mathbf{b}}, \mathbf{T}^{IC}, \mathbf{T}_1^C, \ldots, \mathbf{T}_N^C \triangleq \boldsymbol{\chi} \sim \mathcal{N}_R\left(\bar{\boldsymbol{\chi}}, \boldsymbol{\xi}\right), \tag{19}$$

where $\triangleq$ stands for "identifiable to" and $\bar{(\cdot)}$ for estimated mean value, is partitioned into

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_{\mathbf{R}}^T & \boldsymbol{\xi}_{\mathbf{v}}^T & \boldsymbol{\xi}_{\mathbf{x}}^T & \boldsymbol{\xi}_{\boldsymbol{\omega}}^T & \boldsymbol{\xi}_{\mathbf{a}}^T & \boldsymbol{\xi}_{IC}^T & \boldsymbol{\xi}_{C_1}^T & \cdots & \boldsymbol{\xi}_{C_N}^T \end{bmatrix}^T, \tag{20}$$

and encoded using the right uncertainty (9), i.e., the uncertainty representation is defined as

$$\exp\left(\boldsymbol{\xi}\right)\bar{\boldsymbol{\chi}} \triangleq \begin{cases} \exp_{SE_2(3)}\left(\left[\boldsymbol{\xi}_{\mathbf{R}}^T \ \boldsymbol{\xi}_{\mathbf{v}}^T \ \boldsymbol{\xi}_{\mathbf{x}}^T\right]^T\right)\boldsymbol{\chi}^I \\ \left[\boldsymbol{\xi}_{\boldsymbol{\omega}}^T \ \boldsymbol{\xi}_{\mathbf{a}}^T\right]^T + \boldsymbol{\chi}^{\mathbf{b}} \\ \exp_{SE(3)}\left(\boldsymbol{\xi}_{IC}\right)\mathbf{T}^{IC} \\ \exp_{SE(3)}\left(\boldsymbol{\xi}_{C_i}\right)\mathbf{T}_i^C, i = 1, \ldots, N \end{cases} \tag{21}$$

and we define for convenience the IMU error as $\boldsymbol{\xi}_I = \left[\boldsymbol{\xi}_{\mathbf{R}}^T \ \boldsymbol{\xi}_{\mathbf{v}}^T \ \boldsymbol{\xi}_{\mathbf{x}}^T \ \boldsymbol{\xi}_{\boldsymbol{\omega}}^T \ \boldsymbol{\xi}_{\mathbf{a}}^T\right]^T$. Any unknown feature $\mathbf{p}^j$, 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 $\left(\mathbf{T}_1^C, \mathbf{p}^j\right)$ as a element of the Lie group $SE_2(3)$ [4]. Note that, error $\boldsymbol{\xi}_{\mathbf{p}^j}$ on landmark j then differs from the standard Euclidean error.

Remark 3: using another camera pose than $\mathbf{T}_1^C$ 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 $\bar{\boldsymbol{\chi}}$. To propagate the uncertainty of the state, let us consider the dynamic of the IMU linearized error as

$$\dot{\boldsymbol{\xi}}_I = \mathbf{F}\boldsymbol{\xi}_I + \mathbf{G}\mathbf{n}, \tag{22}$$

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{R} \\ (\mathbf{g})_\times & \mathbf{0} & \mathbf{0} & -\mathbf{R} & -(\mathbf{v})_\times \mathbf{R} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & -(\mathbf{x})_\times \mathbf{R} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{23}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0} & \mathbf{R} & \mathbf{0} & \mathbf{0} \\ \mathbf{R} & (\mathbf{v})_\times \mathbf{R} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{x})_\times \mathbf{R} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{24}$$

and first compute the discrete time state transition matrix

$$\mathbf{\Phi}_n = \mathbf{\Phi}\left(t_{n+1}, t_n\right) = \exp_m\left(\int_{t_n}^{t_{n+1}} \mathbf{F}(\tau) d\tau\right) \tag{25}$$

and discrete time noise covariance matrix

$$\mathbf{Q}_n = \int_{t_n}^{t_{n+1}} \mathbf{\Phi}\left(t_{n+1}, \tau\right) \mathbf{G}\mathbf{Q}\mathbf{G}^T \mathbf{\Phi}\left(t_{n+1}, \tau\right)^T d\tau. \tag{26}$$

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

$$\mathbf{P}_{n+1}^{II} = \mathbf{\Phi}_n \mathbf{P}_n^{II} \mathbf{\Phi}_n + \mathbf{Q}_n, \tag{27}$$

and the full uncertainty propagation can be computed as

$$\mathbf{P}_{n+1} = \begin{bmatrix} \mathbf{P}_{n+1}^{II} & \mathbf{\Phi}_n \mathbf{P}_n^{IC} \\ \mathbf{P}_n^{CI} \mathbf{\Phi}_n^T & \mathbf{P}_n^{CC} \end{bmatrix}. \tag{28}$$

When new images are received, the state should be augmented with the new camera state. The new augmented covariance matrix becomes

$$\mathbf{P}_{n+1} = \begin{bmatrix} \mathbf{I} \\ \mathbf{J} \end{bmatrix} \mathbf{P}_{n+1} \begin{bmatrix} \mathbf{I} \\ \mathbf{J} \end{bmatrix}^T. \tag{29}$$

To obtain the expression of $\mathbf{J}$ in (29), let us denote $\mathbf{T} \in SE(3)$ as the sub-matrix of $\boldsymbol{\chi}^I$ that contain only the IMU pose $(\mathbf{R}, \mathbf{x})$ with $\boldsymbol{\xi}_\mathbf{T} = \left[\, \boldsymbol{\xi}_\mathbf{R}^T \;\; \boldsymbol{\xi}_\mathbf{x}^T \,\right]^T$ its corresponding uncertainties and thus write the current camera pose as

$$\begin{aligned} \mathbf{T}^C &= \mathbf{T}\mathbf{T}^{IC} \\ &= \exp_{SE(3)}\left(\boldsymbol{\xi}_\mathbf{T}\right) \bar{\mathbf{T}} \exp_{SE(3)}\left(\boldsymbol{\xi}_{IC}\right) \bar{\mathbf{T}}^{IC} \\ &= \exp_{SE(3)}\left(\boldsymbol{\xi}_\mathbf{T}\right) \exp_{SE(3)}\left(\mathrm{Ad}_{\bar{\mathbf{T}}}\, \boldsymbol{\xi}_{IC}\right) \bar{\mathbf{T}}\bar{\mathbf{T}}^{IC} \\ &\simeq \exp_{SE(3)}\left(\boldsymbol{\xi}_C\right) \bar{\mathbf{T}}^C, \end{aligned} \tag{30}$$

in which $\bar{\mathbf{T}}^C = \bar{\mathbf{T}}\bar{\mathbf{T}}^{IC}$ and $\boldsymbol{\xi}_C \simeq \boldsymbol{\xi}_\mathbf{T} + \mathrm{Ad}_{\bar{\mathbf{T}}}\, \boldsymbol{\xi}_{IC}$ after using a first order BCH approximation, and where $\mathrm{Ad}_{\bar{\mathbf{T}}}$ is the adjoint notation of $SE(3)$, which finally leads to

$$\mathbf{J} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0}_{3\times 6} & \mathbf{R} & \mathbf{0} & \mathbf{0} & \mathbf{0}_{3\times 6N} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}_{3\times 6} & (\mathbf{x})_\times \mathbf{R} & \mathbf{0} & \mathbf{R} & \mathbf{0}_{3\times 6N} \end{bmatrix}. \tag{31}$$

C. Update Step in the MSCKF Methodology

Let us first consider the observation of a single feature $\mathbf{p}^j$, in which the estimated unbiased feature position ${}^{\mathsf{LS}}\mathbf{p}^j$ is computed using least squares estimate based on the current estimated camera poses [9]. Linearizing the measurement model at the obtained estimates $\bar{\boldsymbol{\chi}}$, ${}^{\mathsf{LS}}\mathbf{p}^j$, the residual of the measurement is approximated as

$$\mathbf{r}_i^j = \mathbf{y}_i^j - {}^{\mathsf{UKF}}\mathbf{y}_i^j = {}^{\mathsf{UKF}}\mathbf{H}_i^j \boldsymbol{\xi} + \mathbf{H}_{\mathbf{p}^j}^i \boldsymbol{\xi}_{\mathbf{p}^j} + \mathbf{n}_i^j, \tag{32}$$

where ${}^{\mathsf{UKF}}\mathbf{H}_i^j$, $\mathbf{H}_{\mathbf{p}^j}^i$ and ${}^{\mathsf{UKF}}\mathbf{y}_i^j$ 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 $\boldsymbol{\xi}$, $\boldsymbol{\xi}_{\mathbf{p}^j}$ related to the Lie group structure we have endowed the state space with. By stacking multiple observations of the same feature $\mathbf{p}^j$, we then dispose of

$$\mathbf{r}^j = \mathbf{y}^j - {}^{\mathsf{UKF}}\mathbf{y}^j = {}^{\mathsf{UKF}}\mathbf{H}^j \boldsymbol{\xi} + \mathbf{H}_{\mathbf{p}^j} \boldsymbol{\xi}_{\mathbf{p}^j} + \mathbf{n}^j. \tag{33}$$

Then, along the lines of [9], to eliminate the landmark errors from the residual, measurements are projected onto the null space $\mathbf{V}$ of $\mathbf{H}_{\mathbf{p}^j}^i$, i.e.,

$$\mathbf{r}_o^j = \mathbf{V}^T \mathbf{r}^j = \mathbf{V}^T\, {}^{\mathsf{UKF}}\mathbf{H}^j \boldsymbol{\xi} + \mathbf{V}^T \mathbf{n}^j = \mathbf{H}_o^j \boldsymbol{\xi} + \mathbf{n}_o^j. \tag{34}$$

Based on (34) and after stacking residual and Jacobians for multiple landmarks in, respectively, $\mathbf{r}_o$ and $\mathbf{H}_o$, the update step is carried out by first computing $\mathbf{S} = \mathbf{R} + \mathbf{H}_o \mathbf{P}_n \mathbf{H}_o^T$ and the gain matrix $\mathbf{K} = \mathbf{P}_n \mathbf{H}_o^T / \mathbf{S}$. We then compute the innovation $\bar{\boldsymbol{\xi}}$ to update the mean state as [3,17]

$$\bar{\boldsymbol{\xi}} = \mathbf{K}\mathbf{r}_o, \tag{35}$$

$$\bar{\boldsymbol{\chi}}^+ = \exp\left(\bar{\boldsymbol{\xi}}\right) \bar{\boldsymbol{\chi}}, \tag{36}$$

which is an update that is consistent with our uncertainties (21) defined using the right multiplication based representation (9). The associated covariance matrix writes

$$\mathbf{P}_n^+ = \mathbf{P}_n \left(\mathbf{I} - \mathbf{K}\mathbf{H}_o\right). \tag{37}$$

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 ${}^{\mathsf{UKF}}\mathbf{H}_i^j$, $\mathbf{H}_{\mathbf{p}^j}^i$ and ${}^{\mathsf{UKF}}\mathbf{y}_i^j$ in (32) following Section III-D.

1) Computation of $\mathbf{H}_{\mathbf{p}^j}^i$: since the covariance of $\boldsymbol{\xi}_\mathbf{p}^j$ is unknown, we can not apply UKF-LG update and consequently we compute $\mathbf{H}_{\mathbf{p}^j}^i$ in closed-form, as $\mathbf{H}_{\mathbf{p}^j}^1 = \mathbf{0}$, and for $i > 1$, as

$$\mathbf{H}_{\mathbf{p}^j}^i = \mathbf{J}_i^j \begin{bmatrix} \left(\bar{\mathbf{R}}_i^L\right)^T \\ \left(\bar{\mathbf{R}}_i^R\right)^T \end{bmatrix}, \tag{38}$$

where $\bar{\mathbf{R}}_i^L$ and $\bar{\mathbf{R}}_i^R$ are the orientations of the i-th left and right cameras, and where

$$\mathbf{J}_i^j = \begin{bmatrix} 1/\bar{z}_l & 0 & -\bar{x}_l/\bar{z}_l^2 \\ 0 & 1/\bar{z}_l & -\bar{y}_l/\bar{z}_l^2 \\ 1/\bar{z}_r & 0 & -\bar{x}_r/\bar{z}_r^2 \\ 0 & 1/\bar{z}_r & -\bar{y}_r/\bar{z}_r^2 \end{bmatrix} \tag{39}$$

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 $\boldsymbol{\xi}_{\mathbf{p}^j}$.

2) Computation of ${}^{\mathsf{UKF}}\mathbf{H}_i^j$ and ${}^{\mathsf{UKF}}\mathbf{y}_i^j$: we apply the UKF-LG [3] to numerically infer the Jacobian ${}^{\mathsf{UKF}}\mathbf{H}_i^j$ and estimated measurement ${}^{\mathsf{UKF}}\mathbf{y}_i^j$ (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 $\mathbf{y}_i$ that can be written in the form $\mathbf{y}_i = g_i\left(\boldsymbol{\xi}_{C_1}, \boldsymbol{\xi}_{C_i}\right)$, when noise and landmark errors are marginalized, as follow. Since

$\mathbf{y}_i$ is the concatenation of (4) for multiple landmarks, it is sufficient to write $\mathbf{y}_i^j$ as a function of $\boldsymbol{\xi}_{C_1}$ and $\boldsymbol{\xi}_{C_i}$ only. Let us first write our alternative state error as

$$\exp_{SO(3)}(\boldsymbol{\xi}_{\mathbf{R}_{C_i}}) = \mathbf{R}_{C_i}\bar{\mathbf{R}}_{C_i}^T, \tag{40}$$

$$\boldsymbol{\xi}_{\mathbf{x}_{C_i}} = \mathbf{R}_{C_i}\bar{\mathbf{R}}_{C_i}^T\mathbf{x}_{C_i} - \bar{\mathbf{x}}_{C_i}, \tag{41}$$

$$\boldsymbol{\xi}_{\mathbf{p}^j} = \mathbf{R}_{C_1}\bar{\mathbf{R}}_{C_1}^T\mathbf{p}^j - {}^{\mathsf{LS}}\mathbf{p}^j, \tag{42}$$

and consider the measurement function of the stereo camera such that

$$\mathbf{y}_i^j = l\left(\mathbf{R}_{C_i}^T\left(\mathbf{p}^j - \mathbf{x}_{C_i}\right)\right), \tag{43}$$

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

$$\begin{aligned}\mathbf{y}_i^j = l(\bar{\mathbf{R}}_{C_i}^T(\exp_{SO(3)}(\boldsymbol{\xi}_{\mathbf{R}_{C_i}})\exp_{SO(3)}(-\boldsymbol{\xi}_{\mathbf{R}_{C_1}})\,{}^{\mathsf{LS}}\mathbf{p}^j \\ -\bar{\mathbf{R}}_{C_i}^T(\mathbf{x}_{C_i} + \boldsymbol{\xi}_{\mathbf{x}_{C_i}})),\end{aligned} \tag{44}$$

which depends on $\boldsymbol{\xi}_{\mathbf{R}_{C_1}}$ and $\boldsymbol{\xi}_{C_i} = [\,\boldsymbol{\xi}_{\mathbf{R}_{C_i}}\ \boldsymbol{\xi}_{\mathbf{x}_{C_i}}\,]$ only, such that we can write $\mathbf{y}_j^i$ and by extension $\mathbf{y}^i$ as function of $\boldsymbol{\xi}_{\mathbf{R}_{C_1}}$ and $\boldsymbol{\xi}_{C_i}$ only. Following then [16], we sample only sigma-points from variables the observations depend on, i.e., from the rotational part of $\boldsymbol{\xi}_{C_1}$ and $\boldsymbol{\xi}_{C_i}$, 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 $\mathbf{y}_i^j = \left(\bar{\mathbf{T}}_i^C, \bar{\mathbf{p}}^j\right)$ is alternatively computed with estimated state, $\mathbf{H}_{\mathbf{p}^j}^i$ in (38) and $\mathbf{H}_i^j$ as

$$\mathbf{H}_i^j = \mathbf{J}_i^j\begin{bmatrix}\mathbf{0} & \mathbf{H}_{C_1}^j & \mathbf{0} & \mathbf{H}_{C_i}^j & \mathbf{0}\end{bmatrix}, \tag{45}$$

where $\mathbf{J}_i^j$ is computed in (39), for $i > 1$,

$$\mathbf{H}_{C_1}^j = \begin{bmatrix}-\left(\bar{\mathbf{R}}_i^L\right)^T\left({}^{\mathsf{LS}}\mathbf{p}^j\right)_\times & \mathbf{0} \\ -\left(\bar{\mathbf{R}}_i^R\right)^T\left({}^{\mathsf{LS}}\mathbf{p}^j\right)_\times & \mathbf{0}\end{bmatrix}, \tag{46}$$

$$\mathbf{H}_{C_i}^j = \begin{bmatrix}\left(\bar{\mathbf{R}}_i^L\right)^T\left({}^{\mathsf{LS}}\mathbf{p}^j\right)_\times & -\left(\bar{\mathbf{R}}_i^L\right)^T \\ \left(\bar{\mathbf{R}}_i^R\right)^T\left({}^{\mathsf{LS}}\mathbf{p}^j\right)_\times & -\left(\bar{\mathbf{R}}_i^R\right)^T\end{bmatrix}, \tag{47}$$

and, for $i = 1$,

$$\mathbf{H}_{C_1}^j = \mathbf{H}_{C_i}^j = \begin{bmatrix}\mathbf{0} & -\left(\bar{\mathbf{R}}_1^L\right)^T \\ \mathbf{0} & -\left(\bar{\mathbf{R}}_1^R\right)^T\end{bmatrix}. \tag{48}$$

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.

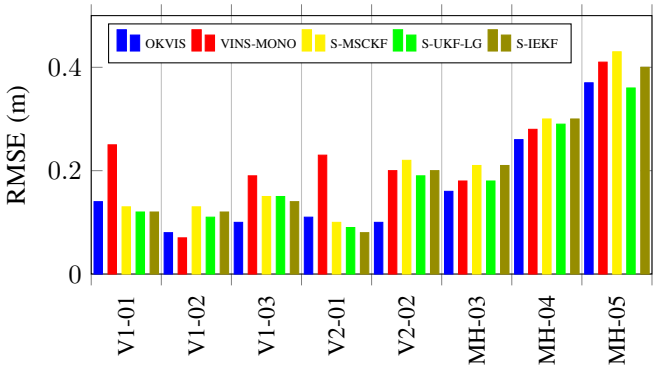

Fig. 2. Root Mean Square Error of the proposed S-UKF-LG and S-IEKF compared to various methods on the EuRoC dataset [24]. Statistics are averaged over ten runs on each sequence.

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-

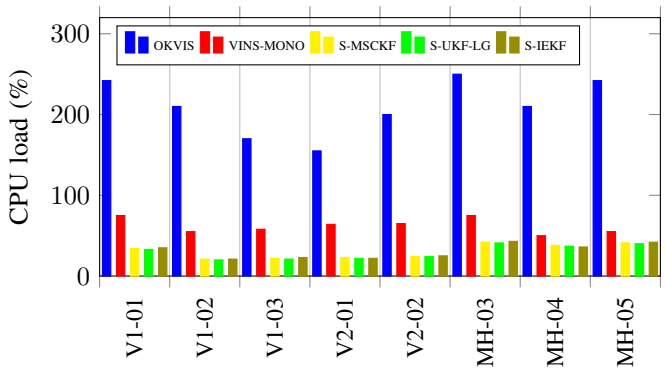

Fig. 3. Average CPU load of the proposed S-UKF-LG and S-IEKF compared to various methods on the EuRoC dataset [24].

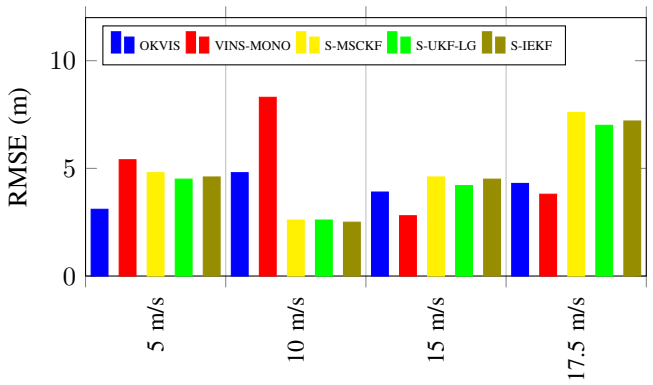

Fig. 4. Root Mean Square Error of the proposed S-UKF-LG and S-IEKF compared to various methods on the dataset [2]. Statistics are averaged over ten runs on each dataset.

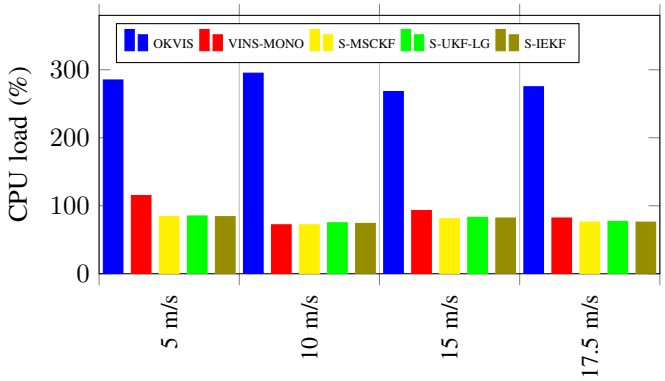

Fig. 5. Average CPU load of the proposed S-UKF-LG and S-IEKF compared to various methods on the dataset [2].

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 $5\,\mathrm{m/s}$, $10\,\mathrm{m/s}$, $15\,\mathrm{m/s}$, and $17.5\,\mathrm{m/s}$ [2]. During each sequence, the quadrotor goes $300\,\mathrm{m}$ straight and returns to the starting point. The configuration includes two cameras running at $40\,\mathrm{Hz}$ and one IMU running at $200\,\mathrm{Hz}$. 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 xy 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 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

[1] N. Michael, S. Shen, K. Mohta, Y. Mulgaonkar, V. Kumar, K. Nagatani, Y. Okada, S. Kiribayashi, K. Otake, K. Yoshida, *et al.*, "Collaborative mapping of an earthquake-damaged building via ground and aerial robots," *Journal of Field Robotics*, vol. 29, no. 5, pp. 832–841, 2012.

[2] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robotics and Automation Letters*, vol. 3, pp. 965–972, April 2018.

[3] M. Brossard, S. Bonnabel, and J.-P. Condomines, "Unscented Kalman filtering on Lie groups," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[4] A. Barrau and S. Bonnabel, "An EKF-SLAM algorithm with consistency properties," *Conditionally accepted in IEEE Transactions on Robotics. Preprint arXiv:1510.06263*, 2015.

[5] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Observability-constrained vision-aided inertial navigation," *University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Tech. Rep*, vol. 1, 2012.

[6] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, "An Invariant-EKF VINS algorithm for improving consistency," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[7] G. Loianno, M. Watterson, and V. Kumar, "Visual inertial odometry for quadrotors on SE(3)," in *IEEE International Conference On Robotics and Automation (ICRA)*, pp. 1544–1551, IEEE, 2016.

[8] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.

[9] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 3565–3572, 2007.

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

[11] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2017.

[12] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[13] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *ArXiv e-prints*, Aug. 2017.

[14] G. Huang, M. Kaess, and J. J. Leonard, "Towards consistent visual-inertial navigation," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4926–4933, May 2014.

[15] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Camera-IMU-based localization: Observability analysis and consistency improvement," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 182–201, 2014.

[16] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "A quadratic-complexity observability-constrained unscented Kalman filter for SLAM," *IEEE Transactions on Robotics*, vol. 29, no. 5, pp. 1226–1243, 2013.

[17] A. Barrau and S. Bonnabel, "The Invariant Extended Kalman Filter as a stable observer," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1797–1812, 2017.

[18] A. Barrau and S. Bonnabel, "Invariant Kalman filtering," *Annual Reviews of Control, Robotics, and Autonomous Systems (in press)*, vol. 1, 2018.

[19] T. D. Barfoot and P. T. Furgale, "Associating uncertainty with three-dimensional poses for use in estimation problems," *IEEE Transactions on Robotics*, vol. 30, no. 3, pp. 679–693, 2014.

[20] T. Zhang, K. Wu, J. Song, S. Huang, and G. Dissanayake, "Convergence and consistency analysis for a 3-D Invariant-EKF SLAM," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 733–740, 2017.

[21] G. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Applied and Numerical Harmonic Analysis, Birkhäuser Boston, 2011.

[22] S. Hauberg, F. Lauze, and K. S. Pedersen, "Unscented kalman filtering on riemannian manifolds," *Journal of mathematical imaging and vision*, vol. 46, no. 1, pp. 103–120, 2013.

[23] T. Lefebvre, H. Bruyninckx, and J. D. Schuller, "Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators" [with authors' reply]," *IEEE Transactions on Automatic Control*, vol. 47, pp. 1406–1409, Aug 2002.

[24] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. Achtelik, and R. Siegwart, "The EuRoC MAV datasets," *The International Journal of Robotics Research*, 2015.

[25] G. Bourmaud, R. Mégret, M. Arnaudon, and A. Giremus, "Continuous-discrete extended Kalman filter on matrix lie groups using concentrated Gaussian distributions," *Journal of Mathematical Imaging and Vision*, vol. 51, no. 1, pp. 209–228, 2015.

[26] M. Zefran, V. Kumar, and C. Croke, "Metrics and connections for rigid-body kinematics," *International Journal of Robotics Research*, vol. 18, no. 2, pp. 243–258, 1999.

[27] S. Bonnabel, "Symmetries in observer design: Review of some recent results and applications to ekf-based slam," in *Robot Motion and Control*, pp. 3–15, Springer, 2011.