2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

An Equivariant Filter for Visual Inertial Odometry

Pieter van Goor[1] and Robert Mahony[1]

Abstract— **Visual Inertial Odometry (VIO) is of great interest due the ubiquity of devices equipped with both a monocular camera and Inertial Measurement Unit (IMU). Methods based on the extended Kalman Filter remain popular in VIO due to their low memory requirements, CPU usage, and processing time when compared to optimisation-based methods. In this paper, we analyse the VIO problem from a geometric perspective and propose a novel formulation on a smooth quotient manifold where the equivalence relationship is the well-known invariance of VIO to choice of reference frame. We propose a novel Lie group that acts transitively on this manifold and is compatible with the visual measurements. This structure allows for the application of Equivariant Filter (EqF) design leading to a novel filter for the VIO problem. Combined with a very simple vision processing front-end, the proposed filter demonstrates state-of-the-art performance on the EuRoC dataset compared to other EKF-based VIO algorithms.**

I. Introduction

Visual Inertial Odometry (VIO) belongs to the more general class of spatial awareness problems often referred to as Simultaneous Localisation and Mapping (SLAM). SLAM algorithms are a core technology in mobile robotics and have been the subject of significant research for at least 30 years [1]. The particular problem of Visual Inertial SLAM (VI-SLAM), where the only available sensors are an Inertial Measurement Unit (IMU) and a monocular camera continues to see substantial interest due the low-cost of the required sensors and the breadth of applications [2]. Visual inertial odometry and visual inertial SLAM share the same formulation, however, the odometry problem focuses on estimating the robot trajectory while the SLAM problem places equal emphasis on the map. In practice, the difference is characterised by how long feature points are stored and whether full loop-closure is considered in the algorithms. State-of-the-art solutions to VIO can be broadly classified into optimisation-based or Extended Kalman Filter (EKF)-based systems. Optimisation-based solutions, including ORB-SLAM 3 [3], OKVIS [4] and VINS-Mono [5], treat VI-SLAM as a non-linear least squares problem and optimise over a moving window of data measurements. In contrast, EKF-based solutions, such as ROVIO [6], MSCKF [7] and SVO [8], model the state estimate as a normal distribution, linearise the state equations and apply an EKF to the resulting error coordinates. While optimisation methods typically achieve the highest accuracy in computing the robot's trajectory, EKF methods remain of interest due to their lower memory requirements and processing times [2].

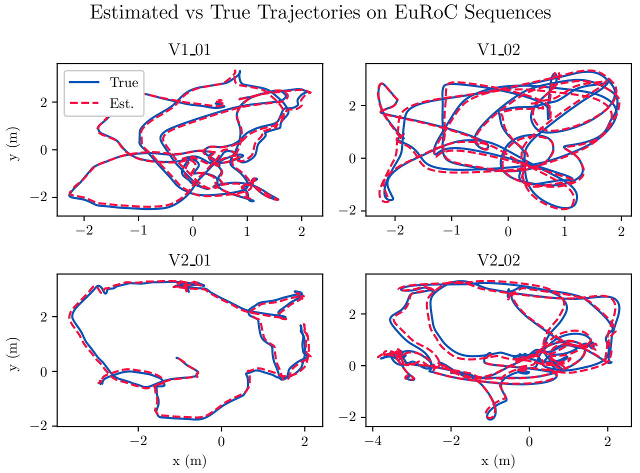

Fig. 1. The true and estimated trajectory of the robot in the considered EuRoC sequences.

Classical EKF designs for VIO are well-known to grow overconfident in their state over time [9], [10]. This inconsistency is a direct consequence of unobservability of the visual inertial SLAM problem expressed in inertial coordinates [11], [12], [13]. In [14], Kelly and Sukhatme provide a complete characterisation of the non-linear observability of visual inertial SLAM, and show that the problem has a four-dimensional unobservable subspace corresponding to the position and yaw of the reference frame. This issue can be mitigated by careful choice of linearisation points of the EKF to ensure the observability of the linearised system reflects that of the true system [13]. While this improves the consistency of the state estimate, it does not avoid the existence of unobservable subspaces in the state space and, as a consequence, the underlying Riccati equation may grow unbounded [15]. Recently, a number of authors have exploited the Invariant Extended Kalman Filter [16] and the novel Lie group structure proposed in [17] to develop filters for SLAM [17], [18], [19], [20]. These filters still suffer from unobservable states, however, invariance properties are exploited to ensure consistency of the filter, and the unbounded growth in the covariance of the unobservable state [15] can be managed using heuristics. A further limitation of the symmetry is that it is not compatible with visual feature outputs, and although the state propagation linearisation is exact, the IEKF suffers from output linearisation error for the visual SLAM problem. Mahony *et al.* [21] introduced a quotient manifold structure, termed the SLAM-manifold, that overcomes the observability issues by providing a fully

[1] Pieter van Goor and Robert Mahony are with the Systems Theory and Robotics group and the Australian Centre for Robotic Vision at the Australian National University. {first name}.{last name}@anu.edu.au

978-1-7281-9077-8/21/$31.00 ©2021 IEEE

observable state space for the SLAM problem that is geometrically motivated. Van Goor *et al.* [22] introduced a new symmetry for the SLAM problem that acts transitively on the SLAM-manifold, and in addition is compatible with visual point feature measurements, overcoming the limitation of the symmetry [17] associated with linearisation of the output function. However, although this symmetry acts transitively on the SLAM-manifold [21], the IEKF is only formulated for systems posed directly on a Lie-group and cannot be applied to systems on homogeneous spaces such as is the case for the SLAM-manifold formulation with either of the symmetries [17], [21] or [22]. In contrast, the recently proposed Equivariant Filter (EqF) [23] is explicitly posed for systems on general homogeneous spaces and can be applied.

In this paper, we derive a novel VIO filter based on equivariance principles that has state-of-the-art performance. We show that there is a natural invariance in the traditional inertial coordinates of the visual inertial SLAM formulation associated with the gauge transformation that leads to the unobservability properties that are well known [14]. We show that the quotient of the inertial coordinates by the gauge transform generates a smooth manifold we term the *VI-SLAM manifold* on which the system is fully observable. The symmetry group first proposed in [22] is easily generalised to a new Lie-group, we term the *VI-SLAM group*, that acts transitively on the *VI-SLAM manifold* and is compatible with the vision measurements of points features. This provides the geometric structure necessary to implement the Equivariant Filter (EqF) [23]. The resulting algorithm benefits from all the advantages of a complete Lie group symmetry that is compatible with the measurements as well as being fully observable, overcoming limitations of previous invariant filter algorithms for visual inertial SLAM problems. Finally, we demonstrate the performance of the proposed system on sequences in the EuRoC dataset and achieve state of the art results compared to EKF-based algorithms in spite of the simplicity of our front-end image processing system.

II. Preliminaries

For a background on smooth manifolds, Lie groups and their actions, the authors recommend [24, Chapter 7]. For a smooth manifold $\mathcal{M}$, let $\mathrm{T}_\xi\mathcal{M}$ denote the tangent space of $\mathcal{M}$ at ξ and let $\mathrm{T}\mathcal{M}$ denote the tangent bundle. Given a differentiable function between smooth manifolds $h : \mathcal{M} \to \mathcal{N}$, the linear map

$$\begin{aligned} \mathrm{D}_\xi|_{\xi'} h(\xi) : \mathrm{T}_{\xi'}\mathcal{M} &\to \mathrm{T}_{h(\xi')}\mathcal{N}, \\ v &\mapsto \mathrm{D}_\xi|_{\xi'} h(\xi)[v], \end{aligned}$$

denotes the differential of h with respect to the argument ξ evaluated at ξ'. The map

$$\begin{aligned} \mathrm{d}h : \mathrm{T}\mathcal{M} &\to \mathrm{T}\mathcal{N}, \\ (\xi', v) &\mapsto (h(\xi'), \mathrm{D}_\xi|_{\xi'} h(\xi)[v]), \end{aligned}$$

denotes the differential of h where the base point is implicit in the argument. That is, given $v \in \mathrm{T}_{\xi'}\mathcal{M}$ for some $\xi' \in \mathcal{M}$, and a function $h : \mathcal{M} \to \mathcal{N}$, we write

$$\mathrm{d}h[v] := \mathrm{D}_\xi|_{\xi'} h(\xi)[v] \in \mathrm{T}_{h(\xi')}\mathcal{N}.$$

We make extensive use of a number of Lie groups. For a Lie group $\mathbf{G}$ we denote the Lie algebra $\mathfrak{g}$. The Special Orthogonal group $\mathbf{SO}(3)$ has elements $R \in \mathbf{SO}(3)$ and acts on $q \in \mathbb{R}^3$ by $R(q) = Rq$. The Special Euclidean group $\mathbf{SE}(3)$ has elements $P = (R_P, x_P) \in \mathbf{SO}(3) \ltimes \mathbb{R}^3$ and acts on $q \in \mathbb{R}^3$ by $P(q) = R_P q + x_P$. The Extended Special Euclidean group $\mathbf{SE}_2(3)$ has elements $(A, w) \in \mathbf{SE}(3) \times \mathbb{R}^3$ with group multiplication $(A_1, w_1) \cdot (A_2, w_2) = (A_1 A_2, w_1 + R_A w_2)$ [16]. The positive multiplicative reals $\mathbf{MR}(1)$ has elements $c > 0$. The Scaled Orthogonal Transformations $\mathbf{SOT}(3)$ has elements $Q = (R_Q, c_Q) \in \mathbf{SO}(3) \times \mathbf{MR}(1)$ and acts on $q \in \mathbb{R}^3$ by $Q(q) = c_Q R_Q q$ [22].

For any $\Omega \in \mathbb{R}^3$ define

$$\Omega^\times := \begin{pmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{pmatrix}.$$

Then the Lie algebra $\mathfrak{so}(3) = \{\Omega^\times \in \mathbb{R}^{3\times 3} | \Omega \in \mathbb{R}^3\}$, and

$$\Omega^\times p = \Omega \times p = -p \times \Omega = -p^\times \Omega,$$

for any $\Omega, p \in \mathbb{R}^3$, where × is the usual cross product.

For any $\Omega, v \in \mathbb{R}^3$ define $U(\Omega, v) \in \mathfrak{se}(3)$ such that for any $P = (R_P, x_P) \in \mathbf{SE}(3)$,

$$\dot{P} = PU(\Omega, v) \ \Leftrightarrow\ \dot{R}_P = \dot{R}\Omega^\times \text{ and } \dot{x}_P = R_P v.$$

A (right) group action of a Lie group $\mathbf{G}$ on a smooth manifold $\mathcal{M}$ is a smooth function $\phi : \mathbf{G} \times \mathcal{M} \to \mathcal{M}$ satisfying

$$\phi(XY, \xi) = \phi(Y, \phi(X, \xi)), \quad \phi(\mathrm{id}, \xi) = \xi.$$

Given such a map, we denote by ϕ_X and ϕ_ξ the partial maps

$$\begin{aligned} \phi_X &: \mathcal{M} \to \mathcal{M}, & \phi_X(\xi) &:= \phi(X, \xi), \\ \phi_\xi &: \mathbf{G} \to \mathcal{M}, & \phi_\xi(X) &:= \phi(X, \xi). \end{aligned}$$

Denote the 2-sphere by $\mathrm{S}^2 = \{y \in \mathbb{R}^3 \mid \|y\| = 1\}$. Let $\pi_{\mathrm{S}^2}(q) := q/\|q\|$ be the sphere projection for all $q \in \mathbb{R}^3$ with $q \neq 0$. For $\mathbf{e}_1 = (1, 0, 0)$ let $\vartheta_{\mathbf{e}_1} : \mathcal{U}_{\mathbf{e}_1} \subset \mathrm{S}^2 \to \mathbb{R}^2$ be the stereographic projection as in [24, Chapter 1], and for every $\eta \in \mathrm{S}^2 \setminus \{\mathbf{e}_1\}$, define $\vartheta_\eta : \mathcal{U}_\eta \subset \mathrm{S}^2 \to \mathbb{R}^2$ by

$$\vartheta_\eta(y) := \vartheta_{\mathbf{e}_1}(y - 2\zeta\zeta^\top y), \quad \zeta := \pi_{\mathrm{S}^2}(\mathbf{e}_1 - \eta), \qquad (1)$$

where $\mathcal{U}_\eta = \mathrm{S}^2 \setminus \{-\eta\}$. Then for each $\eta \in \mathrm{S}^2$, ϑ_η is a local coordinate chart with $\vartheta_\eta(\eta) = 0$.

Let

$$\mathbb{S}_+(n) = \{S = S^\top \in \mathbb{R}^{n\times n} | x^\top S x > 0, \text{ for all } x \neq 0 \in \mathbb{R}^n\}$$

denote the set of positive definite symmetric matrices of dimension of n.

III. Problem Description

Choose an arbitrary inertial reference frame $\{0\}$, and consider a robot equipped with an IMU and a camera, both of which are rigidly attached. For simplicity we identify the IMU frame $\{I\}$ with the robot's body-fixed frame $\{B\}$. The inertial coordinates for the visual inertial SLAM problem are

$$(P, v, p_1, ..., p_n) \in \mathbf{SE}(3) \times \mathbb{R}^3 \times (\mathbb{R}^3)^n, \qquad (2)$$

where,

- $P = (R_P, x_P) \in \mathbf{SE}(3)$ is the pose of the IMU $\{B\}$ with respect to the inertial frame $\{0\}$,
- $v \in \mathbb{R}^3$ is the linear velocity of the robot in the body-fixed frame $\{B\}$,
- $p_i \in \mathbb{R}^3$ is the coordinates of landmark i in the inertial frame $\{0\}$.

We frequently use the notation $(P, v, p_i) \equiv (P, v, p_1, ..., p_n)$ as shorthand. To ensure that the visual measurements are always well defined we assume that the trajectory considered never passes through an exception set $\mathcal{E} \subset \mathbf{SE}(3) \times \mathbb{R}^3 \times (\mathbb{R}^3)^n$ corresponding to all situations where the camera centre coincides with a landmark point. To formalise this, we define the visual inertial SLAM (VI-SLAM) total space

$$\mathcal{T}_n^{\mathrm{VI}}(3) := \mathbf{SE}(3) \times \mathbb{R}^3 \times (\mathbb{R}^3)^n - \mathcal{E}$$

and consider the visual inertial SLAM problem on $\mathcal{T}_n^{\mathrm{VI}}(3)$. Note that $\mathcal{T}_n^{\mathrm{VI}}(3)$ is an open subset of a smooth manifold and as such is itself a smooth manifold.

Let the acceleration due to gravity in the inertial frame $\{0\}$ be $g\mathbf{e}_3$, where $g \approx 9.81$ m/s^2 and $\mathbf{e}_3 \in \mathrm{S}^2$ is standard gravity direction in the inertial frame. The ideal IMU measurements are $(\Omega, a) \in \mathbb{R}^3 \times \mathbb{R}^3$, the angular velocity and linear acceleration of the IMU, respectively. The VIO system dynamics are

$$\frac{\mathrm{d}}{\mathrm{d}t}(P, v, p_i) = f_{(\Omega,a)}(P, v, p_i), \qquad (3)$$

$$\dot{P} = PU(\Omega, v), \quad \dot{v} = -\Omega^\times v + a - gR_P^\top \mathbf{e}3, \quad \dot{p}_i = 0.$$

The camera measurements are modelled as n bearing measurements of the landmarks p_i in the camera frame $\{C\}$ on the manifold $\mathcal{N}_n^{\mathrm{V}}(3) := (\mathrm{S}^2)^n$ where the superscript "V" stands for visual measurements. Let $T_C \in \mathbf{SE}(3)$ denote the pose of the camera frame $\{C\}$ with respect the body frame $\{B\}$. We do not consider the online calibration of T_C in the present work. Then the measurement function $h : \mathcal{T}_n^{\mathrm{VI}}(3) \to \mathcal{N}_n^{\mathrm{V}}(3)$ is given by

$$\begin{aligned} h(P, v, p_i) &:= \left(h^1(P, v, p_i), ...h^n(P, v, p_i)\right), \qquad (4) \\ h^k(P, v, p_i)) &:= \pi_{\mathrm{S}^2}\left((PT_C)^{-1}(p_k)\right). \end{aligned}$$

Modelling the bearing measurements directly on the sphere rather than the image plane enables the proposed system to model a wide variety of monocular cameras.

A. Invariance of Visual Inertial SLAM

Let $\mathbf{e}_3$ be the standard gravity direction and define the semi-direct product group

$$\mathrm{S}^1 \ltimes_{\mathbf{e}_3} \mathbb{R}^3 := \{(\theta, x) \,|\, \theta \in \mathrm{S}^1, x \in \mathbb{R}^3\},$$

with group product, identity and inverse

$$\begin{aligned} (\theta^1, x^1) \cdot (\theta^2, x^2) &= (\theta^1 + \theta^2, x^1 + R_{\mathbf{e}_3}(\theta^1)x^2), \\ \mathrm{id}_{\mathrm{S}^1 \ltimes_{\mathbf{e}_3} \mathbb{R}^3} &= (0, 0_{3\times 1}), \\ (\theta, x)^{-1} &= (-\theta, -R_{\mathbf{e}_3}(\theta)x), \end{aligned}$$

where $R_{\mathbf{e}_3}(\theta) \in SO(3)$ is the anti-clockwise rotation of an angle θ about the axis $\mathbf{e}_3$. Then $\mathrm{S}^1 \ltimes_{\mathbf{e}_3} \mathbb{R}^3$ may be identified with the subgroup

$$\mathbf{SE}_{\mathbf{e}_3}(3) := \{(R, x) \in \mathbf{SE}(3) \,|\, R\mathbf{e}_3 = \mathbf{e}_3\} \leq \mathbf{SE}(3).$$

Define $\alpha : \mathbf{SE}_{\mathbf{e}_3} \times \mathcal{T}_n^{\mathrm{VI}}(3) \to \mathcal{T}_n^{\mathrm{VI}}(3)$ by

$$\alpha(S, (P, v, p_i)) := (S^{-1}P, v, S^{-1}(p_i)).$$

Then α is a (right) group action of $\mathbf{SE}_{\mathbf{e}_3}(3)$ on $\mathcal{T}_n^{\mathrm{VI}}(3)$. For a given $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$, the action $\alpha(S, \cdot)$ represents a change of inertial reference frame from $\{0\}$ to $\{1\}$ where S is the pose of $\{1\}$ with respect to $\{0\}$. Moreover, any change of reference $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$ leaves the direction of gravity $\mathbf{e}_3$ unchanged.

Proposition 3.1: The system function (3) and measurement function (4) are invariant with respect to α, that is,

$$\begin{aligned} f_{(\Omega,a)}(\alpha(S, (P, v, p_i))) &= \mathrm{d}\alpha_S f_{\Omega,a}(P, v, p_i), \\ h(\alpha(S, (P, v, p_i))) &= h(P, v, p_i), \end{aligned}$$

for any $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$.

A proof is provided in Appendix I.

B. VI-SLAM Manifold

We exploit the invariance of the dynamics and measurements to propose a new state space where the system is fully observable. Given any $(P, v, p_i) \in \mathcal{T}_n^{\mathrm{VI}}(3)$, define the equivalence class

$$[P, v, p_i] = \{\alpha(S, (P, v, p_i)) \,|\, S \in \mathbf{SE}_{\mathbf{e}_3}(3)\}.$$

Since α is a proper group action the associated quotient is a smooth manifold that we term the *Visual Inertial SLAM (VI-SLAM) manifold*

$$\mathcal{M}_n^{\mathrm{VI}}(3) := \mathcal{T}_n^{\mathrm{VI}}(3)/\alpha = \{[P, v, p_i] \,|\, (P, v, p_i) \in \mathcal{T}_n^{\mathrm{VI}}(3)\},$$

with projection map $\pi(P, v, p_i) := [P, v, p_i]$. The induced system and measurements functions on $\mathcal{M}_n^{\mathrm{VI}}(3)$ are well-defined due to their invariance with respect to α. Transformation by the subgroup $\mathbf{SE}_{\mathbf{e}_3}(3)$ corresponds directly the unobservable states in the inertial SLAM coordinates [14]. It follows that the SLAM problem posed on the VI-SLAM manifold is fully observable since the quotient operation factors out the unobservable states while preserving the observable information.

IV. Equivariant Filter for VI-SLAM

A. Symmetry of VI-SLAM

The Equivariant Filter (EqF) [23] exploits symmetries of systems on homogeneous spaces to design a filter about a fixed linearisation point on the state manifold with a constant output linearisation.

Let $\mathbf{SLAM}_n^{\text{VI}}(3) = \mathbf{SE}_2(3) \times \mathbf{SOT}(3)^n$ denote the *VI-SLAM Group* [22] with group product, identity and inverse given by

$$(A^1, w^1, Q_i^1) \cdot (A^2, w^2, Q_i^2) = (A^1A^2, w^1 + R_{A^1}w^2, Q_i^1Q_i^2),$$
$$\text{id} = (I_4, 0_{3\times1}, (I_3)_i), \quad (A, w, Q_i)^{-1} = (A^{-1}, -R_Aw, Q_i^{-1}).$$

This Lie group is a symmetry group that acts transitively on $\mathcal{M}_n^{\text{VI}}(3)$ and $\mathcal{N}_n^{\text{V}}(3)$ in a compatible manner that makes both the system function f (3) and the output function h (4) equivariant. The following lemmas are proved in Appendix I.

Lemma 4.1: The map $\Phi : \mathbf{SLAM}_n^{\text{VI}}(3)\times\mathcal{T}_n^{\text{VI}}(3) \to \mathcal{T}_n^{\text{VI}}(3)$ defined by

$$\begin{aligned}&\Phi((A, w, Q_i), (P, v, p_i))\\ &\quad := (PA, R_A^\top(v - w), PAT_CQ_i^{-1}T_C^{-1}P^{-1}(p_i)),\end{aligned} \tag{5}$$

is a transitive (right) group action. Moreover, the induced action $\phi : \mathbf{SLAM}_n^{\text{VI}}(3) \times \mathcal{M}_n^{\text{VI}}(3) \to \mathcal{M}_n^{\text{VI}}(3)$, given by

$$\phi((A, w, Q_i), [P, v, p_i]) := [\Phi((A, w, Q_i), (P, v, p_i))], \tag{6}$$

is well-defined.

Lemma 4.2: The map $\rho : \mathbf{SLAM}_n^{\text{VI}}(3)\times\mathcal{N}_n^{\text{V}}(3) \to \mathcal{N}_n^{\text{V}}(3)$ defined by

$$\rho((A, w, Q_i), (\eta_i)) := (R_{Q_i}^\top \eta_i), \tag{7}$$

is a (right) group action. Additionally, the measurement function (4) is equivariant with respect to the actions ϕ (6) and ρ, that is,

$$h(\phi((A, w, Q_i), [P, v, p_i])) = \rho((A, w, Q_i), h([P, v, p_i])),$$

for all $(A, w, Q_i) \in \mathbf{SLAM}_n^{\text{VI}}(3)$ and $[P, v, p_i] \in \mathcal{M}_n^{\text{VI}}(3)$.

The existence of a transitive action by the VI-SLAM group on the VI-SLAM manifold guarantees the existence of an equivariant lift [25]. That is, the system dynamics may be lifted to the symmetry group.

Lemma 4.3: The map $\Lambda : \mathcal{T}_n^{\text{VI}}(3) \times (\mathbb{R}^3 \times \mathbb{R}^3) \to \mathfrak{slam}_n^{\text{VI}}(3)$, given by

$$\begin{aligned}&\Lambda((P, v, p_i), (\Omega, a)) \\ &\quad := \left(U(\Omega, v),\ -a + gR_P^\top\mathbf{e}_3,\ \left(\Omega_C + \frac{q_i^\times v_C}{\|q_i\|^2}, \frac{q_i^\top v_C}{\|q_i\|^2}\right)_i\right),\\ &q_i := (PT_C)^{-1}(p_i), \qquad (\Omega_C, v_C) := \text{Ad}_{T_C}^{-1}(\Omega, v),\end{aligned} \tag{8}$$

is a lift [25] of the system function (3). That is,

$$\text{D}_E|_{\text{id}}\phi_{(P,v,p_i)}(E)\Lambda((P, v, p_i), (\Omega, a)) = f_{(\Omega,a)}(P, v, p_i).$$

Moreover, the induced map $\Lambda : \mathcal{M}_n^{\text{VI}}(3) \times (\mathbb{R}^3 \times \mathbb{R}^3) \to \mathfrak{slam}_n^{\text{VI}}(3)$ is well-defined and also a lift for the system on the VI-SLAM manifold.

B. Origin Choice and Local Coordinates

The EqF design procedure requires a choice of origin configuration and local coordinates. Let $\Xi^\circ = (P^\circ, v^\circ, p_i^\circ) \in \mathcal{T}_n^{\text{VI}}(3)$ denote a fixed *origin configuration* and set $\xi^\circ = [\Xi^\circ] = [P^\circ, v^\circ, p_i^\circ] \in \mathcal{M}_n^{\text{VI}}(3)$. The filter state for the EqF is an element of the VI-SLAM group, $\hat{X} \in \mathbf{SLAM}_n^{\text{VI}}(3)$, and the associated state estimate is obtained by applying the group action $\hat{\Xi} = \Phi(\hat{X}, \Xi^\circ)$ (5) [25].

Choose the map $\varepsilon : \mathcal{U}_{\xi^\circ} \subset \mathcal{M}_n^{\text{VI}}(3) \to \mathbb{R}^{5+3n}$ defined by

$$\varepsilon([P, v, p_i]) := \begin{pmatrix} \vartheta_{R_{P^\circ}^\top\mathbf{e}_3}(R_P^\top\mathbf{e}_3) \\ v - v^\circ \\ T_C^{-1}(P^{-1}(p_1) - P^{\circ-1}(p_1^\circ)) \\ \vdots \\ T_C^{-1}(P^{-1}(p_n) - P^{\circ-1}(p_n^\circ)) \end{pmatrix}, \tag{9}$$

to be the coordinate chart for $\mathcal{M}_n^{\text{VI}}(3)$. Let $(y_i^\circ) = h(\xi^\circ)$. Choose the map $\delta : \mathcal{U}_{(y_i^\circ)} \subset \mathcal{N}_n^{\text{V}}(3) \to \mathbb{R}^{2n}$ defined by

$$\delta(y_1, ..., y_n) := \left(\vartheta_{y_1^\circ}(y_1), ..., \vartheta_{y_n^\circ}(y_n)\right), \tag{10}$$

to be the local coordinate chart for $\mathcal{N}_n^{\text{V}}(3)$ where ϑ is the stereographic projection of the sphere (1). Note that $\varepsilon(\xi^\circ) = 0$ and $\delta(y_i^\circ) = 0$.

C. Input Bias

We model real-world IMU measurements as having a constant (or slowly time-varying) bias,

$$\begin{aligned}\Omega_m &= \Omega + b_\Omega, & a_m &= a + b_a,\\ \dot{b}_\Omega &= 0, & \dot{b}_a &= 0,\end{aligned}$$

where $\Omega_m, a_m \in \mathbb{R}^3$ are the measured angular velocity and linear acceleration, respectively, and $b_\Omega, b_a \in \mathbb{R}^3$ are the biases. Let $b = (b_\Omega, b_a) \in \mathbb{R}^6$.

D. EqF with Bias Dynamics

Let $\hat{X} \in \mathbf{SLAM}_n^{\text{VI}}(3)$ be the observer state [25] and let $\hat{b} = (\hat{b}_\Omega, \hat{b}_a) \in \mathbb{R}^6$ be the estimated bias with dynamics

$$\begin{aligned}\dot{\hat{X}} &= \hat{X}\Lambda(\phi(\hat{X}, \xi^\circ), (\hat{\Omega}, \hat{a})) - \Delta\hat{X}, & \hat{X}(0) &= \text{id},\\ \dot{\hat{b}} &= -\beta, & \hat{b}(0) &= 0,\end{aligned}$$

where $\hat{\Omega} = \Omega_m - \hat{b}_\Omega$, $\hat{a} = a_m - \hat{b}_a$ and $\Delta \in \mathfrak{slam}_n^{\text{VI}}(3)$ and $\beta \in \mathbb{R}^6$ are correction terms.

Let A_t°, B_t, C° be the EqF state, input, and output matrices, respectively, as described in Appendix II. Let $\Sigma \in \mathbb{S}_+(11 + 3n)$ be the Riccati term of the EqF with bias, with dynamics

$$\begin{aligned}\dot{\Sigma} = &\begin{pmatrix} 0 & 0 \\ -B_t & A_t \end{pmatrix}\Sigma + \Sigma\begin{pmatrix} 0 & -B_t^\top \\ 0 & A_t^\top \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & B_tR_tB_t^\top \end{pmatrix}\\ &+ P_t - \Sigma\begin{pmatrix} 0 & 0 \\ 0 & C^{\circ\top}Q_t^{-1}C^\circ \end{pmatrix}\Sigma, \qquad \Sigma(0) = \Sigma_0\end{aligned}$$

where $\Sigma_0 \in \mathbb{S}_+(11 + 3n)$ is the initial Riccati term, and $P_t \in \mathbb{S}_+(11 + 3n)$, $R_t \in \mathbb{S}_+(6)$, and $Q_t \in \mathbb{S}_+(2n)$ are positive definite gain matrices.

Let $\xi = [P, v, p_i] \in \mathcal{M}_n^{\text{VI}}(3)$ denote the true state of the system and let $e = \phi(\hat{X}^{-1}, \xi) \in \mathcal{M}_n^{\text{VI}}(3)$ denote the global EqF error. The correction term of the EqF with bias is determined by the lift of the Kalman update to the Lie-group. That is

$$\Delta := \mathrm{D}_E|_{\mathrm{id}} \phi_{\xi^\circ}(E)^\dagger \cdot \mathrm{D}_\xi|_{\xi^\circ} \varepsilon(\xi)^{-1} \Gamma, \tag{11}$$

$$\begin{pmatrix} \beta \\ \Gamma \end{pmatrix} := \Sigma \begin{pmatrix} 0 & C^{\circ\top} \end{pmatrix} Q_t^{-1} \delta(\rho(\hat{X}^{-1}, h(\xi))), \tag{12}$$

where $\mathrm{D}_E|_{\mathrm{id}} \phi_{\xi^\circ}(E)^\dagger$ is a suitably chosen right-inverse of $\mathrm{D}_E|_{\mathrm{id}} \phi_{\xi^\circ}(E)$. Then the EqF state estimate is given by

$$(\hat{P}, \hat{v}, \hat{p}_i) := \Phi((\hat{A}, \hat{w}, \hat{Q}_i), (P^\circ, v^\circ, p_i^\circ)). \tag{13}$$

E. Bundle Lift

The EqF is designed directly on the VI-SLAM manifold to overcome the unobservability of the problem. However, in practice it is usual to report the state as an element of the total space. Moreover, in lifting to the total space it is desirable to do so in such a manner to minimize the error introduced into the trajectory estimation. The correction term is therefore lifted from the manifold to the total space by minimising the motion of landmark points with respect to the current choice of inertial frame, subject to a weighting term derived from the EqF Riccati matrix Σ.

Define the weighted cost function $J : \mathrm{T}_{(\hat{P}, \hat{v}, \hat{p}_i)} \mathcal{T}_n^{\mathrm{VI}}(3) \to \mathbb{R}^+$ by

$$J(\hat{P}U, u_v, u_{p_i}) := \left\| \mathrm{d}\varepsilon \cdot \mathrm{d}\pi \cdot \mathrm{d}\Phi^{-1}_{(\hat{A}, \hat{w}, \hat{Q}_i)}(0, 0, u_{p_i}) \right\|^2_\Sigma ,$$

where Σ is the EqF Riccati term, and $\|\cdot\|_\Sigma$ is the Mahalanobis norm.

Let $\Gamma \in \mathbb{R}^{5+3n}$ be the correction term defined in (11). The correction on the total space $\Gamma' \in \mathrm{T}_{(P^\circ, v^\circ, p_i^\circ)} \mathcal{T}_n^{\mathrm{VI}}(3)$ is the solution of the optimisation problem

$$\begin{aligned} &\text{minimise } J(\mathrm{D}_\Xi|_{(P^\circ, v^\circ, p_i^\circ)} \Phi_{(\hat{A}, \hat{w}, \hat{Q}_i)}(\Xi)(\Gamma')), \\ &\text{subject to } \mathrm{D}_\xi|_{[P^\circ, v^\circ, p_i^\circ]} \varepsilon(\xi) \cdot \mathrm{D}_\Xi|_{(P^\circ, v^\circ, p_i^\circ)} \pi(\Xi) \Gamma' = \Gamma. \end{aligned}$$

This may be solved using weighted linear least-squares.

Finally, the correction term $\Delta \in \mathfrak{slam}_n^{\mathrm{VI}}(3)$ is chosen by

$$\Delta = \mathrm{D}_E|_{\mathrm{id}} \Phi_{(P^\circ, v^\circ, p_i^\circ)}(E)^\dagger \Gamma',$$

where $\mathrm{D}_E|_{\mathrm{id}} \Phi_{(P^\circ, v^\circ, p_i^\circ)}(E)^\dagger$ is an arbitrary fixed right-inverse of $\mathrm{D}_E|_{\mathrm{id}} \Phi_{(P^\circ, v^\circ, p_i^\circ)}(E)$.

V. EXPERIMENTS

To demonstrate practical performance, we evaluated the proposed EqF on a number of sequences from the EuRoC dataset [26]. Vision measurements were obtained by applying OpenCV functions `goodFeaturesToTrack` and `calcOpticalFlowPyrLK` to detect and track features. The maximum number of features was kept to 50, and new features were detected whenever the number of features being tracked fell below 40. The EqF gain matrices and parameters were kept consistent across all trials, and the observer dynamics were discretised using Euler integration. The proposed system was implemented in `c++` and our code is available online[1].

We limited our attention to the easy and medium sequences in the Vicon rooms, as we found the Machine Hall and "hard" sequences to be too challenging for our vision processing front-end to track features reliably. Figure 1 shows the estimated trajectories compared with the ground truth. Table I shows the Root Mean Square Error (RMSE) between ground truth trajectory of the robot and the estimated position reported by our system for each sequence. The RMSE of popular EKF-based VI-SLAM solutions SVO-MSF, MSCKF, and ROVIO (obtained from [2]), and the invariance based R-UKF-LG (obtained from [27]) are shown for comparison. Due to significant differences in system architectures, the tuning parameters for each system cannot be compared directly, and this contributes to the differences in outcomes in Table I.

	RMSE (m)				
	R-UKF-LG[2]	SVO-MSF	MSCKF	ROVIO	EqF
	[27]	[8]	[7]	[6]	*
V1 01	0.55	0.40	0.34	0.10	**0.07**
V1 02	0.40	0.63	0.20	**0.10**	0.11
V2 01	0.37	0.20	0.10	0.12	**0.08**
V2 02	0.47	0.37	0.16	0.14	**0.13**

TABLE I
COMPARISON OF RMSE ON THE EUROC DATASET.

The proposed EqF clearly outperforms competitor EKF-based algorithms shown in Table I. ROVIO [6] achieves the lowest RMSE on `V1_02`, likely thanks to the tight-coupling between the vision front-end and the filter back-end. Figure 2 shows the distribution of the error of the trajectory estimated by our system for each sequence.

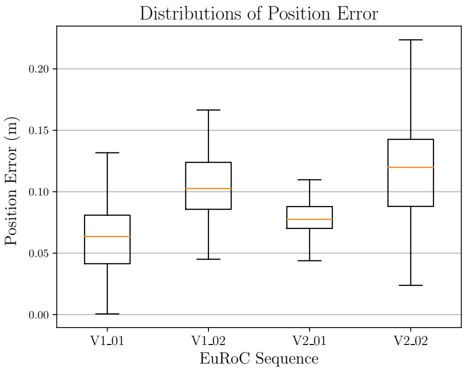

Fig. 2. The distributions of position error of the estimated trajectories.

The pie chart in Figure 3 shows the mean processing time per frame of each component of the full system. The processing times were recorded on a desktop computer with an Intel®Core™i7-8700 CPU @ 3.20GHz × 12 and 16GB of RAM. The total processing time per frame was recorded at 5.4ms or 183.7Hz. Of the total time, the vision front-end consumes 4.2ms and the filter consumes 1.2ms. This high speed combined with the high accuracy reported in Table I

[1] https://github.com/pvangoor/eqf_vio

[2] The R-UKF-LG [27] uses both the left and right cameras for stereo vision rather than monocular vision.

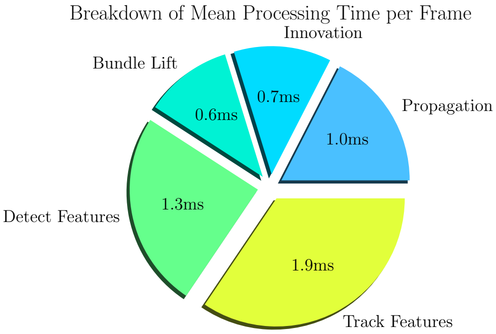

Fig. 3. The mean processing time per frame for each component of the proposed system.

clearly demonstrate the real-world potential of the proposed system.

VI. Conclusion

This paper makes the following contributions.

- The VI-SLAM manifold is constructed as a state space for visual inertial SLAM where the unobservability of the system associated with a change of reference frame is factored out.
- The VI-SLAM group is developed and shown to act equivariantly on the VI-SLAM total space, manifold and output space.
- An Equivariant Filter (EqF) is designed according to [23], and coupled with a simple vision processing front-end to achieve state-of-the-art results on the EuRoC MAV dataset [26].

Appendix I
Proofs

Proof of Proposition 3.1: Choosing $S \in \mathbf{SE}_{\mathbf{e}_3}(3)$, $(P, v, p_i) \in \mathcal{T}_n^{\mathrm{VI}}(3)$ and $(\Omega, a) \in \mathbb{R}^3 \times \mathbb{R}^3$ arbitrary, one has

$$\begin{aligned} f_{(\Omega,a)}(\alpha(S,(P,v,p_i))) &= f_{(\Omega,a)}(S^{-1}P, v, S^{-1}(p_i)), \\ &= \left(S^{-1}PU,\ -\Omega^\times v + a - gR_P^\top R_S \mathbf{e}_3,\ 0\right), \\ &= \left(S^{-1}(PU),\ -\Omega^\times v + a - gR_P^\top \mathbf{e}_3,\ 0\right), \\ &= \mathrm{d}\alpha_S f_{\Omega,a}(P, v, p_i), \end{aligned}$$

as required. Similarly, for any k, the partial measurement function h^k (4) satisfies

$$\begin{aligned} h^k(\alpha(S,(P,v,p_i))) &= h^k(S^{-1}P, v, S^{-1}(p_i)), \\ &= \pi_{\mathrm{S}^2}\left((S^{-1}PT_C)^{-1}S^{-1}(p_k)\right), \\ &= \pi_{\mathrm{S}^2}\left((PT_C)^{-1}(p_k)\right), \\ &= h^k(S,(P,v,p_i)), \end{aligned}$$

and the invariance of the full measurement function h follows immediately. ■

Proof of Lemma 4.1: The proof that Φ is a group action closely follows that of [22, Lemma 4.2], and has been omitted from this paper to save space. To see that ϕ is well-defined, observe that

$$\begin{aligned} &\phi((A, w, Q_i), [S^{-1}P, v, S^{-1}(p_i)]) \\ &= [\Phi((A, w, Q_i), (S^{-1}P, v, S^{-1}(p_i)))], \\ &= [S^{-1}PA, R_A^\top(v - w), S^{-1}PAT_C Q_i^{-1} T_C^{-1} P^{-1} S S^{-1}(p_i)], \\ &= [PA, R_A^\top(v - w), PAT_C Q_i^{-1} T_C^{-1} P^{-1}(p_i)], \\ &= [\Phi((A, w, Q_i), (P, v, p_i))], \\ &= \phi((A, w, Q_i), [P, v, (p_i)]), \end{aligned}$$

as required. ■

The proofs of Lemmas 4.2 and 4.3 closely follow proofs previously published in [22] and have been omitted from this paper to save space.

Appendix II
EqF Matrices

Here we present the state, input, and output matrices of the EqF proposed in Section IV. Let $(P^\circ, v^\circ, p_i^\circ) \in \mathcal{T}_n^{\mathrm{VI}}(3)$ denote the origin coordinates, let $(\hat{A}, \hat{w}, \hat{Q}_i) \in \mathbf{SLAM}_n^{\mathrm{VI}}(3)$ denote the observer state, let $(\hat{P}, \hat{v}, \hat{p}_i)$ denote the estimated state as in (13), and let the input to the system be $(\Omega, a) \in \mathbb{R}^3 \times \mathbb{R}^3$. Define $(\hat{\Omega}_C, \hat{v}_C) := \mathrm{Ad}_{T_C}^{-1}(\Omega, \hat{v})$, $\hat{q}_i := (\hat{P}T_C)^{-1}(\hat{p}_i)$.

The EqF state matrix A_t° is given by [23, Lemma A.1],

$$A_t^\circ = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -g\mathrm{D}_z|_0 \vartheta_{\mathbf{e}_3}^{-1}(z) & 0 & 0 & \cdots & 0 \\ 0 & -\hat{Q}_1 R_{\hat{A}T_C}^\top & A_{\hat{q}_1} & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & -\hat{Q}_n R_{\hat{A}T_C}^\top & 0 & 0 & A_{\hat{q}_n} \end{pmatrix},$$

where,

$$A_{\hat{q}_i} := -\|\hat{q}_i\|^{-2} \hat{Q}_i (\hat{q}_i^\times v_C^\times - 2v_C \hat{q}_i^\top + \hat{q}_i v_C^\top) \hat{Q}_i^{-1}.$$

The EqF input matrix B_t is given by

$$\begin{aligned} B_t &= \mathrm{D}_\xi|_{[P^\circ, v^\circ, p_i^\circ]} \varepsilon(\xi) \cdot \mathrm{D}_\xi|_{[\hat{P}, \hat{v}, \hat{q}_i]} \phi_{(\hat{A}, \hat{w}, \hat{Q}_i)^{-1}}(\xi) \\ &\qquad \cdot \mathrm{D}_E|_{\mathrm{id}} \phi_{[\hat{P}, \hat{v}, \hat{q}_i]}(E) \cdot \mathrm{D}_u|_{(\Omega, a)} \Lambda([\hat{P}, \hat{v}, \hat{q}_i], u), \\ &= \begin{pmatrix} \mathrm{D}_\eta|_{R_{P^\circ}^\top \mathbf{e}_3} \vartheta_{R_{P^\circ}^\top \mathbf{e}_3}(\eta) R_{\hat{A}} (R_{\hat{A}}^\top \mathbf{e}_3)^\times & 0 \\ R_{\hat{A}} \hat{v}^\times & R_{\hat{A}} \\ \hat{Q}_1 (\hat{q}_1^\times R_{T_C}^\top + R_{T_C}^\top x_{T_C}^\times) & 0 \\ \vdots & \vdots \\ \hat{Q}_n (\hat{q}_n^\times R_{T_C}^\top + R_{T_C}^\top x_{T_C}^\times) & 0 \end{pmatrix}. \end{aligned}$$

Let $q_i^\circ = (P^\circ T_C)^{-1}(p_i^\circ)$ and $(y_i^\circ) = h(P^\circ, v^\circ, p_i^\circ)$. Then the (constant) EqF output matrix C° is given by

$$C^\circ = \begin{pmatrix} 0 & 0 & C_1^\circ & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & C_n^\circ \end{pmatrix},$$

where each $C_i^\circ \in \mathbb{R}^{2 \times 3}$ is given by

$$C_i^\circ = \mathrm{D}_\eta|_{y_i^\circ} \vartheta_{y_i^\circ}(\eta) \frac{1}{\|q^\circ\|} \left(I_3 - \frac{q^\circ q^{\circ\top}}{q^{\circ\top} q^\circ} \right).$$

References

[1] H. F. Durrant-Whyte and T. Bailey, "Simultaneous Localisation and Mapping (SLAM): Part I," *IEEE Robotics and Automation Magazine (RAM)*, vol. 13, no. 2, pp. 99–110, 2006.

[2] J. Delmerico and D. Scaramuzza, "A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[3] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM," *arXiv preprint arXiv:2007.11898*, 2020.

[4] 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, Mar. 2015.

[5] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[6] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 298–304.

[7] 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*. IEEE, 2007, pp. 3565–3572.

[8] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect visual odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2016.

[9] J. A. Castellanos, J. Neira, and J. D. Tardós, "Limits to the consistency of EKF-based SLAM," *IFAC Proceedings Volumes*, vol. 37, no. 8, pp. 716–721, 2004.

[10] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 3562–3568.

[11] J. Andrade-Cetto and A. Sanfeliu, "The effects of partial observability in SLAM," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, vol. 1, Apr. 2004, pp. 397–402 Vol.1.

[12] K. W. Lee, W. S. Wijesoma, and J. I. Guzman, "On the observability and observability analysis of SLAM," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 3569–3574.

[13] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *The International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, 2010.

[14] 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, Jan. 2011.

[15] W. M. Wonham, "On a matrix Riccati equation of stochastic control," *SIAM Journal on Control*, vol. 6, no. 4, pp. 681–697, 1968.

[16] 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, 2015.

[17] ——, "An EKF-SLAM algorithm with consistency properties," *arXiv:1510.06263 [cs]*, Sep. 2016.

[18] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, "An invariant-EKF VINS algorithm for improving consistency," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1578–1585.

[19] 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.

[20] S. Heo and C. G. Park, "Consistent EKF-based visual-inertial odometry on matrix Lie group," *IEEE Sensors Journal*, vol. 18, no. 9, pp. 3780–3788, 2018.

[21] R. Mahony and T. Hamel, "A geometric nonlinear observer for simultaneous localisation and mapping," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, Dec. 2017, pp. 2408–2415.

[22] P. van Goor, R. Mahony, T. Hamel, and J. Trumpf, "A Geometric Observer Design for Visual Localisation and Mapping," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, Dec. 2019, pp. 2543–2549.

[23] P. van Goor, T. Hamel, and R. Mahony, "Equivariant Filter (EqF)," *arXiv:2010.14666 [cs, eess]*, Oct. 2020.

[24] J. M. Lee, "Smooth Manifolds," in *Introduction to Smooth Manifolds*, ser. Graduate Texts in Mathematics, J. M. Lee, Ed. New York, NY: Springer, 2012.

[25] R. Mahony, T. Hamel, and J. Trumpf, "Equivariant Systems Theory and Observer Design," *arXiv:2006.08276 [cs, eess]*, Aug. 2020.

[26] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, Sep. 2016.

[27] M. Brossard, S. Bonnabel, and A. Barrau, "Invariant Kalman Filtering for Visual Inertial SLAM," in *2018 21st International Conference on Information Fusion (FUSION)*, Jul. 2018, pp. 2021–2028.