Decentralized Collaborative State Estimation for Aided Inertial Navigation

Roland Jung[1], Christian Brommer[2] and Stephan Weiss[2]

***Abstract*— In this paper, we present a Quaternion-based Error-State Extended Kalman Filter (Q-ESEKF) based on IMU propagation with an extension for Collaborative State Estimation (CSE) and a communication complexity of $\mathcal{O}(1)$ (in terms of required communication links). Our approach combines a versatile filter formulation with the concept of CSE, allowing independent state estimation on each of the agents and at the same time leveraging and statistically maintaining interdependencies between agents, after joint measurements and communication (i.e. relative position measurements) occur. We discuss the development of the overall framework and the probabilistic (re-)initialization of the agent's states upon initial or recurring joint observations. Our approach is evaluated in a simulation framework on two prominent benchmark datasets in 3D.**

Video – `https://youtu.be/igvWLbwnk7M`

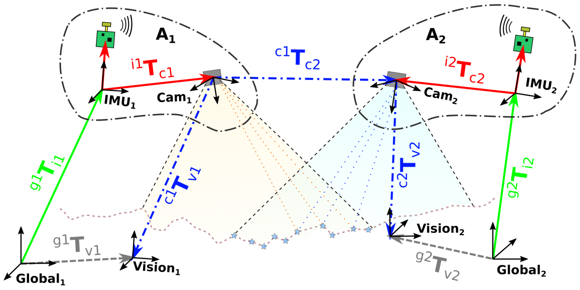

Fig. 1: Visualization of the different reference frames involved in a relative pose measurement between two Agents A_1 and A_2. Each frame is represented as a pose. Blue are measurements. Gray ones are short term stable calibrations states.

I. INTRODUCTION

In this work we present an approach for decentralized *Collaborative State Estimation* (CSE) for aided *Inertial Navigation Systems* (INS). In our case, the INS is based on the widely used *Quaternion-based feedback Error-State (indirect) EKF* (Q-ESEKF) formulation, which can be used in combination with different exteroceptive sensors, e.g. GNSS module, camera, lidar, sonar, pressure, magnetometer, *ultra-wideband* (UWB) range measurements, etc., to correct the estimated states. Hausman et al. demonstrated in [1] the versatility, modularity, the gained robustness, and self-calibration capabilities of a modular multi-sensor state-estimator. Brommer et al. demonstrated in [2] the benefit of using multiple sensors to repetitively recharging a drone for long-term autonomy. In this work, we go a step further and extend the idea of a multi-sensor fusion framework on a single agent to a multi-sensor collaborative sensor fusion framework incorporating not only different sensors, but also agents that essentially acting as opportunistic virtual sensors.

The basic idea of CSE is as follows: Gather multiple a priori estimates of multivariate random variables and apply joint observation(s) to correct the estimates. The challenges in performing this fusion is discussed in the field of *Distributed Collective Localization* (DCL) and extends the localization question: *"Where am I?"* in a group of collaboratively localizing entities to the *two* questions *"Where are we?"* and *"What do we know from each other?"*. The challenges remain in the maintenance of interdependencies between distributed states, as those need to be considered, otherwise the result may be over-optimistic [3]. In a centralized architecture, all required information to perform an optimal information fusion is available.

In [3], Roumeliotis and Bekey show that state propagation of an EKF formulation can be performed locally as long as there are no interdependencies between the process noises. The state correction is performed in a central fusion entity, requiring a full and permanent connection between the server and all robots/clients for each individual or joint observation. To avoid the full connection requirement, Luft et al. proposed recently in [4] approximations, allowing individual observations being processed locally and joint observations requiring a bidirectional communication among participants. The joint observations are exact for participating estimators, but not for non-participating estimators with interdependencies (i.e. previously seen ones). The authors demonstrated their concept using the UTIAS dataset [5] for wheeled robots. Our work in this paper is inspired by this approach due to the proven good performance and reduced communication, despite the approximations made. We extend the approach to 3D space for aerial vehicle and demonstrate with real data that our solution is suitable for (global) localization of aerial robots in their 6-DoF pose and 3D velocity. More precisely our contributions are:

- Development of a Q-ESEKF with Collaborative State

[1]Karl Popper Kolleg on Networked Autonomous Aerial Vehicles, University of Klagenfurt, Austria. Email: `roland.jung@aau.at`

[2]Control of Networked Systems Group, University of Klagenfurt, Austria. Email: `christian.brommer@ieee.org`, `stephan.weiss@ieee.org`

The research leading to these results has received funding from the University of Klagenfurt within the doctoral school KPK-NAV, from the ARL within the BAA W911NF-12-R-0011 under grant agreement W911NF-16-2-0112, and from the European Union's Horizon 2020 research and innovation program under grand agreement No 871260

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/ICRA40945.2020.9197178

Estimation (CSE) extension for multiple agents in 3D space
- Development of a initialization strategy for state uncertainties using first order pose composition and inversion on the Lie algebra upon relative measurement reception
- Evaluation of the decentralized formulation with real-world datasets in 3D.

II. Related Work

The challenge in DCL is a decoupled, decentralized architecture that performs equivalently to a joint centralized solution (denoted as *exact* solution). To reduce communication, different strategies were shown using either complex bookkeeping algorithms, approximations or optimizations techniques like *Covariance Intersection* (CI). The advantage of CI (introduced by Julier and Uhlmann in [6]) is that interdependencies can be neglected, as the joint covariance is computed by a convex combination of the existing uncertainties, resulting in provable consistent but overly pessimistic/conservative covariances. This approach cannot be used to directly recover unknown cross-correlation between agents, but to fuse estimates and measurements with unknown interdependencies. To be exact, a decentralized fusion typically comes at the cost of additional communication and/or of bookkeeping (see [3] [7] [8] [9]). Comprehensive performance and theoretical upper bound analysis for the covariances of CL can be found in [10] and [11]. Avoiding the maintenance of interdependencies by performing CI was presented in [12] [13] and split-CI solutions in [14] [15]. Centralized *Collaborative Localization* (CL), based on imaging sensors and overlapping field of views was presented in [16] [17]. In [16], Melnyk et al. present a centralized collaborative MSC-KF (a well known VIO approach), holding current states of two agents and a history of both past camera poses in a centralized estimation architecture. Overlapping camera views makes their relative pose, under some mild conditions, observable and therefore improves the localization accuracy. Due to the strong coupling and the centralized architecture, this approach is not scalable. In [18], Karrer et al. present an approach to recover a 6-DoF relative pose between two UAVs based on overlapping field of views of monocular cameras and the local odometry information provided by an IMU. This 6-DoF relative pose can then be used as joint observation in a CSE approach. In [19], Karrer et al. presented a centralized collaborative *Simultaneous Localization And Mapping* (SLAM) approach that allows agents, similar to our approach, to navigate w.r.t a common reference/map, based on overlap detection/loop detection, map merging and bundle adjustment. The merging and bundle adjustment is a computational expensive task with strong data dependencies and coupling. In contrast to that, the presented approach is decentralized and allows a seamless navigation in a common reference frame due to a lightweight global frame elimination.

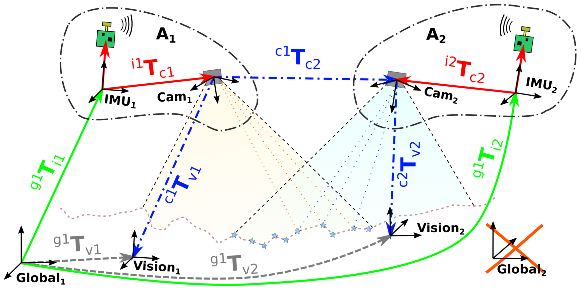

Fig. 2: The first relative pose measurement is used to eliminate the global frame of the agent with the higher reference ID, and would allow all agents to refer to a single reference frame.

III. Problem Formulation

Consider a team of N communicating agents, equipped with at least an inertia measurement unit (IMU) and optionally different exteroceptive sensors, navigating in 3D. Due to the IMU, each robot can estimate its orientation and position with respect to an initial coordinate reference frame. Different exteroceptive sensors can be used to correct each agent's locally drifting estimate. In our case, we assume robots are equipped with either no exteroceptive sensing capabilities, a sensor providing absolute position, a camera (monocular RGB-D or stereo) onboard, or both of them. The camera setup on each agent can be used to provide a metrically scaled six *Degrees of Freedom* (DoF) ego motion-estimation (e.g. by RGB-D *Visual Odometry* (VO) or a feature-based stereo VO/SLAM) with respect to the vision reference frame $\mathcal{V}$. If an overlap between two camera views (each on a different agent) is detected, one agent acts as an interim master, requests the information from the other agent, performs an joint update locally, and sends the correction back. The overlap detection and relative pose computation can be done from feature correspondences in two views. We consider both, the VO and the relative pose computation based on overlapping field of views in metric scale provided. Further, the communication between agents can be established e.g. over WiFi and due to the fact that communication is only required when camera views are close and/or overlapping, the link quality is inherently high. Generally, we make the following assumptions:

- system clocks, IMUs and cameras are synchronized, e.g. by network based synchronization protocol or decentralized methods [20],
- extrinsic calibration between the camera and IMU is known,
- period of exteroceptive sensors is an integer multiple of the IMU period, and
- exteroceptive sensor measurements arrive without delay.

In the beginning, agents are uncorrelated and each of them is navigating with respect to its own reference frame $\mathcal{G}_i$. If an agent receives an absolute position measurement, it can reposition its pose with respect to an absolute reference frame $\mathcal{G}$. If an agent only has vision-based pose updates related to its local vision frame $\mathcal{V}$, the absolute pose will be still unobservable. If two agents have overlapping camera views and a relative pose was computed for the first time, the reference frame of the agent without global information (or with the higher reference ID, in case no agent has global sensors) will be eliminated and its vision frame will refer to the reference frame of the other agent as shown in Figure 2. As a result, if agents are traversing VO-based in GPS-denied environments, having sporadically overlapping camera views, it allows them to navigate with respect to a common reference frame, if all of them converged the lowest reference ID.

A. Notation

Vectors are lower case bold, matrices capitalized bold. The mean and covariance of multivariate random variable is defined as $\mathbf{X}_i \sim \mathcal{N}(\hat{\mathbf{x}}_i, \Sigma_{ii})$. The relation between coordinate reference frames are represented as ${}^{\mathcal{A}}\mathbf{T}_{\mathcal{B}} \in SE(3) := \left\{ \begin{bmatrix} {}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}} & {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p}_{\mathcal{B}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \middle| \mathbf{R} \in SO(3), \mathbf{p} \in \mathbb{R}^3 \right\}$. The composition of homogeneous transformation is ${}^{\mathcal{A}}\mathbf{T}_{\mathcal{C}} = {}^{\mathcal{A}}\mathbf{T}_{\mathcal{B}}{}^{\mathcal{B}}\mathbf{T}_{\mathcal{C}}$ and the transformation of a coordinate vector ${}^{\mathcal{C}}_{\mathcal{C}}\mathbf{p}_{P_1}$ pointing from the origin of the reference frame $\mathcal{C}$ to a point P_1, expressed in $\mathcal{C}$, can be transformed into the frame $\mathcal{A}$ by $\begin{bmatrix} {}^{\mathcal{A}}_{\mathcal{A}}\mathbf{p}_{P_1} \\ 1 \end{bmatrix} = {}^{\mathcal{A}}\mathbf{T}_{\mathcal{C}} \begin{bmatrix} {}^{\mathcal{C}}_{\mathcal{C}}\mathbf{p}_{P_1} \\ 1 \end{bmatrix}$ (read as ${}^{from}_{in}\mathbf{x}\ _{to}$). To differentiate between different agents, we use right subscript to specify the identifier $\{\mathsf{A}_i, i \in 1, \dots, N\}$, where N is the total number of robots. To specify the time indices of state variables, we use the right superscript, e.g. $\mathbf{X}^k$, denoting the state at the time $t(k)$. Names of reference frames are capitalized and italic, e.g. $\mathcal{I}$ for IMU.

IV. Collaborative State Estimation

An EKF framework performs either a state prediction or an individual/private update steps. For CSE, it can be distinguished between individual and joint updates, and cross-correlations between agents need to be maintained by propagating and updating them. Each agent requires an unique identifier and, as explained later, a reference identifier. To maintain the interdependencies between agents, and inspired by [21], we propose storing the factorized cross-covariances in a container that accesses elements by an unique key (known as dictionary). It is denoted as $\sigma_{\mathrm{dict},i}$, with σ emphasizing that it holds factorized cross-covariances between agent A_i to its known agents. A cross-covarinace can be factorized (e.g. by a Cholesky decomposition) into $\Sigma^k_{ij} = \sigma^k_{ij}(\sigma^k_{ji})^{\mathsf{T}}$. Such container can grow dynamically, can inherently give information about known keys, and simplifies the container maintenance during the propagation and update steps. To propagate the sigmas by the transition matrix $\mathbf{F}$ one can iterate over the keys of the dictionary $\sigma^{k(-)}_{\mathrm{dict},i}(key) = \mathbf{F}\sigma^{k-1}_{\mathrm{dict},i}(key)$ and update them, in case of a private observation, by $\sigma^{k(+)}_{\mathrm{dict},i}(key) = \mathbf{U}\sigma^{k(-)}_{\mathrm{dict},i}(key)$ with $\mathbf{U} = (\mathbf{I} - \mathbf{K}\mathbf{H}_i)$. In case of joint observations, the cross-correlations of non-participant agents are updated by $\sigma^{k(+)}_{\mathrm{dict},i}(key) = \mathbf{P}^{k(+)}_{ii}(\mathbf{P}^{k(-)}_{ii})^{-1}\sigma^{k(-)}_{\mathrm{dict},i}(key)$. An element is accessed by $\sigma^k_{ij} = \sigma^k_{\mathrm{dict},i}(id_j)$ and setting one by $\sigma^k_{\mathrm{dict},i}(id_j) = \sigma^k_{ij}$. Due to space limitation, we would like to refer the interested readers for further details to the algorithms $1-3$ provided by Luft et al. in [21].

V. Quaternion-based Error-State EKF

The state space representation and the system propagation model is inspired by [22]. We extended it by an additional calibration state between the global $\mathcal{G}$ and vision reference frame $\mathcal{V}$, yielding a 30-element state vector:

$$\mathbf{X} = [{}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{I}}, {}^{\mathcal{G}}_{\mathcal{G}}\mathbf{v}_{\mathcal{I}}, {}^{\mathcal{G}}\mathbf{q}_{\mathcal{I}}, {}_{\mathcal{I}}\mathbf{b}_{\omega}, {}_{\mathcal{I}}\mathbf{b}_{a}, {}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{C}}, {}^{\mathcal{I}}\mathbf{q}_{\mathcal{C}}, {}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{V}}, {}^{\mathcal{G}}\mathbf{q}_{\mathcal{V}}], \quad (1)$$

with ${}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{I}}, {}^{\mathcal{G}}_{\mathcal{G}}\mathbf{v}_{\mathcal{I}}$, and ${}^{\mathcal{G}}\mathbf{q}_{\mathcal{I}}$ as the position, velocity and orientation of the IMU $\mathcal{I}$ w.r.t. the global frame $\mathcal{G}$. ${}_{\mathcal{I}}\mathbf{b}_{\omega}$ and ${}_{\mathcal{I}}\mathbf{b}_{a}$ are the estimated gyroscope and accelerometer biases. ${}^{\mathcal{I}}_{\mathcal{I}}\mathbf{p}_{\mathcal{C}}$ and ${}^{\mathcal{I}}\mathbf{q}_{\mathcal{C}}$ are the calibration states between the IMU $\mathcal{I}$ and the camera $\mathcal{C}$, and ${}^{\mathcal{G}}_{\mathcal{G}}\mathbf{p}_{\mathcal{V}}$ and ${}^{\mathcal{G}}\mathbf{q}_{\mathcal{V}}$ are the calibration states between the global $\mathcal{G}$ and the vision reference frame $\mathcal{V}$. The corresponding error-state vector using the small angle error approximation for rotations is

$$\widetilde{\mathbf{X}} = [{}^{\mathcal{G}}_{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}}, {}^{\mathcal{G}}_{\mathcal{G}}\tilde{\mathbf{v}}_{\mathcal{I}}, {}^{\mathcal{G}}_{\mathcal{I}}\boldsymbol{\theta}_{\mathcal{I}}, {}_{\mathcal{I}}\tilde{\mathbf{b}_{\omega}}, {}_{\mathcal{I}}\tilde{\mathbf{b}_{a}}, {}^{\mathcal{I}}_{\mathcal{I}}\tilde{\mathbf{p}}_{\mathcal{C}}, {}^{\mathcal{I}}_{\mathcal{C}}\boldsymbol{\theta}_{\mathcal{C}}, {}^{\mathcal{G}}_{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{V}}, {}^{\mathcal{G}}_{\mathcal{V}}\boldsymbol{\theta}_{\mathcal{V}}]. \quad (2)$$

Note, that the calibration states between the IMU and the camera can be omitted and set to a calibrated constant.

Due to our indirect formulation of the EKF, the filter uses the measurement error $\widetilde{\mathbf{z}}$ which is modeled as $\widetilde{\mathbf{z}} = \mathbf{z} - \hat{\mathbf{z}}$, which can be linearized to $\widetilde{\mathbf{z}} = \mathbf{H}\widetilde{\mathbf{x}}$, where $\mathbf{H} = \frac{\partial \tilde{h}}{\partial \widetilde{\mathbf{x}}}|_{\hat{\mathbf{x}}}$ is the Jacobian of the measurement function h with respect to the error state $\widetilde{\mathbf{x}}$ and evaluated at the nominal-states $\hat{\mathbf{x}}$. As the error is additive $\mathbf{p} = \hat{\mathbf{p}} + \tilde{\mathbf{p}}$, except of the rotational, which is right multiplicative $\mathbf{R} = \hat{\mathbf{R}}(\mathbf{I}_3 + [\boldsymbol{\theta}]_{\times})$, the residual $\mathbf{r}$ and measurement error $\widetilde{\mathbf{z}}$ needs to be computed respectively. If a pose is measured, e.g. $\mathbf{z} = \begin{bmatrix} \mathbf{z}_R & \mathbf{z}_p \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}$, the residual is $\mathbf{r} = \begin{bmatrix} \mathbf{z}_p - \hat{\mathbf{p}} \\ (\hat{\mathbf{R}}^{\mathsf{T}}\mathbf{z}_R)^{\vee} \end{bmatrix}$, with $^{\vee}$ as the commonly called *vee* operator.

A. Camera Pose update

For a metrically scaled camera pose measurement ${}^{\mathcal{C}}\mathbf{T}_{\mathcal{V}}$ we derive the following measurement model:

$$ {}^{\mathcal{C}}\mathbf{T}_{\mathcal{V}} = {}^{\mathcal{I}}\mathbf{T}_{\mathcal{C}}^{-1}\ {}^{\mathcal{G}}\mathbf{T}_{\mathcal{I}}^{-1}\ {}^{\mathcal{G}}\mathbf{T}_{\mathcal{V}} = h_{cv}(\mathbf{X}) + \mathbf{n} \quad (3)$$

with $\mathbf{n} \sim \mathcal{N}(0, \mathbf{R})$. The measurement matrix $\mathbf{H}_{cv} = \frac{\partial \tilde{h_c v}}{\partial \widetilde{\mathbf{x}}}|_{\hat{\mathbf{x}}}$ and is presented in Section VII-A.

B. Relative Pose update

For a metrically scaled pose ${}^{\mathcal{C}_1}\mathbf{T}_{\mathcal{C}_2}$ between two agent's cameras $\mathcal{C}_1$ and $\mathcal{C}_2$, we derive the following measurement model:

$$ {}^{\mathcal{C}_1}\mathbf{T}_{\mathcal{C}_2} = {}^{\mathcal{I}_1}\mathbf{T}_{\mathcal{C}_1}^{-1}\ {}^{\mathcal{G}}\mathbf{T}_{\mathcal{I}_1}^{-1}\ {}^{\mathcal{G}}\mathbf{T}_{\mathcal{I}_2}{}^{\mathcal{I}_2}\mathbf{T}_{\mathcal{C}_2} = h_{cc}(\mathbf{X}_1, \mathbf{X}_2) + \mathbf{n} \quad (4)$$

with $\mathbf{n} \sim \mathcal{N}(0, \mathbf{R})$. The measurement matrix $\mathbf{H}_{cc} = \frac{\partial \tilde{h_{cc}}}{\partial \widetilde{\mathbf{x}}}|_{\hat{\mathbf{x}}}$ and is shown in Section VII-B.

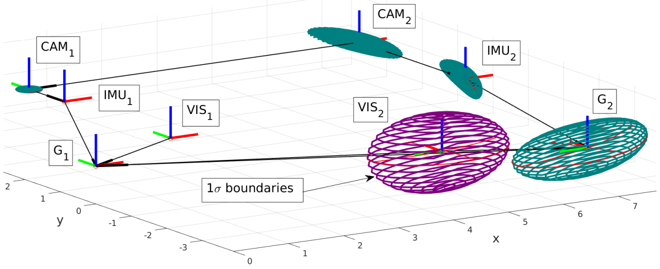

Fig. 3: Banana-shaped covariance ellipsoids showing the 1σ bounds of the propagated compounded poses. The final pose and uncertainty by computing the chain ${}^{\mathcal{G}_1}\mathcal{X}_{\mathcal{V}_2}$ is held in purple.

C. Elimination of global reference

In Figure 2 the initialization strategy is presented and consists of two steps, the absolute pose update and the estimation of the uncertainties for the vision reference frame with respect to the other agent's global reference. We define the estimated poses and calibration states with their uncertainties as Gaussian variables ${}^{\mathcal{A}}\mathcal{X}_{\mathcal{B}} \sim \mathcal{N}({}^{\mathcal{A}}\mathbf{T}_{\mathcal{B}}, {}^{\mathcal{A}}\mathbf{\Sigma}_{\mathcal{B}})$ in $SE(3)$. By combining these Gaussian variables, we are able to compute the pose and, more important, an exact uncertainty. In Figure 3 the resulting covariance of the computed chain between ${}^{\mathcal{G}_1}\mathcal{X}_{\mathcal{V}_2}$ is visualized as deformed ellipsoid (the deformation results from uncertainties in the orientation). This approach allows to set the exact mean and covariance of calibration states. In Algorithm 1 we present our proposed initialization strategy using the composition operators $\oplus$ and $\ominus$ defined in [23] and [24]. The function set_gv() sets the mean and covarinace of the corresponding states.

```
Algorithm 1: init_gv
input : $\mathbf{X}^k_{\{i,j\}}, refID_{\{i,j\}}, {}^{\mathcal{C}_i}\mathbf{T}_{\mathcal{C}_j}, {}^{\mathcal{C}_i}\mathbf{\Sigma}_{\mathcal{C}_j}$
output: $\mathbf{X}^k_{\{i,j\}}, refID_{\{i,j\}}$
if $refID_i < refID_j$ then
    ${}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{V}_j} = {}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{I}_i} \oplus {}^{\mathcal{I}_i}\mathcal{X}_{\mathcal{C}_i} \oplus {}^{\mathcal{C}_i}\mathcal{X}_{\mathcal{C}_j} \ominus {}^{\mathcal{I}_j}\mathcal{X}_{\mathcal{C}_j} \ominus {}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{I}_j} \oplus {}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{V}_j}$ ;
    ${}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{I}_j} = {}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{I}_i} \oplus {}^{\mathcal{I}_i}\mathcal{X}_{\mathcal{C}_i} \oplus {}^{\mathcal{C}_i}\mathcal{X}_{\mathcal{C}_j} \ominus {}^{\mathcal{I}_j}\mathcal{X}_{\mathcal{C}_j}$ ;
    update_abs_pose($\mathbf{X}^k_j, {}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{I}_j}$) ;
    set_gv($\mathbf{X}^k_j, {}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{V}_j}$); $refID_j = refID_i$ ;
else if $refID_i > refID_j$ then
    ${}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{V}_i} = {}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{I}_j} \oplus {}^{\mathcal{I}_j}\mathcal{X}_{\mathcal{C}_j} \ominus {}^{\mathcal{C}_i}\mathcal{X}_{\mathcal{C}_j} \ominus {}^{\mathcal{I}_i}\mathcal{X}_{\mathcal{C}_i} \ominus {}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{I}_i} \oplus {}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{V}_j}$ ;
    ${}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{I}_i} = {}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{I}_j} \oplus {}^{\mathcal{I}_j}\mathcal{X}_{\mathcal{C}_j} \ominus {}^{\mathcal{C}_i}\mathcal{X}_{\mathcal{C}_j} \ominus {}^{\mathcal{I}_i}\mathcal{X}_{\mathcal{C}_i} \ominus {}^{\mathcal{G}_i}\mathcal{X}_{\mathcal{I}_i}$ ;
    update_abs_pose($\mathbf{X}^k_i, {}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{I}_i}$) ;
    set_gv($\mathbf{X}^k_i, {}^{\mathcal{G}_j}\mathcal{X}_{\mathcal{V}_i}$); $refID_i = refID_j$ ;
```

VI. EXPERIMENTAL RESULTS

Our experiments on the estimation performance are performed in a simulation environment using real-world datasets [25] [26] (actually rendering, it is rather an emulation framework than a simulation). The datasets can be loaded between a start and stop timestamp, where the first measurement provided by the ground-truth (GT) sensor is considered to be $t(0)$. The dataset can be shifted via a position offset with respect to it's global reference frame. This offset helps e.g. to increase the distance between close trajectories. The exteroceptive measurements are generated based on the ground-truth trajectory, the sensors calibration states and noise parameters. The ground-truth data of the TUM-VIO dataset suffers from discontinuities, which we mask and a cube spline data interpolation is applied to recover missing data. Based on the mask, we discard measurements, as those are likely to be outliers which may lead to filter inconsistency. The modifier allows to add jitter, latency, and to specify a drop rate. The real-world IMU samples and camera poses provided by the datasets are directly (without modification) used as measurement. Finally, all measurements are sorted chronologically and are locally processed in the multi-instance manager. The multi-instance manager is maintaining a filter instance per simulation data and communication between filter instances is handled locally. We evaluate our approach in two scenarios $\mathsf{S}_{\{1,2\}}$ explained further below. For all our tests, we initialize the state with a proportional offset to the ground-truth value (30 %) and initialize the covariance with realistic values, to demonstrate the self-calibration of states.

In the trajectory plots (Figure 5 and Figure 8), solid lines represent the estimated trajectory of the agents, the dashed ones the corresponding ground-truth trajectory. The orange lines represent a relative observation between agents. Offsets between GT and estimates may occur due to inconsistency or missing updates. As the robots are initialized wrongly, a jump in the estimated pose is visible. The initial values for the biases are causing the filter to drift heavily until they converge and are main source for the estimation error. These are also causing higher RMSE values in the legends of the pose plots, as the RMSE is computed over the entire trajectory. The pose plots (Figure 4 and Figure 6) show the 3σ bounds of the estimated uncertainty as well as the position (x, y, z) and orientation error $(roll, pitch, yaw)$ in red, blue, and orange respectively.

A. Scenario S_1

In CL with agents using an INS only, neither the absolute position nor the orientation about the gravity vector is observable. If a single agents has global position information, e.g. provided by a GNSS, and observing linear acceleration in two axis, all 6-DoF can be recovered [22]. Furthermore, if one recovers its absolute pose, relative position measurements are sufficient for other robots to observe their absolute pose as well [3]. In scenario S_1, we want to evaluate, if an agent A_2 can recover its absolute pose by relative measurements from A_1 with IMU and absolute position measurements. We evaluate this on two Machine Hall (MH) sequences (MH4 for A_1 and MH5 for A_2) of the EuRoC dataset [25]. A_1 is provided by absolute position measurements at a rate of 10 Hz with a standard deviation of $\sigma = 0.1\,\mathrm{m}$. Relative position measurements between A_1 and

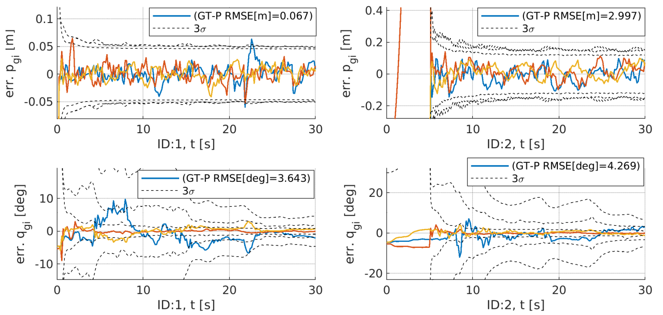

Fig. 4: The error of the estimated position (top row) and attitude (bottom row) in yellow, blue, red for position x, y, z respectively roll, pitch, yaw of agent A_1 (left) and A_2 (right)) in scenario S_1. At $t = 5\,\mathrm{s}$ the relative position measurements between the robots began and A_2 can restore all 6-DoF w.r.t. the absolute position. The estimation errors of both remain between the 3σ boundaries.

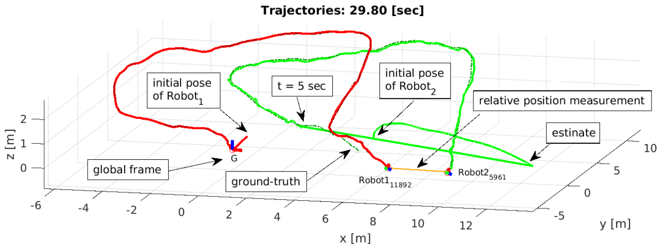

Fig. 5: Estimated trajectories recorded in scenario S_1 of the agents A_1 and A_2 in red and green respectively. A_1 obtains absolute position measurements, whereas A_2 only receives relative position measurements from $t = 5\,\mathrm{s}$ on (explaining the large estimation errors in Figure 4 before the relative measurements).

A_2 are performed from $t = 5\,\mathrm{s}$ on with a standard deviation of $\sigma = 0.1\,\mathrm{m}$ and at a rate of $10\,\mathrm{Hz}$. In Figure 5 the traversed trajectory and in Figure 4 the pose errors are shown. In the first $5\,\mathrm{s}$, the estimated pose of A_2 is heavily drifting, due to wrongly initialized gyroscope and accelerometer biases (${}_{\mathcal{I}}\mathbf{b}_{\omega}$ and ${}_{\mathcal{I}}\mathbf{b}_{a}$). From the first relative position measurement on, the position error of A_2 remains bounded and the uncertainty converged fast. The uncertainties of the orientation converges slower, but the error remains bounded and is not exceeding$10\,°$.

B. Scenario S_2

In this scenario we want to show that our approach works with real overlapping camera views. Therefore we selected three different room sequences of the TUM-VIO dataset [26], namely *room2*, *room3* and *room6*. The movement in these sequences is circular and either clock- or counterclockwise, increasing the chance of overlapping camera views (Figure 8). Relative pose between the cameras was perturbed by a Gaussian noise of $\sigma_p = 0.1\,\mathrm{m}$ and $\sigma_R = 1\,°$. Additionally we compute the overlap between camera views at a depth of $3\,\mathrm{m}$ as a fair assumption for real camera setups and accepted the camera tuple if the overlap is $> 40\,\%$. A_1 is provided by absolute pose measurements at a rate of $10\,\mathrm{Hz}$ with the same noise as the relative pose measurement, whereas A_2 and A_3 are provided by VO pose updates at the same rate, but with a noise of $\sigma_p = 0.3\,\mathrm{m}$ and $\sigma_R = 3\,°$. Furthermore, the initial pose of their vision reference frame was shifted by an arbitrary offset to emphasize their unawareness of their real position. In Figure 7, missing measurement are clearly visible and relative measurements appear intermittently. Especially between A_1 and A_3, as they take place in the beginning and end of the evaluation. Figure 6 shows the pose error of the agents. The missing data in the provided dataset clearly reduce performance on all agents. Based on the 3σ bounds, one can see when relative measurements took place and the absolute error was reduced. We marked the these events with blue and green ellipsoids. Summarized, all agents are navigation with respect to a common global reference frames and the absolute estimation errors remain bounded.

VII. CONCLUSIONS

Driven by the need for truly autonomous and robust navigation for groups of Micro Aerial Vehicles (MAVs), we have presented a versatile Q-ESEKF formulation that supports joint observation considering interdependencies between distributed estimators. Due to the used approximation, communication between estimators is just required while the joint observations is processed. Aiming for 3D reconstruction in GPS-denied environments, we assume that the MAVs are equipped with cameras, a IMU, and optionally with a GNSS module. Overlapping camera views can be used to estimate the relative pose between agents cameras. We presented a method to process these in a filter formulation and introduced a method to change the global reference frames of the estimators. In the field of *Decentralized Collaborative Localization* (DCL), evaluations are mostly performed on ground robots, whereas our evaluations are based on benchmark datasets for *Visual-Inertial Odometry* (VIO) in 3D using real camera overlaps. In the first scenario we shown that the system is observable if one robot has access to absolute position information. In the second, we evaluated our initialization strategy and the relative pose measurement model under difficult and realistic conditions.

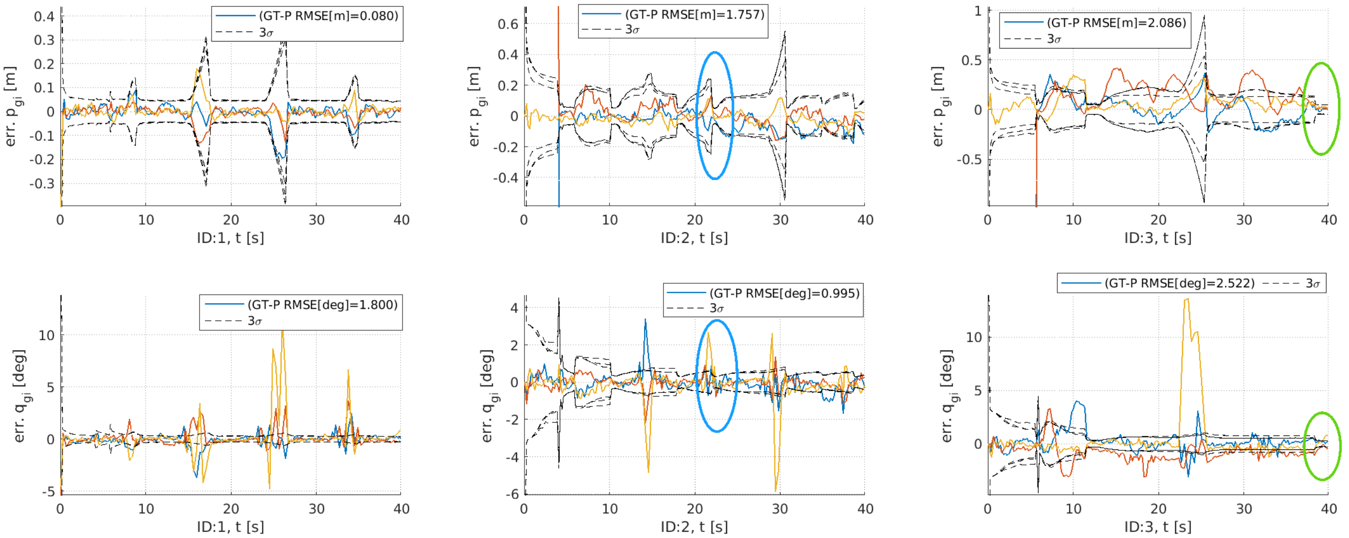

Fig. 6: The error of the estimated position (top row) and attitude (bottom row) in yellow, blue, red for position x, y, z respectively roll, pitch, yaw of agent A_1 (left), A_2 (middle) and A_3 (right) in scenario S_2. Solid lines show the state's mean, dashed ones its covariance. The missing data in the dataset is reflected in the growing error and uncertainty per agent. Nonetheless, relative position measurements for both agents (sample events shown with blue and green ellipses) clearly show the estimation improvement bounding the error through global information propagation from the sensors of A_1.

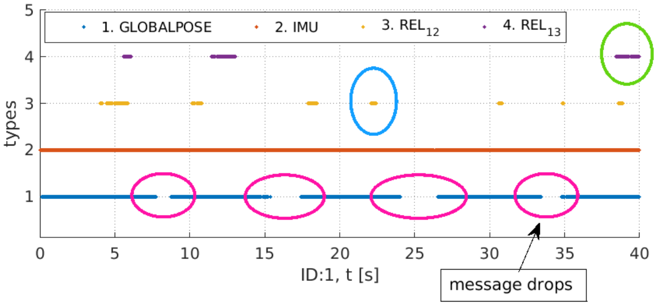

Fig. 7: Measurements performed on agent A_1 in scenario S_2. It clearly shows that relative pose measurements between agents happened intermittently.

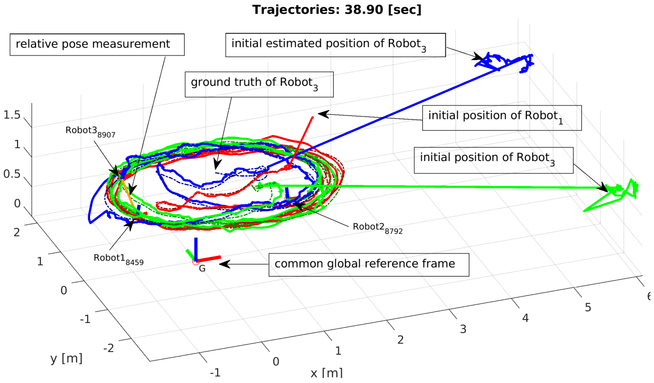

Fig. 8: Shows the trajectories recorded in scenario S_2 of the three agents $\mathsf{A}_{\{1,2,3\}}$ in red, green and blue. The initial positions of A_2 and A_3 were estimated wrongly and could be corrected with the first relative observations. Even though the agents suffered from massive message drops, their estimated pose could converge to ground-truth. Hardly visible is the orange line indicating a relative observation between A_1 and A_2.

APPENDIX

A. Measurement matrix for vision updates

The measurement matrix resulting from Equation (3) is $H_{cv} = \begin{bmatrix}\mathbf{H}_{\tilde{p}} \\ \mathbf{H}_{\tilde{\theta}}\end{bmatrix}$ of

$$\mathbf{H}_{\tilde{p}}({}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}}) = -{}^{\mathcal{I}}\hat{\mathbf{R}}_{\mathcal{C}}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}}^{\mathsf{T}} \tag{5a}$$

$$\mathbf{M} = {}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}}^{\mathsf{T}}(-{}^{\mathcal{G}}\hat{\mathbf{p}}_{\mathcal{I}} + {}^{\mathcal{G}}\hat{\mathbf{p}}_{\mathcal{V}}) \tag{5b}$$

$$\mathbf{H}_{\tilde{p}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}) = {}^{\mathcal{I}}\hat{\mathbf{R}}_{\mathcal{C}}^{\mathsf{T}}\,[\mathbf{M}]_{\times} \tag{5c}$$

$$\mathbf{H}_{\tilde{p}}({}^{\mathcal{I}}\tilde{\mathbf{p}}_{\mathcal{C}}) = -{}^{\mathcal{I}}\hat{\mathbf{R}}_{\mathcal{C}}^{\mathsf{T}} \tag{5d}$$

$$\mathbf{H}_{\tilde{p}}({}^{\mathcal{I}}\tilde{\boldsymbol{\theta}}_{\mathcal{C}}) = \left[{}^{\mathcal{I}}\hat{\mathbf{R}}_{\mathcal{C}}^{\mathsf{T}}(-{}^{\mathcal{I}}\hat{\mathbf{p}}_{\mathcal{C}} + \mathbf{M})\right]_{\times} \tag{5e}$$

$$\mathbf{H}_{\tilde{p}}({}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{V}}) = {}^{\mathcal{I}}\hat{\mathbf{R}}_{\mathcal{C}}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}}^{\mathsf{T}} \tag{5f}$$

$$\mathbf{H}_{\tilde{p}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{V}}) = \mathbf{0}_{3\times 3} \tag{5g}$$

$$\mathbf{H}_{\tilde{\theta}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}) = -{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{V}}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}} \tag{5h}$$

$$\mathbf{H}_{\tilde{\theta}}({}^{\mathcal{I}}\tilde{\boldsymbol{\theta}}_{\mathcal{C}}) = -{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{V}}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}}\,{}^{\mathcal{I}}\hat{\mathbf{R}}_{\mathcal{C}} \tag{5i}$$

$$\mathbf{H}_{\tilde{\theta}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{V}}) = \mathbf{I}_{3\times 3} \tag{5j}$$

with e.g. $\mathbf{H}_{\tilde{p}}({}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}})$ describing the partial derivative of the measured position error $\tilde{\mathbf{z}}_p$ with respect to the position error ${}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}}$. Note, this pose measurement can be regarded as general pose update with respect to a know reference frame, e.g provided by a motion capture system.

B. Measurement matrix for relative pose updates

The measurement matrix resulting from Equation (4) is $H_{cc} = \begin{bmatrix}\mathbf{H}_{1,\tilde{p}} & \mathbf{H}_{2,\tilde{p}} \\ \mathbf{H}_{1,\tilde{\theta}} & \mathbf{H}_{2,\tilde{\theta}}\end{bmatrix}$ with

$$\mathbf{H}_{1,\tilde{p}}({}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}}) = -{}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1}{}^{\mathsf{T}}\,{}^{\mathcal{G}_1}\hat{\mathbf{R}}_{\mathcal{I}_1}{}^{\mathsf{T}} \tag{6a}$$

$$\mathbf{M} = {}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_1}{}^{\mathsf{T}}(-{}^{\mathcal{G}}\hat{\mathbf{p}}_{\mathcal{I}_1} + {}^{\mathcal{G}}\hat{\mathbf{p}}_{\mathcal{I}_2} + {}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_2}\,{}^{\mathcal{I}_2}\hat{\mathbf{p}}_{\mathcal{C}_2}) \tag{6b}$$

$$\mathbf{H}_{1,\tilde{p}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}) = {}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1}{}^{\mathsf{T}}\,[\mathbf{M}]_{\times} \tag{6c}$$

$$\mathbf{H}_{1,\tilde{p}}({}^{\mathcal{I}}\tilde{\mathbf{p}}_{\mathcal{C}}) = -{}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1}{}^{\mathsf{T}} \tag{6d}$$

$$\mathbf{H}_{1,\tilde{p}}({}^{\mathcal{I}}\tilde{\boldsymbol{\theta}}_{\mathcal{C}}) = \left[{}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1}{}^{\mathsf{T}}(-{}^{\mathcal{I}_1}\hat{\mathbf{p}}_{\mathcal{C}_1} + \mathbf{M})\right]_{\times} \tag{6e}$$

$$\mathbf{H}_{2,\tilde{p}}({}^{\mathcal{G}}\tilde{\mathbf{p}}_{\mathcal{I}}) = {}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_1}{}^{\mathsf{T}} \tag{6f}$$

$$\mathbf{H}_{2,\tilde{p}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}) = {}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_1}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_2}\left[-{}^{\mathcal{I}_2}\hat{\mathbf{p}}_{\mathcal{C}_2}\right]_{\times} \tag{6g}$$

$$\mathbf{H}_{2,\tilde{p}}({}^{\mathcal{I}}\tilde{\mathbf{p}}_{\mathcal{C}}) = {}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_1}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_2} \tag{6h}$$

$$\mathbf{H}_{1,\tilde{\theta}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}) = -{}^{\mathcal{I}_2}\hat{\mathbf{R}}_{\mathcal{C}_2}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_2}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_1} \tag{6i}$$

$$\mathbf{H}_{1,\tilde{\theta}}({}^{\mathcal{I}}\tilde{\boldsymbol{\theta}}_{\mathcal{C}}) = -{}^{\mathcal{I}_2}\hat{\mathbf{R}}_{\mathcal{C}_2}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_2}{}^{\mathsf{T}}\,{}^{\mathcal{G}}\hat{\mathbf{R}}_{\mathcal{I}_1}\,{}^{\mathcal{I}_1}\hat{\mathbf{R}}_{\mathcal{C}_1} \tag{6j}$$

$$\mathbf{H}_{2,\tilde{\theta}}({}^{\mathcal{G}}\tilde{\boldsymbol{\theta}}_{\mathcal{I}}) = {}^{\mathcal{I}_2}\hat{\mathbf{R}}_{\mathcal{C}_2}{}^{\mathsf{T}} \tag{6k}$$

$$\mathbf{H}_{2,\tilde{\theta}}({}^{\mathcal{I}}\tilde{\boldsymbol{\theta}}_{\mathcal{C}}) = \mathbf{I}_{3\times 3} \tag{6l}$$

Note, by setting ${}^{\mathcal{I}_{\{1,2\}}}\hat{\mathbf{R}}_{\mathcal{C}_{\{1,2\}}}$ and ${}^{\mathcal{I}_{\{1,2\}}}\hat{\mathbf{p}}_{\mathcal{C}_{\{1,2\}}}$ to the neutral element and neglecting the terms $\mathbf{H}_{\{1,2\},\tilde{p}}({}^{\mathcal{I}}\tilde{\mathbf{p}}_{\mathcal{C}})$ and $\mathbf{H}_{\{1,2\},\tilde{p}}({}^{\mathcal{I}}\tilde{\boldsymbol{\theta}}_{\mathcal{C}})$, this measurement matrix can be used for relative pose measurements between two IMU frames ${}^{\mathcal{I}_1}\mathbf{T}_{\mathcal{I}_2}$. Note, by neglecting the rotational part entirely, it can be regarded as relative position measurement between two cameras ${}^{\mathcal{C}_1}_{\mathcal{C}_1}\mathbf{p}_{\mathcal{C}_2}$ and with the aforementioned modifications between the IMUs ${}^{\mathcal{I}_1}_{\mathcal{I}_1}\mathbf{p}_{\mathcal{I}_2}$.

References

[1] K. Hausman, S. Weiss, R. Brockers, L. Matthies, and G. S. Sukhatme, "Self-calibrating multi-sensor fusion with probabilistic measurement validation for seamless sensor switching on a UAV," in *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2016-June, 2016, pp. 4289–4296.

[2] C. Brommer, D. Malyuta, D. Hentzen, and R. Brockers, "Long-Duration Autonomy for Small Rotorcraft UAS Including Recharging," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018, pp. 7252–7258.

[3] S. I. Roumeliotis and G. A. Bekey, "Distributed multirobot localization," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 5, pp. 781–795, Oct. 2002.

[4] L. Luft, T. Schubert, S. I. Roumeliotis, and W. Burgard, "Recursive Decentralized Collaborative Localization for Sparsely Communicating Robots," in *Robotics: Science and Systems XII*. Robotics: Science and Systems Foundation, 2016. [Online]. Available: http://www.roboticsproceedings.org/rss12/p16.pdf

[5] K. Y. Leung, Y. Halpern, T. D. Barfoot, and H. H. Liu, "The UTIAS multi-robot cooperative localization and mapping dataset," *The International Journal of Robotics Research*, vol. 30, no. 8, pp. 969–974, Jul. 2011. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364911398404

[6] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proceedings of the 1997 American Control Conference (Cat. No.97CH36041)*, vol. 4, Jun. 1997, pp. 2369–2373 vol.4.

[7] K. Leung, T. Barfoot, and H. Liu, "Decentralized Localization of Sparsely-Communicating Robot Networks: A Centralized-Equivalent Approach," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 62–77, Feb. 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5371975/

[8] G. P. Huang, N. Trawny, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based consistent EKF estimators for multi-robot cooperative localization," *Autonomous Robots*, vol. 30, no. 1, pp. 99–122, Jan. 2011. [Online]. Available: https://doi.org/10.1007/s10514-010-9207-y

[9] S. S. Kia, S. F. Rounds, and S. Martinez, "A centralized-equivalent decentralized implementation of Extended Kalman Filters for cooperative localization," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Chicago, IL, USA: IEEE, Sep. 2014, pp. 3761–3766. [Online]. Available: http://ieeexplore.ieee.org/document/6943090/

[10] S. I. Roumeliotis and I. M. Rekleitis, "Analysis of multirobot localization uncertainty propagation," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, vol. 2, Oct. 2003, pp. 1763–1770 vol.2.

[11] A. I. Mourikis and S. I. Roumeliotis, "Performance analysis of multirobot Cooperative localization," *IEEE Transactions on Robotics*, vol. 22, no. 4, pp. 666–681, Aug. 2006.

[12] L. C. Carrillo-Arce, E. D. Nerurkar, J. L. Gordillo, and S. I. Roumeliotis, "Decentralized multi-robot cooperative localization using covariance intersection." IEEE, Nov. 2013, pp. 1412–1417. [Online]. Available: http://ieeexplore.ieee.org/document/6696534/

[13] H. Li and F. Nashashibi, "Cooperative Multi-Vehicle Localization Using Split Covariance Intersection Filter," *IEEE Intelligent Transportation Systems Magazine*, vol. 5, no. 2, pp. 33–44, 2013.

[14] H. Li, F. Nashashibi, and M. YANG, "Split Covariance Intersection Filter: Theory and Its Application to Vehicle Localization," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 14, pp. 1860–1871, Dec. 2013.

[15] T. R. Wanasinghe, G. K. I. Mann, and R. G. Gosine, "Decentralized Cooperative Localization for Heterogeneous Multi-robot System Using Split Covariance Intersection Filter," in *2014 Canadian Conference on Computer and Robot Vision*, May 2014, pp. 167–174.

[16] I. V. Melnyk, J. A. Hesch, and S. I. Roumeliotis, "Cooperative vision-aided inertial navigation using overlapping views," in *2012 IEEE International Conference on Robotics and Automation*, May 2012, pp. 936–943.

[17] N. Piasco, J. Marzat, and M. Sanfourche, "Collaborative localization and formation flying using distributed stereo-vision," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden: IEEE, May 2016, pp. 1202–1207. [Online]. Available: http://ieeexplore.ieee.org/document/7487251/

[18] M. Karrer, M. Agarwal, M. Kamel, R. Siegwart, and M. Chli, "Collaborative 6dof Relative Pose Estimation for Two UAVs with Overlapping Fields of View," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018, pp. 6688–6693.

[19] M. Karrer, P. Schmuck, and M. Chli, "CVI-SLAM—Collaborative Visual-Inertial SLAM," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2762–2769, Oct. 2018.

[20] A. Barciś, M. Barciś, and C. Bettstetter, "Robots that Sync and Swarm: A Proof of Concept in ROS 2," *arXiv:1903.06440 [cs]*, Mar. 2019, arXiv: 1903.06440. [Online]. Available: http://arxiv.org/abs/1903.06440

[21] L. Luft, T. Schubert, S. I. Roumeliotis, and W. Burgard, "Recursive decentralized localization for multi-robot systems with asynchronous pairwise communication," *The International Journal of Robotics Research*, p. 0278364918760698, Mar. 2018. [Online]. Available: https://doi.org/10.1177/0278364918760698

[22] S. Weiss, M. W. Achtelik, M. Chli, and R. Siegwart, "Versatile distributed pose estimation and sensor self-calibration for an autonomous MAV." IEEE, May 2012, pp. 31–38. [Online]. Available: http://ieeexplore.ieee.org/document/6225002/

[23] J. Solà, J. Deray, and D. Atchuthan, "A micro Lie theory for state estimation in robotics," *arXiv:1812.01537 [cs]*, Dec. 2018, arXiv: 1812.01537. [Online]. Available: http://arxiv.org/abs/1812.01537

[24] J. G. Mangelson, M. Ghaffari, R. Vasudevan, and R. M. Eustice, "Characterizing the Uncertainty of Jointly Distributed Poses in the Lie Algebra," *arXiv:1906.07795 [cs]*, Jun. 2019, arXiv: 1906.07795. [Online]. Available: http://arxiv.org/abs/1906.07795

[25] 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. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364915620033

[26] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, “The TUM VI Benchmark for Evaluating Visual-Inertial Odometry,” *arXiv:1804.06120 [cs]*, Apr. 2018, arXiv: 1804.06120. [Online]. Available: http://arxiv.org/abs/1804.06120