2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France

# Vision-based Multi-MAV Localization with Anonymous Relative Measurements Using Coupled Probabilistic Data Association Filter

Ty Nguyen[1*], Kartik Mohta[2*], Camillo J. Taylor[1], Vijay Kumar[1]

***Abstract*— We address the localization of robots in a multi-MAV system where external infrastructure like GPS or motion capture systems may not be available. Our approach lends itself to implementation on platforms with several constraints on size, weight, and power (SWaP). Particularly, our framework fuses the onboard VIO with the anonymous, visual-based robot-to-robot detection to estimate all robot poses in one common frame, addressing three main challenges: 1) the initial configuration of the robot team is unknown, 2) the data association between each vision-based detection and robot targets is unknown, and 3) the vision-based detection yields false negatives, false positives, inaccurate, and provides noisy bearing, distance measurements of other robots. Our approach extends the Coupled Probabilistic Data Association Filter [1] to cope with nonlinear measurements. We demonstrate the superior performance of our approach over a simple VIO-based method in a simulation with the measurement models statistically modeled using the real experimental data. We also show how onboard sensing, estimation, and control can be used for formation flight.**

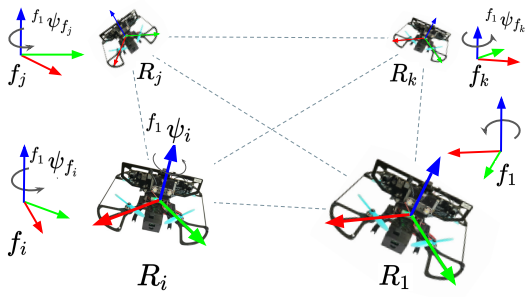



Fig. 1: An example of a homogeneous 4-MAV system featuring our Falcon 250 platform [22]. Robot $i^{th}$ is associated with a body frame $i$ and a fixed frame $f_i$. $f_1$ is chosen to be the common fixed frame. Each robot can detect other robots but not their identity. There are errors in detection and tracking. The robots can communicate with each other.

## I. Introduction

Multi-robot systems are of interest for their potential in performing tasks which may not be feasible or desirable to do with only a single robot in applications such as perimeter surveillance [2], [3], patrolling missions [4], [5], searching operations [6], [7], and formation control [8], [9], [10]. For example, the task of surveilling a large area is often infeasible for one robot due to the robot's limited coverage but can be accomplished by a team of robots under proper coordination. A major requirement for these applications is that the robots need localization information within a common reference frame. That way, each robot can correctly execute its designated subtask and the team can collaboratively complete the full task.

This localization problem becomes trivial when there is a single global coordinate system that can provide the state estimate for all robots, such as GPS, motion capture systems [11], and aerial-image-based localization systems [12]. However, such systems are often not available or reliable. Another solution to this problem is to launch the robots in a predetermined spatial configuration with a common frame [13]. This solution requires significant time and effort to set up and launch the system. Alternative solutions rely on exchanging environment features between robots in the system [14], [15], [16], [17], [18]. Montijano et al. [15] and Leahy et al. [19], for example, propose to use the homography estimation [20], [21] to compute the relative pose between two robots. The main problem with this approach is that the robots either need to maintain a shared feature map or have overlapping views with shared features. The challenge in finding good features in textureless environments is also an obstacle in using this approach.

Our work is related to Franchi et al. [23] in which relative robot-to-robot measurements are used to bootstrap the localization system. Our study, however, focuses on localization for systems of multiple micro Aerial Vehicles (MAVs). These robot systems are highly suitable for applications such as surveillance, search, and rescue operations. MAVs are often subject to size, weight, and power (SWaP) constraints. As a result, cameras are often preferred over range finders and LiDARs to do the localization task due to their compactness and light weight. The system of robots studied in this work, as shown in Fig. 1, is equipped with two types of sensor modalities: the vision-based detection and measurement of other robots within each robot's field of view; and the visual-inertial odometry (VIO) using stereo cameras and inertial measurement unit measurements. These robots are measured $400\,\mathrm{cm}$ from tip to tip and weigh less than $1\mathrm{kg}$. They are also inexpensive, making them a great candidate for research and civilian applications.

We gratefully acknowledge the support of ARL grant DCIST CRA W911NF-17-2-0181, NSF Grants CNS-1446592, and CNS-1521617, ARO grant W911NF-13-1-0350, Qualcomm Research, United Technologies, Intel Corporation, and C-BRIC, a Semiconductor Research Corporation Joint University Microelectronics Program cosponsored by DARPA, and NVIDIA NVAIL

[1] The authors are with the GRASP Lab, University of Pennsylvania, Philadelphia, PA 19104 USA. email: {tynguyen, cjtaylor, kumar}@seas.upenn.edu

[2] The author is with Autel Robotics, USA

[*] Authors contributed equally to the paper

Nevertheless, the localization problem in such multi-MAV systems is challenging due to a couple of reasons. First, robots are homogeneous and visually similar. Thus, the vision-based detection provides no identity information, leading to the data association ambiguity. Secondly, the vision-based detection is usually imperfect, yielding both false positives and false negatives. And unlike range finders used in previous studies such as [23], using only vision-based detection to measure the distance and bearing also leads to noisy measurements. These factors make the data association problem rooted in the localization even more challenging.

In short, we study the problem of multi-MAV localization under the following assumptions,

- Initial relative poses between robots are unknown.
- Robot detection provides no identity information.
- Robot detection can include false negatives and false positives.
- Vision-based distance and bearing measurements are noisy.
- Robot can communicate to exchange the state estimates.

Our main contributions in this work include:

- We introduce a fusion scheme that leverages the onboard VIO and relative distance, bearing measurements for localization of a system of SWaP-constrained robot platforms.
- We propose an extension of Coupled Probabilistic Data Association Filter (CPDAF) with a simple but effective gating and evaluating mechanism to keep the number of hypotheses manageable.
- We demonstrate how onboard sensing, estimation, and control can be used for formation flight.

This work focuses on using VIO for onboard odometry and camera sensors for distance, bearing measurements. However, it is straightforward to apply to a system equipped with other kinds of distance, bearing sensors, and onboard odometry.

## II. Related Work

The localization of robots in a multi-robot system using relative measurements has attracted a large number of research works. For example, Spica et al. [24] address the problem of estimating the formation scale in the context of bearing-based formation localization for multiple robots. In [25], authors propose an Extended Kalman Filter to estimate each follower's position and orientation with respect to the leader, using the bearing information only. These works, however, assume that the data association is known.

The literature has investigated in providing relative measurements along with the robot identities via tagging. For example, in [26], authors put colored circular markers on the robots to obtain relative bearings between them. Dias et al. [27] utilize active markers to identify unique identities of quadrotors based on pulsating at a predefined frequency. The main disadvantage of these methods is that they do not scale well with the number of robots.

State-of-the-art approaches directly deal with unknown robot identities by estimating these identities together with the robot localization. Chang et al. [28], for instance, propose a maximum likelihood data association algorithm with a threshold gating on the Mahalanobis distance between the incoming measurement and the expected measurement. This method's downside is that the selected measurement may not be the correct one due to the noisy measurements, causing the filter divergence. Our approach, based on a probabilistic data association framework, can handle the noisy measurement situation.

The localization problem for multi-robot systems using anonymous relative measurements was first considered by Franchi et al. [29], [30], [23]. In [29], for example, authors introduce a two-phase localization system with a multiple registration algorithm followed by a multi-hypothesis EKF. The registration algorithm establishes data association hypotheses which are then pruned out based on the consistency with the current state belief. Their successive work [30] utilizes the belief in the system's spatial configuration to handle the worst-case computation time. In [23], they improve these frameworks further by replacing the EKF with particle filters. Compared to ours, their later works rely on the assumption that the posterior probability distribution functions of robot states are independent so that each particle filter can be feasibly updated in a separate manner. These frameworks also suffer from adding heavy computation to maintain and update particle filters. This computation is problematic when scaling up the robot system or in case the robot states are high dimensional.

## III. Problem Formulation

Let us consider a problem formulation with a team of homogeneous $N-$robots $\{\mathcal{R}_1, \mathcal{R}_2, \dots \mathcal{R}_N\}$, $N$ is known. Each robot $\mathcal{R}_i$ is attached to a moving body frame $i$, and a fixed frame $f_i$ as shown in Fig. 1 such that the $Z_i$ axis of the frame $f_i$ is on the same direction with the gravity. In the following, we describe the localization problem whose objective is to localize every robot $\mathcal{R}_i$ within a common frame, which can be any in the set $\{f_1, f_2, \dots f_N\}$. Without losing the generalization, we choose the common frame to be $f_1$.

Let ${}^{f_1}\mathbf{p}_{i,t} \subset \mathbb{R}^3$ and ${}^{f_1}\mathbf{R}_{i,t} \subset \mathcal{SO}3$, $i \in \{1, \dots N\}$ denote the translation and rotation of robot $\mathcal{R}_i$ in frame $f_1$ at time $t$, respectively. A set $\bar{\mathcal{G}}_t = \{({}^{f_1}\mathbf{p}_{i,t}, {}^{f_1}\mathbf{R}_{i,t}) \,|i \in \{1, \dots N\}, t \in [0, T]\}$ represents the localization of all robots within frame $f_1$ at time $t$. Our problem becomes,

$$\bar{\mathcal{G}}_{t|t\in[0,T]} = \arg\min_{\bar{\mathcal{G}}_t} \; E_t^t \; = \arg\min_{\bar{\mathcal{G}}_t} \; ||\bar{\mathcal{G}}_t - \mathcal{G}_t||_2^2 \qquad (1)$$

where $\mathcal{G}_t$ is the ground truth. For the simplicity, from now on, we ignore the subscript $t$ in the translation and rotation notations.

## IV. The Stochastic Model

Before representing the proposed approach, we first define discrete models for the system dynamics and observation measurements.

### A. The System State Model

Within the coordinate system associated with the common frame $f_1$, we have,

$$\begin{aligned} {}^{f_1}\mathbf{p}_i &= {}^{f_1}\mathbf{R}_{f_i}\ {}^{f_i}\mathbf{p}_i \ +\ {}^{f_1}\mathbf{p}_{f_i} \quad \forall i \in 1,\dots,N \\ {}^{f_1}\mathbf{R}_i &= {}^{f_1}\mathbf{R}_{f_i}\ {}^{f_i}\mathbf{R}_i \qquad\qquad \forall i \in 1,\dots,N \\ {}^{f_1}\mathbf{p}_j &= {}^{f_1}\mathbf{R}_i\ {}^{i}\mathbf{p}_j \ +\ {}^{f_1}\mathbf{p}_i \qquad \forall i,j \in 1,\dots,N \end{aligned} \tag{2}$$

where ${}^{f_1}\mathbf{p}_{f_i}$ and ${}^{f_1}\mathbf{R}_{f_i}$ represent the translation and rotation of frame $f_i$ within frame $f_1$. Thus, the odometry measurement coming from the VIO system of robot $\mathcal{R}_i$, described in frame $f_i$ can be represented as follows,

$$\mathbf{z}_i = \begin{bmatrix} {}^{f_i}\mathbf{p}_i \\ {}^{f_i}\mathbf{R}_i \end{bmatrix} = \begin{bmatrix} {}^{f_1}\mathbf{R}_{f_i}^T({}^{f_1}\mathbf{p}_i \ -\ {}^{f_1}\mathbf{p}_{f_i}) \\ {}^{f_1}\mathbf{R}_{f_i}^T\ {}^{f_1}\mathbf{R}_i \end{bmatrix} \tag{3}$$

The detection measurement generating from robot $\mathcal{R}_j$ detected by robot $\mathcal{R}_i$, and described in frame $i$, can be presented as,

$${}^{i}\mathbf{z}_j = \begin{bmatrix} {}^{i}\mathbf{p}_j \end{bmatrix} = \begin{bmatrix} {}^{f_1}\mathbf{R}_i^T\ ({}^{f_1}\mathbf{p}_j \ -\ {}^{f_1}\mathbf{p}_i) \end{bmatrix} \tag{4}$$

The first two equations in Eq. 2 show that to achieve the rotation and translation of robot $\mathcal{R}_i$ in frame $f_1$, given only local measurements, we need to know the rotation and translation of frame $f_i$ in $f_1$. Thus, we define the state of robot $\mathcal{R}_i$ as,

$$\mathbf{x}_i = \begin{bmatrix} {}^{f_i}\mathbf{p}_i^T\ {}^{f_1}\mathbf{R}_i\ {}^{f_1}\mathbf{p}_{f_i}^T\ {}^{f_1}\mathbf{R}_{f_i} \end{bmatrix}^T$$

Note that, we substitute ${}^{f_1}\mathbf{p}_i$ by ${}^{f_i}\mathbf{p}_i$ to make it convenient to define the state equation and that these two variables can be derived from each other. The coupled state system can be defined as,

$$\mathbf{x} = [\mathbf{x}_1^T\ \mathbf{x}_2^T\ \dots\ \mathbf{x}_i^T \dots\ \mathbf{x}_{N-1}^T \mathbf{x}_N^T]$$

We can decompose the rotation ${}^{f_1}\mathbf{R}_i$ into two parts, $\mathbf{R}_z({}^{f_1}\psi_i)$ corresponds to the rotation around the gravity vector, and ${}^{f_1}\mathbf{R}_{i,xy}$ corresponds to the rotation on the plane perpendicular to the gravity vector,

$${}^{f_1}\mathbf{R}_i = \ \mathbf{R}_z({}^{f_1}\psi_i){}^{f_1}\mathbf{R}_{i,xy}$$

In a VIO system, only rotation along $Z$ axis is unobservable. Thus, we can assume that ${}^{f_1}\mathbf{R}_{i,xy}$ is known, leaving only ${}^{f_1}\psi_i$ needs to estimate. Furthermore, every frame $f_i$ is defined to be different only in the rotation around $Z$ axis, denoted as ${}^{f_1}\psi_{f_i}$. Thus, ${}^{f_1}\mathbf{R}_{f_i} = \ \mathbf{R}_z({}^{f_1}\psi_{f_i})$ and we can rewrite the individual robot state,

$$\mathbf{x}_i = \begin{bmatrix} {}^{f_i}\mathbf{p}_i^T\ {}^{f_1}\psi_i\ {}^{f_1}\mathbf{p}_{f_i}^T\ {}^{f_1}\psi_{f_i} \end{bmatrix}^T$$

We utilize a linear system whose input is the velocity. The velocity input is assumed to be corrupted with i.i.d zero-mean Gaussian noise. The robots' position and yaw can be modeled as follows

$${}^{f_i}\dot{\mathbf{p}}_i = {}^{f_i}\ \mathbf{v}_i \ +\ {}^{f_i}\eta_{\mathbf{v_i}}\ , \quad {}^{f_1}\dot{\psi}_i = {}^{f_1}\ \omega_i \ +\ {}^{f_1}\eta_{\omega_\mathbf{i}}$$

Frames $f_2, \dots f_N$ have unknown, fixed transformations with respect to frame $f_1$ but there can be drifts due to errors from VIO systems,

$${}^{f_1}\dot{\mathbf{p}}_{f_i} = 0 \ +\ {}^{f_1}\eta_{\mathbf{p_i}}\ , \quad {}^{f_1}\dot{\psi}_{f_i} = 0 \ +\ {}^{f_1}\eta_{\psi_\mathbf{i}}$$

We can write the state equation in a standard form

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u} + \omega$$

where $A = \mathbf{0}_{7N},\ \omega\ \sim \mathcal{N}(0, Q)$

$$B = \begin{bmatrix} I_4 & \mathbf{0}_4 & \mathbf{0}_4 & \dots & \mathbf{0}_4 \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{0}_4 & \dots & \mathbf{0}_4 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{0}_4 & \dots & I_4 \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{0}_4 & \dots & \mathbf{0}_4 \end{bmatrix},\ \mathbf{u} \ = \begin{bmatrix} {}^{f_1}\mathbf{v}_1 \\ {}^{f_1}\omega_1 \\ \vdots \\ {}^{f_N}\mathbf{v}_N \\ {}^{f_1}\omega_N \end{bmatrix}$$

$Q$ is the covariance matrix of the i.i.d Gaussian noise. We discretize this continuous time system using zero-hold for the input

$$\mathbf{x}_k = F\mathbf{x}_{k-1} + G\mathbf{u}_{k-1} + \omega_{k-1} \tag{5}$$

where $k = 1, \dots, T$ is the current time step, $F = I_{7N}, G = B\Delta t$, $\omega_{k-1} \sim \mathcal{N}(0, Q_d)$, $Q_d = Q\Delta t$, $\Delta t$ is the sampling time.

### B. The Measurement Model

During the update process, we update the robot state estimates using two types of measurements in a decoupled manner.

*1) Odometry Measurements:* VIO systems directly provide each robot its state represented in its fixed frame. We rewrite Eq. 3 as follows, taking into account a Gaussian noise,

$$\mathbf{z}_i \ = \begin{bmatrix} {}^{f_i}\mathbf{p}_i \ +\ {}^{f_i}\epsilon_{\mathbf{p_i}} \\ \mathbf{R}_z({}^{f_1}\psi_{f_i} + {}^{f_1}\ \epsilon_{\psi_{\mathbf{f_i}}})^T\ \mathbf{R}_z({}^{f_1}\psi_i + {}^{f_i}\ \epsilon_{\psi_\mathbf{i}})\ {}^{f_1}\mathbf{R}_{i,xy} \end{bmatrix}$$

where ${}^{(.)}\epsilon_{(.)}$ denotes noises. There is no data association involved in this partial update step.

*2) Detection Measurements:* We assume that each robot can detect all other robots with detection probability $P_D$ and some probability of false positive, false negative. This assumption can be achieved using omnidirectional cameras and that robots are in the range of detection. For every true measurement, the one generated from an actual robot, Eq. 4 can be rewritten with white noise added,

$${}^{i}\mathbf{z}_j = \ \mathbf{R}_z({}^{f_1}\psi_i)^T\ ({}^{f_1}\mathbf{p}_j \ -\ {}^{f_1}\mathbf{p}_i) \ +\ {}^{i}\eta_\mathbf{j}$$

Let $F_{t=0,\dots,K}$ denote random variables representing the number of false positives at time $t$. We assume $F$ to have Poisson distribution,

$$P_{F_t}(F) = \exp(-\lambda V)(\lambda V)^F/(F!)$$

*3) Detection Measurement Permutation:* To handle the unknown data association as well as the false positives, false negatives, we define the following helper variables, similar to those defined in [1]. Note that each robot has exactly $N-1$ targets.

- $M_i$, number of measurements at current time on robot $\mathcal{R}_i$
- $\phi_{i,j} \in \{0, 1\}$, an indicator that tells whether robot $\mathcal{R}_j$ is detected by robot $\mathcal{R}_i$ among $M_i$ target measurements
- $\boldsymbol{\phi}_i$, a $(N-1)$-dimensional vector stack of all $\phi_{i,j}$

- $D_i = \sum_{j=1}^{N} \phi_{i,j}$, the number of detected robots on robot $\mathcal{R}_i$
- $\tilde{\chi}_i$, a $D_i \times M_i$, a permutation of $D_i$ true measurements among $M_i$ relative measurements at current time

Given $M_i$ measurements on robot $\mathcal{R}_i$, it can be understood that $\boldsymbol{\phi}_i$ is a random variable possible that represents which robots are actually detected among $(N-1)$ targets and there can be $2^{N-1}$ such $\boldsymbol{\phi}_i$ outcomes. For each outcome $\boldsymbol{\phi}_i$, there could be many possible ways to match it with the detection measurements. This data association is unknown since there is no identity information available. If we call $\tilde{\chi}_i$ a possible match, all possible association events happened the $M_i$ detection measurements on robot $\mathcal{R}_i$ can be represented by tuples $(\boldsymbol{\phi}_i, \tilde{\chi}_i)$. We call these tuples data association hypotheses, or simply hypotheses. The approach introduced in the next section centers around how to find all feasible $(\boldsymbol{\phi}_i, \tilde{\chi}_i)$-hypotheses and update the system state based on the probability of each hypothesis, for every robot $\mathcal{R}_i$.

## V. Method

As can be seen from Eq. 4, each detection measurement depends on the state of multiple robots, making it improper to use the standard JPDA filter [31] for doing state estimation. Instead, we develop an extension of CPDA filter [1] for the nonlinear measurement model. When a robot receives a set of relative measurements, it is considered the station while other robots are considered targets. To simplify the notations, in the next section, we represent our extension of CPDAF in case a robot $\mathcal{R}$ served as the station with $L$ targets, $M$ detection measurements and 1-dimensional states.

### A. CPDAF Extension for Nonlinear Measurement Model

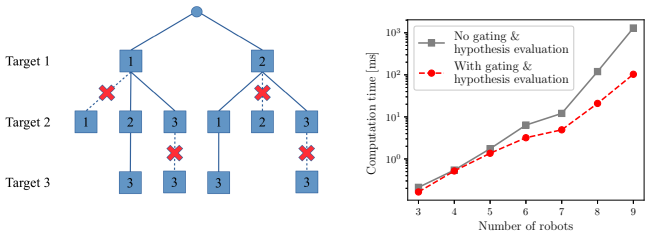
*1) CPDAF Step 1 - Prediction:* We denote $\mathbf{x}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}$ as the state and covariance of the state at time step $k-1$, respectively. The prior estimate for the state and covariance at time step $k$ is obtained, based on the system model in Eq. 5,

$$\begin{aligned} \mathbf{x}_{k|k-1} &= F\mathbf{x}_{k-1|k-1} + G\mathbf{u}_{k-1} \\ \mathbf{P}_{k|k-1} &= F\mathbf{P}_{k-1|k-1}F^T + Q_d \end{aligned} \tag{6}$$

*2) CPDAF Step 2 - Gating:* This step aims to reduce the number of possible measurements in the $M$-measurement set that can be assigned to each target robot. In our problem, the detection measurement is nonlinear, due to Eq. 4. Thus the gating for target robot $j$

$$G_j := \{z_j | (z_j - h_j(\mathbf{x}))^T P_{z_j z_j}^{-1} (z_j - h_j(\mathbf{x})) \leq \gamma\} \tag{7}$$

Where $P_{zz}$ is the innovation covariance matrix computed as same as in UKF [32] and $P_{z_j z_j}$ is the block of this matrix corresponding to target $j$. $\gamma$ is a threshold taken from the inverse chi-squared cumulative distribution at a significance level $P_G$ and the degree of freedom equal to dimension of $h_j(\mathbf{x})$.




(a) (b)

Fig. 2: (a): Hypothesis tree for three targets and three measurements after gating. $\emptyset$ is omitted for simplicity. The set of valid hypotheses is $\{(1,2,3),(2,1,3)\}$
(b): Comparison of the computation time required for the CPDAF update step with and without gating and hypothesis evaluation. Note that the Y-axis has log scale.

*3) CPDAF Step 3 - Evaluation of Hypotheses:* As [33] pointed out, the total number of $(\boldsymbol{\phi}, \tilde{\chi})$-hypotheses for a set of measurement $M$ on a robot $\mathcal{R}$ is

$$\sum_{D}^{\min(M,L)} \binom{L}{D} \binom{L}{D} D! \tag{8}$$

This can make evaluating them over time intractable when $L$ and $M$ are large. To tackle this problem, we propose an efficient evaluating algorithm that is inspired by [34] and starts with an association hypothesis tree of depth $L$. Each level of this tree features nodes - detection measurements - which can be associated with the target robot represented at this level. There can be $\emptyset$ node, indicating that the target is not detected at all. A valid hypothesis is a path connecting the root to one single node at every level, such that each node, except $\emptyset$, exists exactly one time on this path. To find these valid paths, the remaining part of this algorithm executes a depth-first traversal. An illustration of the algorithm is shown in Fig. 2(a). We omit its details due to the space limit.

*4) CPDAF Step 4 - Measurement-based Update:* As mentioned in section IV, the measurement-based update is separated into two update steps. The first update is based on the odometry measurement, as same as in UKF [32], and the second update is based on the detection measurement.

The later update is based upon the list of valid hypotheses obtained after step 2 and step 3. We extend CPDAF to compute the probability of a hypothesis $(\boldsymbol{\phi}, \tilde{\chi})$ [1] in case the measurements are nonlinear,

$$\beta(\boldsymbol{\phi}, \tilde{\chi}) = \frac{1}{c} F(\boldsymbol{\phi}, \tilde{\chi}) . \lambda^{L-D} \prod_{i=1}^{L} (1 - P_D)^{1-\boldsymbol{\phi}_i} (P_D)^{\boldsymbol{\phi}_i}$$

where $c$ is the normalization factor, $\lambda$ is the false observation spatial density, $P_D$ is the detection probability of a target robot. $P_D$ can vary among robots.

$$F(\boldsymbol{\phi}, \tilde{\chi}) = \frac{\exp\left(\mu(\boldsymbol{\phi}, \tilde{\chi})^T S(\boldsymbol{\phi})^{-1} \mu(\boldsymbol{\phi}, \tilde{\chi})\right)}{\sqrt{(2\pi)^D \det(S(\boldsymbol{\phi}))}}$$

where

$$\mu(\boldsymbol{\phi}, \tilde{\chi}) = \tilde{\chi}\mathbf{z} - \Phi(\boldsymbol{\phi})\mathbf{h}(\mathbf{x})$$
$$S(\boldsymbol{\phi}) = \Phi(\boldsymbol{\phi}) P_{zz} \Phi(\boldsymbol{\phi})^T$$

with $\mathbf{x}$ is the system state at current time which is $(L+1)$ dimensional, $H$ is the $L \times (L+1)$ dimensional observation matrix, $\mathbf{z}$ is the $M$-dimensional vector of stacked detection observations, $\Phi(\boldsymbol{\phi})$ is a $D \times N$ binary matrix with $r^{th}$ row equal to $r^{th}$ non-zero row of $\text{diag}(\boldsymbol{\phi})$, and $P_{zz}$ is the innovation covariance matrix, computed as same as in UKF [32].

Based on [1], we derive the state and covariance udpate as follows.

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \sum_{\boldsymbol{\phi}} K(\boldsymbol{\phi}) \sum_{\tilde{\chi}} \beta(\boldsymbol{\phi}, \tilde{\chi}) \mu(\boldsymbol{\phi}, \tilde{\chi})$$

$$\begin{aligned}\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} &- \sum_{\boldsymbol{\phi}} K(\boldsymbol{\phi}) \phi(\boldsymbol{\phi}_i) P_{zx} \sum_{\tilde{\chi}} \beta(\boldsymbol{\phi}_i, \tilde{\chi}) \\ &+ \sum_{\boldsymbol{\phi}} K(\boldsymbol{\phi}) \left( \sum_{\tilde{\chi}} \beta(\boldsymbol{\phi}, \tilde{\chi}) \mu(\boldsymbol{\phi}, \tilde{\chi}), \mu(\boldsymbol{\phi}, \tilde{\chi})^T \right) K(\boldsymbol{\phi})^T \\ &- \left( \sum_{\boldsymbol{\phi}} K(\boldsymbol{\phi}) \sum_{\tilde{\chi}} \beta(\boldsymbol{\phi}, \tilde{\chi}) \mu(\boldsymbol{\phi} \right) \\ &\times \left( \sum_{\boldsymbol{\phi}} K(\boldsymbol{\phi}') \sum_{\tilde{\chi}'} \beta(\boldsymbol{\phi}', \tilde{\chi}') \mu(\boldsymbol{\phi}' \right)\end{aligned}$$

where $K(\boldsymbol{\phi}) = (\Phi(\boldsymbol{\phi}) P_{zx})^T P_{zz}^{-1}$ is the Kalman gain, $P_{zx}$ is the measurement-state cross-covariance matrix.

### B. Time Complexity Analysis

Step 1 - executing model prediction takes $\mathcal{O}(N)$.

Step 2 iterates over $M$ target measurements on every robot $\mathcal{R}$ to prune out measurements that are not in its validation gate. Step 2 has $\mathcal{O}(\sum_{i=1}^{N} M) = \mathcal{O}(MN)$ time complexity.

Step 3 essentially carries out a depth-first traversal over all nodes and edges. In the worst case, each level, excepting the root, consists of $M+1$ nodes. Since every node in a level is connected to every node in the next level, there will be $\mathcal{O}(M^2)$ edges connecting two consecutive levels. Thus, the total of nodes and edges, or the worst-case time complexity for traversing the tree is $\mathcal{O}(NM + NM^2) = \mathcal{O}(NM^2)$.

The running time for the first update in step 4 is as same as in UKF [32]. The second update depends on how many valid hypotheses selected. In the worst case, this number is exponential with respect to the number of robots and measurements, making step 4 exponential in time of computation. In practice, this number can be largely reduced thanks to step 2 and step 3. Also, some techniques such as $k$-best hypotheses can make step 4 manageable. $k$-best hypotheses algorithm [35], [36] is $\mathcal{O}(kN^3)$. In our case, the complexity of finding $k$-best hypotheses on a tree is $\mathcal{O}(k \times \max(N, M)^3)$.

## VI. Experiments

We evaluate the efficiency of the proposed framework and algorithms on a simulation.

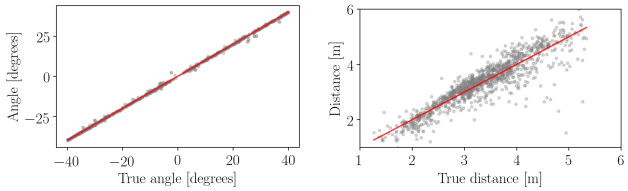



Fig. 3: Vision-based measurement models: (a) bearing; (b) distance. Red line: model v.s. true value, gray dots: measurement v.s. true value.

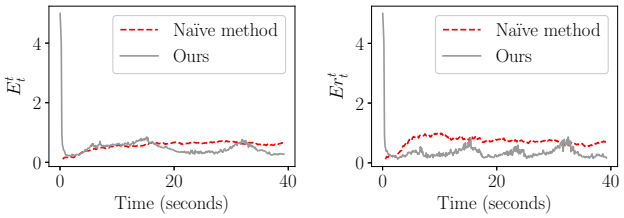



Fig. 4: From left to right: absolute and relative state errors during changes from 1.35 m-radius circle → 2.7 m-radius circle → 1.35 m-radius circle → 2.7 m-radius circle → 1.35 m-radius circle.

### A. Simulation Settings

The simulation is in ROS. The robots simulate our FLA Falcon 250 platforms as shown in Fig. 1. In reality, each robot is featured with an Open Vision Computer [37]. The measurements that robots receive simulate measurement models that we obtain by doing real experiments on real robot platforms.

*1) Odometry Measurement Model:* The VIO error is modeled as a multivariate Gaussian distribution,

$$\mathbf{z}_i|_{measured} = \mathbf{z}_i|_{true} + \mathcal{N}(0, \sigma_o^2)$$

where the standard deviation $\sigma_o$ is $0.01\,\text{m}$ for elements in the transition and $0.002$ rad for euler angle elements in the orientation.

*2) Detection Model:* We utilize MAVNet [38], a light-weight and fast network for vision-based robot detection. The output segmentation is used to estimate the distance from the camera to the target as well as the bearing, assuming the robot's dimension in the 3D world is known.

The bearing measurement, as shown in Fig. 3(a), is modelled as,

$$bearing|_{measured} = bearing|_{true} + \mathcal{N}(0, \sigma_b^2)$$

where $\sigma_b = 0.008$ rad.

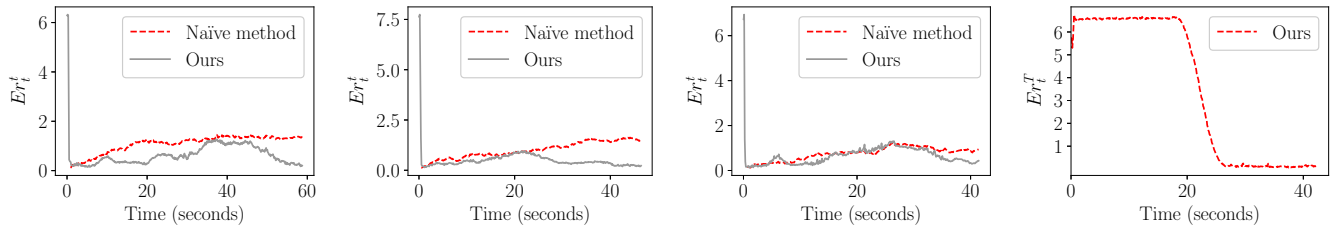
The distance measurement as shown in Fig. 3(b), is modelled as,

$$distance|_{measure} = distance|_{true} + \mathcal{N}(0, \sigma_d^2)$$

with $\sigma_d = 0.0495 * distance|_{true} + 0.0336$ m.

### B. Evaluation Metrics

We use two metrics: 1) $E_t^t$ in Eq. 1 that evaluates the *absolute state* - the state of robots within the common fixed frame, $f_1$, and 2) $Er_t^t$ that evaluates the *relative state* - the




(a) (b) (c) (d)

Fig. 5: (a,b,c): Relative state errors during the system's configuration changes. (a): 1.35 m-radius circle → v-line → line of 12 m → 1.35 m-radius circle → line of 6 m; (b): line of 6 m → line of 12 m → v-line → line of 6 m; (c): line of 6 m → 1.35 m-radius circle → 2.7 m-radius circle → line of 12 m → line of 6 m
(d): Convergence of relative states to the desirable relative state as the robots take off and form a line of 6 m.

state of robots within the common moving frame, i.e. frame 1 of robot $\mathcal{R}1$.

$$Er_t^t = ||\bar{\mathcal{G}}r_t - \mathcal{G}r_t||_2^2 \tag{9}$$

where $\bar{\mathcal{G}}r_t = \{({}^1\mathbf{p}_i, {}^1\mathbf{R}_i) \mid i \in \{1, \dots N\}\}$ is the set of relative poses, $\mathcal{G}r_t$ is the ground truth.

### C. State Estimation Performance

We evaluate the performance of our algorithm in comparison with a naïve method. This comparison method assumes a known initial system configuration and integrates the VIO measurements over time. In each experiment, a team of homogeneous 7-robots are controlled to form different spatial configurations in a centralized manner. Each robot is controlled to follow a predefined path using its true state. The first experiment results, as depicted in Fig. 4, show that our state estimation converges very fast and matches the naïve method in the absolute state error while performing better in the relative state error. Fig. 5 also shows our method's superior performance in the other three experiments in relative state errors. Absolute state errors are omitted since both methods perform similarly.

### D. Formation Control Using Estimated State

We demonstrate a use case of our approach by designing an open-loop controller for formation flight. Our controller takes as input the initial estimate of the system state and determines an optimal path for every robot using minimum snap trajectory generation [39]. Each robot is then controlled to the final state using its current estimated state as the feedback. All steps, from state estimation to control is done on-board.

We launch 7 robots forming a line of 6 m and evaluate $Er_t^T$ which measures the convergence of the relative state estimate versus the final *relative state*.

$$Er_t^T = ||\bar{\mathcal{G}}r_t - \mathcal{G}r_T||_2^2 \tag{10}$$

Fig. 5(d) shows the error converges to 0 illustrating that our controller can converge to the desired state when the robot team forms the desirable line.

### E. Effectiveness of Gating and Hypothesis Evaluation

Fig. 2(b) shows the computation time of our proposed approach with various numbers of robots in two different settings: with and without using gating and hypothesis evaluation steps. Since these two steps eliminate unnecessary hypotheses, they help the algorithm run much faster. The higher number of robots, the more time saving can be achieved, starting from 1.29 times with 3 robots to 12.6 times for 9 robots.

## VII. Conclusions

This work extends the CPDAF algorithm to address the localization of a system of SWaP-constrained aerial platforms. We address three main challenges: 1) unknown data association between vision-based relative measurements and robot targets, 2) the need to bootstrap the system from an unknown initial configuration, and 3) noisy vision-based measurements with false negatives and false positives. Experiments on a simulator driven by measurement models statistically derived from the experimental data demonstrate the superior performance of our approach. We also illustrate how the proposed algorithm can be used in a simple open-loop controller, extending the capability of using onboard sensing for estimation and control in formation flight. Our future work is to develop a closed-loop controller as well as to improve the hypothesis evaluation step.

## References

[1] H. A. Blom and E. A. Bloem, "Probabilistic data association avoiding track coalescence," *IEEE Transactions on Automatic Control*, vol. 45, no. 2, pp. 247–259, 2000.

[2] L. C. Pimenta, G. A. Pereira, M. M. Gonçalves, N. Michael, M. Turpin, and V. Kumar, "Decentralized controllers for perimeter surveillance with teams of aerial robots," *Advanced Robotics*, vol. 27, no. 9, pp. 697–709, 2013.

[3] D. Saldana, R. J. Alitappeh, L. C. Pimenta, R. Assunçao, and M. F. Campos, "Dynamic perimeter surveillance with a team of robots," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 5289–5294.

[4] A. Marino, L. E. Parker, G. Antonelli, and F. Caccavale, "A decentralized architecture for multi-robot systems based on the null-space-behavioral control with application to multi-robot border patrolling," *Journal of Intelligent & Robotic Systems*, vol. 71, no. 3-4, pp. 423–444, 2013.

[5] D. Portugal and R. P. Rocha, "Cooperative multi-robot patrol with bayesian learning," *Autonomous Robots*, vol. 40, no. 5, pp. 929–953, 2016.

[6] A. Marjovi, J. G. Nunes, L. Marques, and A. de Almeida, "Multi-robot exploration and fire searching," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 1929–1934.

[7] N. Basilico and F. Amigoni, "Exploration strategies based on multi-criteria decision making for searching environments in rescue operations," *Autonomous Robots*, vol. 31, no. 4, p. 401, 2011.

[8] A. Franchi, C. Masone, V. Grabe, M. Ryll, H. H. Bülthoff, and P. R. Giordano, "Modeling and control of uav bearing formations with bilateral high-level steering," *The International Journal of Robotics Research*, vol. 31, no. 12, pp. 1504–1525, 2012.

[9] M. Turpin, N. Michael, and V. Kumar, "Decentralized formation control with variable shapes for aerial robots," in *2012 IEEE international conference on robotics and automation*. IEEE, 2012, pp. 23–30.

[10] N. Michael, M. M. Zavlanos, V. Kumar, and G. J. Pappas, "Distributed multi-robot task assignment and formation control," in *2008 IEEE International Conference on Robotics and Automation*. IEEE, 2008, pp. 128–133.

[11] A. Kushleyev, D. Mellinger, C. Powers, and V. Kumar, "Towards a swarm of agile micro quadrotors," *Autonomous Robots*, vol. 35, no. 4, pp. 287–300, 2013.

[12] O. Poulet, F. Guérin, and F. Guinand, "Self-localization of anonymous mobile robots from aerial images," in *2018 European Control Conference (ECC)*. IEEE, 2018, pp. 1094–1099.

[13] A. Weinstein, A. Cho, G. Loianno, and V. Kumar, "Visual inertial odometry swarm: An autonomous swarm of vision-based quadrotors," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1801–1807, 2018.

[14] V. Indelman, E. Nelson, N. Michael, and F. Dellaert, "Multi-robot pose graph localization and data association from unknown initial relative poses via expectation maximization," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 593–600.

[15] E. Montijano, E. Cristofalo, D. Zhou, M. Schwager, and C. Saguees, "Vision-based distributed formation control without an external positioning system," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 339–351, 2016.

[16] A. Wasik, P. U. Lima, and A. Martinoli, "A robust localization system for multi-robot formations based on an extension of a gaussian mixture probability hypothesis density filter," *Autonomous Robots*, pp. 1–20, 2019.

[17] K. Guo, Z. Qiu, W. Meng, L. Xie, and R. Teo, "Ultra-wideband based cooperative relative localization algorithm and experiments for multiple unmanned aerial vehicles in gps denied environments," *International Journal of Micro Air Vehicles*, vol. 9, no. 3, pp. 169–186, 2017.

[18] S. Vemprala and S. Saripalli, "Vision based collaborative localization for multirotor vehicles," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 1653–1658.

[19] K. Leahy, E. Cristofalo, C.-I. Vasile, A. Jones, E. Montijano, M. Schwager, and C. Belta, "Control in belief space with temporal logic specifications using vision-based localization," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 702–722, 2019.

[20] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[21] T. Nguyen, S. W. Chen, S. S. Shivakumar, C. J. Taylor, and V. Kumar, "Unsupervised deep homography: A fast and robust homography estimation model," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2346–2353, 2018.

[22] K. Mohta, M. Watterson, Y. Mulgaonkar, S. Liu, C. Qu, A. Makineni, K. Saulnier, K. Sun, A. Zhu, J. Delmerico *et al.*, "Fast, autonomous flight in gps-denied and cluttered environments," *Journal of Field Robotics*, vol. 35, no. 1, pp. 101–120, 2018.

[23] A. Franchi, G. Oriolo, and P. Stegagno, "Mutual localization in multi-robot systems using anonymous relative measurements," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1302–1322, 2013.

[24] R. Spica and P. R. Giordano, "Active decentralized scale estimation for bearing-based localization," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 5084–5091.

[25] G. L. Mariottini, G. Pappas, D. Prattichizzo, and K. Daniilidis, "Vision-based localization of leader-follower formations," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 635–640.

[26] R. Tron, J. Thomas, G. Loianno, K. Daniilidis, and V. Kumar, "A distributed optimization framework for localization and formation control: Applications to vision-based measurements," *IEEE Control Systems Magazine*, vol. 36, no. 4, pp. 22–44, 2016.

[27] D. Dias, R. Ventura, P. Lima, and A. Martinoli, "On-board vision-based 3d relative localization system for multiple quadrotors," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. Ieee, 2016, pp. 1181–1187.

[28] C.-H. Chang, S.-C. Wang, and C.-C. Wang, "Vision-based cooperative simultaneous localization and tracking," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 5191–5197.

[29] A. Franchi, G. Oriolo, and P. Stegagno, "Mutual localization in a multi-robot system with anonymous relative position measures," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3974–3980.

[30] ——, "On the solvability of the mutual localization problem with anonymous position measures," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 3193–3199.

[31] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, no. 6, pp. 82–100, 2009.

[32] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. Ieee, 2000, pp. 153–158.

[33] D. F. Crouse, Y. Bar-Shalom, P. Willett, and L. Svensson, "The jpdaf in practical systems: Computation and snake oil," in *Signal and Data Processing of Small Targets 2010*, vol. 7698. International Society for Optics and Photonics, 2010, p. 769813.

[34] B. Zhou and N. Bose, "Multitarget tracking in clutter: Fast algorithms for data association," *IEEE Transactions on aerospace and electronic systems*, vol. 29, no. 2, pp. 352–363, 1993.

[35] K. G. Murty, "An algorithm for ranking all the assignments in order of increasing cost," *Operations Research*, vol. 16, no. 3, pp. 682–687, 1968. [Online]. Available: http://www.jstor.org/stable/168595

[36] M. L. Miller, H. S. Stone, and I. J. Cox, "Optimizing murty's ranked assignment method," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 3, pp. 851–862, 1997.

[37] M. Quigley, K. Mohta, S. S. Shivakumar, M. Watterson, Y. Mulgaonkar, M. Arguedas, K. Sun, S. Liu, B. Pfrommer, V. Kumar *et al.*, "The open vision computer: An integrated sensing and compute system for mobile robots," in *2019 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.

[38] T. Nguyen, S. S. Shivakumar, I. D. Miller, J. Keller, E. S. Lee, A. Zhou, T. Özaslan, G. Loianno, J. H. Harwood, J. Wozencraft, C. J. Taylor, and V. Kumar, "Mavnet: An effective semantic segmentation micro-network for mav-based tasks," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3908–3915, Oct 2019.

[39] D. Mellinger and V. Kumar, "Minimum snap trajectory generation and control for quadrotors," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 2520–2525.