2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

# Bayesian Fusion of Unlabeled Vision and RF Data for Aerial Tracking of Ground Targets

Ramya Kanlapuli Rajasekaran[1], Nisar Ahmed[1] and Eric Frew[1]

***Abstract*— This paper presents a method for target localization and tracking in clutter using Bayesian fusion of vision and Radio Frequency (RF) sensors used aboard a small Unmanned Aircraft System (sUAS). Sensor fusion is used to ensure tracking robustness and reliability in case of camera occlusion or RF signal interference. Camera data is processed using an off-the-shelf algorithm that detects possible objects of interest in a given image frame, and the true RF emitting target needs to be identified from among these if it is present. These data sources, as well as the unknown motion of the target, lead to a heavily non-linear non-Gaussian target state uncertainties, which are not amenable to typical data association methods for tracking. A probabilistic model is thus first rigorously developed to relate conditional dependencies between target movements, RF data and visual object detections. A modified particle filter is then developed to simultaneously reason over target states and RF emitter association hypothesis labels for visual object detections. Truth model simulations are presented to compare and validate the effectiveness of the RF + visual data fusion filter.**

## I. Introduction

Small unmanned aerial systems (sUAS) have been used for surveillance, search and rescue and reconnaissance missions in recent times due to their small size, and their ability to fly over obstacles and terrain unfit for humans. Small, lightweight, easy to use sensors like radio frequency (RF) sensors and cameras are used popularly on UAVs. For this application, RF receivers on the sUAS receive signals from RF transmitters like cellphones [1]. A vision sensor (camera) mounted on a sUAS provides good visual awareness of the surroundings. Cameras are also inexpensive, lightweight, and provide a continuous stream of information. However, camera sensors on sUAS are subject to occlusion from trees and low clouds. Alternatively, RF receivers are subject to interference and other signals that can throw off the measurement. This paper describes the fusion of information from RF and vision sensors and how that can lead to accurate, robust, reliable and continuous ground target localization and tracking using sUAS. Performing sensor fusion helps performance stay robust to sensor error, as the information from each sensor is verified by the other, and other artifacts from noise and clutter can be filtered out for target state estimation (Fig 1).

This work is funded by the Center for Unmanned Aircraft Systems (C-UAS), a National Science Foundation Industry/University Cooperative Research Center (I/UCRC) under NSF Award No. CNS-1650468 along with signicant contributions from C-UAS industry members.

[1]Ramya Kanlapuli Rajasekaran, Nisar Ahmed and Eric Frew are with the Ann & H.J. Smead Department of Aerospace Engineering Sciences at the University of Colorado Boulder, Boulder Colorado. raka1840@colorado.edu, nisar.ahmed@colorado.edu,eric.frew@colorado.edu

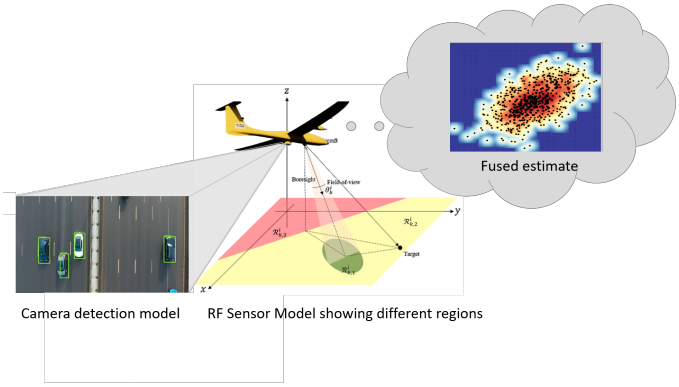


Fig. 1. Overview diagram showing RF sensor model and camera observation

This paper targets a two-tiered problem - data association in a cluttered environment, and target localization and tracking of a single RF emitter. The data stream from the camera is post-processed using an object detection algorithm to provide a set of objects of interest (OOIs). OOIs in this case, are cars that might contain the RF emitter. The RF emitter can change positions between different OOIs (for example, being tossed from one car to another). This detection set consists of the true mobile RF emitter target and clutter. Clutter comprises the other detected OOIs in the camera field of view, which are unlabeled and thus otherwise indistinguishable from each other. As such, data association is necessary to differentiate and identify the true emitter from the clutter(other detected OOIs). After identifying the true target, estimating the position and velocity of the target comprises the second part of the problem. The sensor models of both the RF and camera data are heavily non-linear and described by highly non-Gaussian uncertainties. The true target motion model is also unknown. These features limit the applicability of conventional target state estimation and data association techniques for solving this problem.

This paper proposes a Bayesian framework for fusion of cluttered camera and RF data to perform tracking of ground targets from sUAS. In this paper we extend the concept of using Sequential Monte Carlo filters (i.e. particle filters) [2]–[4] for non-linear non-Gaussian target localization and tracking. Since reliable data association in clutter for non-linear, non-Gaussian sensor models is a complex, difficult problem, with no "one-size-fits-all" solution, the estimation architecture developed in [1] for RF-only tracking is significantly further developed here to incorporate data association for unlabeled cluttered visual OOI measurements.

The state-of-the-art tracking solution that has been studied and cited repeatedly in the context of sUAS applications is vision-based tracking. [5]–[7]. The fusion of RF and camera sensor data for target tracking is an open problem with sparse documentation. The major contribution of this paper therefore is the development of a new robust target state estimation and data association architecture for onboard sUAS operations that support fusion of RF and camera-based measurements.

The specific technical contributions of this paper are: 1) a probabilistic model for data association and fusion that rigorously establishes statistical dependencies between the visual data association and RF emitting target tracking problem; 2) an approach for performing track data association in clutter for unlabeled OOI observations produced by off-the-shelf visual object detectors; and 3) a rigorous technique for approximate online Bayesian fusion of RF and vision data, which also accounts for data association to track the specific RF-emitting target of interest. The rest of the paper is organized as follows: Section II outlines background and related work, and formulates the sUAS-based RF emitter tracking problem; Section III describes the probabilistic model, the data association and estimation method; Section IV presents simulation results and discussion for various implementations of the filter; and Section V concludes the paper and discusses future work.

## II. Background and Problem Formulation

### A. Background and related work

Localization and tracking of moving ground targets from fixed wing UAS using vision-based systems has been previously studied and outlined in various publications, e.g. [5]–[7]. Data association in clutter is a field that has also been explored extensively for different tracking problems involving multiple targets and different sensing systems. The Probabilistic Data Association filter (PDAF) and the Joint Probabilistic Data Association filter (JPDAF) [8] performs well segregating multi target observations and clutter for linear systems (and non-linear systems that are at least weakly linear). However, JPDAFs are computationally complex and expensive for more evolved, non-linear multi-modal sensor models [9], which makes them not ideal for our RF target tracking problem. Approaches like PDA and JPDA, run linear(ized) Kalman Filter (KF)s/Extended Kalman Filter (EKF)s to perform state estimation, and the performance tends to deteriorate as non-linearities are more severe. KFs along with a data association methods have long been used to solve aerial tracking problems [10] [11], but previous applications have mostly focused on linear motion models and Gaussian assumptions about process and measurement uncertainties. The vision-based clutter and data association problem for RF-vision data fusion described in this paper requires both highly non-linear measurement models and highly non-Gaussian models for state and measurement uncertainties. In particular, the combination of RF data and occasional 'negative' visual object detection information introduces severe non-Gaussianity into track measurement updates [12]. Hence, conventional data association methods derived for KF/EKF state estimators are ill-suited to the problem at hand.

To cope with data association in tracking problems with severe non-linear/non-Gaussian characteristics, Rao-Blackwellized Particle filtering strategies (RBPFs) [13], [14] have been used. However, RBPFs require closed-form analytical solutions for some part of the target state estimation problem. Since this cannot be achieved in an obvious way for combining RF and visual detection data, RBPF techniques are not applicable here.

Recursive-Random Sample Consensus (R-RANSAC) is a method that has been used to perform multiple target tracking in clutter using vision based sensors aboard UAS [7], [15]. However, the problem presented in this paper differs from the problems discussed and assumptions made in [7], [15]. In [7] the RANSAC algorithm establishes and identifies multiple tracks for different moving objects in the frame respectively. But, the target track to be followed is picked by a human controller, hence eliminating the need for the type of automated data association considered in this paper. Ref. [16] claims that R-RANSAC is modular and can be used with non-linear system models, explaining that R-RANSAC uses an inlier threshold to compute an association matrix, which in turn computes weights for tracks. The inlier function is a process of gating measurements and the paper uses a fixed rectangular gate that works because of the linear model assumption. However, the process of gating and building a gate shape is not trivial for a non-linear, non-Gaussian function as described in [17]. Rectangular and ellipsoidal gates fail, and an irregularly shaped complex gate would have to be approximated to implement R-RANSAC for our problem.

Recent work in [18] examines tracking of vehicles constrained to simple road networks using vision sensors and nested particle filters. Particles in this approach are constrained to the road, which makes convergence easier but also places strong assumptions on target behavior. The measurement likelihood model in [18] is assumed to be a mixture of a Gaussian and uniform PDF, which oversimplifies the complexities of the true RF and vision measurement likelihoods, and are thus far removed from the measurement models described by this work. The computational cost of the nested particle filter also scales as $O(HMN)$, where $H$ is the number of history particles, $N$ is the number of particles used and $M$ number of targets. From the simulations presented in [18], approximately 10,000 particles are used per target, which leads to a tremendous computational demand for onboard sUAS implementation.

Relevant literature addressing the fusion of RF and vision data for aerial target tracking is sparse. Ref. [19] implemented vision and RFID fusion for tracking people using mobile ground robots, which differ from aerial tracking and introduce more flexibility in computation constraints. Ref. [20] fused WiFi and surveillance camera measurements to perform object tracking. Both the above mentioned references used particle filters, however the sensors, implemen-

tations, and robots used differ significantly from the current problem. In Ref. [20] tracking is performed using information from multiple stationary cameras in different locations and the wireless signal strength values of multiple WiFi access points, as compared to our application that involves the use of a single camera, and a single RF receiver mounted on a moving aircraft. Ref. [20] also "hands over" information about the subject being tracked, identifying it for the camera, which is not the case in the specific scenario addressed in this paper. [19] and [20] do not mention any computation power constraints as these methods involve processing on ground computers.

Thus, it is clear that there is a need for developing state estimation techniques that can handle RF and unlabeled vision data association in clutter for the aerial tracking problem, which requires reliable localization and target tracking. The next section talks about filling this need, by outlining a novel probabilistic model for unlabeled sensor fusion.

### B. Target dynamics

The position state of the target at discrete time step $k$ is $r_k = [x_k, y_k]^T$ . where $x_k$ is the Easting and $y_k$ Northing in inertial coordinates. To simplify the initial development of a rigorous tracking approach, it is assumed that the position and orientation of the sUAS is known at all times and that the sUAS flies at a constant altitude. Consequently, the velocity state at time $k$ is given by $\dot{r}_k = [\dot{x}_k, \dot{y}_k]^T$, where $\dot{x}_k$ and $\dot{y}_k$ are components of the target velocity in the East and North direction. The state $R_k$ is defined as $R_k = [r_k^T, \dot{r}_k^T]^T$. The motion model is described as a function of the previous state and noise.

$$R_k = f_k(R_{k-1}, w_{k-1}), \tag{1}$$

This function does not necessarily have to be linear, but for simplicity, the motion model assumes a random acceleration noise model such that

$$R_k = F_k \cdot R_{k-1} + w_{k-1}, \tag{2}$$

where $w_{k-1} \sim \mathcal{N}(0, Q_{k-1})$ is the process noise at time $k-1$, and for time step $dt$,

$$F_k = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad G = \begin{bmatrix} dt^2/2 & 0 \\ 0 & dt^2/2 \\ dt & 0 \\ 0 & dt \end{bmatrix}. \tag{3}$$

The additive white Gaussian process noise vector $w_{k-1}$ is zero mean and has covariance matrix $Q = \alpha \cdot GG^T$, where $\alpha$ is the variance of the acceleration noise [1]. As mentioned earlier, the target entity carrying the RF emitter can change between OOIs, and this is hard to explicitly account for in this target motion model. Hence, it is necessary to solve the data association problem constantly, and there is no relationship between the associations at different time steps.

### C. RF sensor description

Radio frequency (RF) receivers are useful because they can receive signals from common electronic devices, and tracking the RF signals helps track these devices. The region-based non-linear RF sensor model that was previously developed in [1] is briefly described here. It is assumed that an RF transmitter is placed in a mobile platform (e.g. a car) and that it transmits at a known frequency, e.g. 2.4GHz for our applications. The sUAS has an omni-directional antenna and a directional antenna, tuned to receive at this frequency, and each of these antennae give a Received Power (RP) measurement. A detection measurement $z_k$ is derived by comparing the difference between the two received power measurements to a pre-defined, experimentally derived threshold $\tau$ such that,

$$z_k = \begin{cases} Detect\ (D) & \text{if } y_{RP}^{dir} - y_{RP}^{omni} > \tau \\ No\ Detect\ (ND) & \text{otherwise} \end{cases} \tag{4}$$

The region, where $y_{RP}^{dir} - y_{RP}^{omni} > \tau$ defines a cone, which projected on to the ground is an ellipse, which we refer to as the *sensor ellipse*. This is illustrated in Figs 1 and 3. The total area of interest $\mathbb{A}$, where the target could be present, is divided into $n_r$ regions.

$$\mathbb{A} = \bigcup_{r=1}^{n_r} \mathcal{G}_{k,r} \tag{5}$$

The probability of target detection can now be depicted as a conditional probability $\mathbb{P}(z_k|R_k, \mathcal{G}_{k,r})$. In this case, the regions are defined as: the projected sensor ellipse, $\mathcal{G}_{k,1}$; region ahead of the sensor but not in the projected ellipse, $\mathcal{G}_{k,2}$; and region behind the ellipse $\mathcal{G}_{k,3}$. The probability can be reformulated as:

$$\mathbb{P}(z_k|R_k; \mathcal{G}_{k,r}) = \begin{cases} \beta_r & \text{if } z_k = D \text{ given } r_k \in \mathcal{G}_{k,r} \\ 1 - \beta_r & \text{if } z_k = ND \text{ given } r_k \in \mathcal{G}_{k,r} \end{cases} \tag{6}$$

where $\beta_r$ represents the probability that the target is detected by the sensor of aircraft when the target is in the region $\mathcal{G}_{k,r}$. The probabilities $\beta_r$ are determined experimentally from multiple flight tests [1].

### D. Vision Sensor description and sensor model

Vision sensors mounted on a sUAS provide high visual awareness, making them a popular choice for tracking problems. The field of view of the camera is determined by the sUAS position, orientation, and camera mounting angle. The sUAS and the camera angle also determines whether the target is in the camera frame. The projection of the camera frame on to the inertial frame is referred to as $F_{in}$. The aircraft (with both visual and RF sensors) sweeps over the total surveillance area of interest to find and track the RF-emitting target.

An offline trained neural network detection algorithm detects objects of interest (OOIs) in the camera's frame/field of view. The OOIs are vehicles that could possibly carry the RF transmitter. The camera detects vehicles in an image frame and marks them with detection boxes. It does this

for each frame, and a frame is captured at every time step, leading to the observation $O_k$. $O_k$ consists of $B_k^1$, $B_k^2$, ... $B_k^{n_k}$, where $n_k$ is the number of object bounding boxes drawn over detected OOIs, and $B_k^m$ is the $m^{th}$ detection box at time $k$. The width and height of the bounding box is determined by the size of the car in the image frame.

The camera can image multiple vehicles in the camera frame, but the object detection system has no knowledge of which detected vehicles the transmitter is in. Furthermore, the OOI detector has no 'memory' of which objects it has seen previously, as the emitter can move between OOIs – i.e. $B_m$ are all unlabeled. To compensate for this lack of information, a latent variable $L_k$ will eventually be introduced to model the identity of the true target among the detected OOIs. By fusing RF and camera data to infer $L_k$ along with $R_k$, the sUAS can then differentiate and identify which OOI box $B_k^m$ in $O_k$ (if any) is the true RF emitting target.

*1) Observation model and information update:* Detection boxes exist in the image frame. However, the state $R_k$ exists in the inertial frame. Projecting the detection box from the camera field of view at time $k$ to the inertial frame gives us the inertial projected region $IPR_k$. To write out the observation model, we convert all dependent variables into inertial frame.

The observation $O_k$ is a function of the number of boxes detected at every time step $n_k$,

$$O_k = \begin{cases} \{B_k^1, B_k^2 .... B_k^{n_k}\}, & \text{if } n_k > 0 \\ \{\phi_{TN}, \phi_{FN}\}, & \text{if } n_k = 0 \end{cases} \quad (7)$$

Here, $\phi_{TN}$ and $\phi_{FN}$ are used to differentiate between a true negative and false negative when no objects were detected in frame ($n_k = 0$). A true negative refers to a measurement where no objects are detected in the frame because, in actuality, there were no OOIs in the frame. A false negative refers to a measurement where no objects were detected in a frame because the object detection algorithm did not pick up any OOIs in the frame, despite there being at least one actual OOI in the frame. The separation of true negatives and false negatives allows us to account for negative information updates [12]. That is, when the camera and the algorithm find OOIs in a frame, the resulting bounding boxes can give measurements that allow us to reduce tracking uncertainty. However, when the detector does not pick up any OOIs in a frame, the negative information provided to us by the visual data must be fused efficiently and judiciously to perform a probabilistic update on the target's state uncertainty.

## III. Probabilistic Model for Unlabeled Vision and RF Data Fusion

In this section, the formal probabilistic graphical model in Fig. 2 for data association and fusion is first described. The RF measurement is described by the $z_k$ block and the vision measurement is described by the $O_k$ block. $L_k$ is the emitter association variable for vision data at time $k$, which must be inferred during process of fusing unlabeled vision and RF data. Next, the process for fusing multi-sensor data through Bayesian interference is described using a Sequential Monte Carlo (SMC) particle filter. The two sensors used in this implementation are RF and vision sensors, but the graphical model and SMC/particle filter approach can be extended to any other non-linear, non-Gaussian sensor models. This model and filtering method are tractable for onboard sUAS implementation, and does not require a complex or nested particle filter solution.

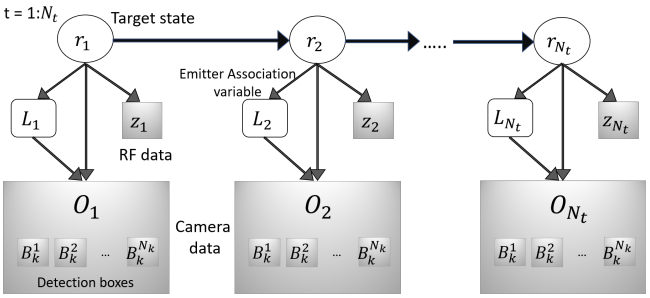


Fig. 2. Probabilistic model for data (shaded) and unknown (clear) variables.

### A. Emitter association variable, $L_k$

The latent variable $L_k$ is introduced to identify the detection boxes (in the image frame) or detection regions (in the inertial frame) observed by the detection and classification system. It is conditionally dependent on the location of the target and conveys what the data association hypothesis should be. In other words, $L_k$ identifies which (if any) of the unlabeled OOI detection boxes from the vision data is the actual RF emitting target. $L_k$ can be represented by a one hot vector encoding, where element $L_k[j] = 1$ means box $j$ is the true target and $L(1 : n_k) = 0$ means that none of the boxes is the true target. $L(1 : n_k) = \phi_{TN}, \phi_{FN}$ imply a true and false negative.

$$L_k = \{0, [1, 0, 0, ..], [0, 1, 0, 0, ...], \phi_{TN}, \phi_{FN}\} \quad (8)$$

The proper accounting of true and false negative measurement events is essential to fusing negative information about the state of the RF emitting target whenever it is not detected in camera images.

$L_k$ depends on the number of boxes detected in frame $n_k$ and the position of the true target relative to $F_{in}$. This model assumes the existence of misdetections. The probability of misdetection $P_{MD}$ is

$$P(L_k = 0 | r_k \text{ in } IPR_k, n_k > 0) = P_{MD} \quad (9)$$

The Conditional Probability Table (CPT) for $P(L_k|r_k, n_k)$ is given in Table I.

To perform an informed measurement update, from Eqn. 7, we can see that the CPT of $B_k^m$ is necessary. The conditional probability table for $P(B_k^m|r_k, L_k)$, is given in Table II. $B_k^m$ takes the value of the center coordinates of the detection box, and fixed box sizes are assumed for OOI simplicity. The value of the image pixel noise covariance matrix $\Sigma$ is determined by the vision sensor specifications and the environment the vision sensor operates in. In Table III-A, $\mathcal{U}(image)$ is a uniform distribution over the image frame. If

TABLE I

CONDITIONAL PROBABILITY TABLE FOR ASSOCIATION VARIABLE L

| $r_k$ | $n_k$ (# of OOI boxes) | $L_k$ | $P(L_k \mid r_k, n_k)$ |
|---|---|---|---|
| In $IPR_k$ | $> 0$ | $\{1, ..., n_k\}$ | $\frac{1-P_{MD}}{n_k}$ |
| In $IPR_k$ | $>0$ | 0 | $P_{MD}$ |
| Not in $IPR_k$ | $>0$ | $\{1....n_k\}$ | 0 |
| Not in $IPR_k$ | $>0$ | 0 | 1 |
| Not in $IPR_k$ | 0 | $\phi_{TN}$ | 0 |
| Not in $IPR_k$ | 0 | $\phi_{FN}$ | 1 |
| In $IPR_k$ | 0 | $\phi_{TN}$ | 0 |
| In $IPR_k$ | 0 | $\phi_{FN}$ | 1 |

TABLE II

CONDITIONAL PROBABILITY TABLE FOR DETECTION BOXES $B_k^m$

| $r_k$ | $L_k$ | $P(B_k^m \mid r_k, L_k)$ |
|---|---|---|
| $\in IPR_k$ | $m$ | $\mathcal{N}(H(x_k) - B_k^m, \Sigma)$ |
| $\in IPR_k$ | $n \in L \ \& \ n \neq m$ | $\mathcal{U}(image)$ |
| $\notin IPR_k$ | $m$ | 0 |
| $\notin IPR_k$ | $n \in L \ \& \ n \ \neq m$ | $\mathcal{U}(image)$ |

the area of the image frame is $A$, the value associated with $\mathcal{U}(image)$ is $1/A$.

### B. Particle filter approximation

To localize and track the target, we must obtain the posterior PDF of the RF emitter target state given the data from the RF sensor and OOIs detected in the vision sensor images. From the probabilistic graphical model and the conditional dependencies described in Tables I and II, we can see that finding the desired target state filtering posterior requires solving an analytically intractable inference problem in order to fuse information from multiple heterogeneous data sources. Specifically, the posterior uncertainties for $L_k$ and target states are conditionally dependent given the RF and OOI detector data, and thus cannot be treated in isolation. In addition to the fact that the sensor models are heavily non-linear and non-Gaussian, the necessity of further extracting negative information from 'no detection' RF and vision data (to aid in target acquisition when it goes out of either visual or RF sensing range) makes this a hard estimation problem [12].

An approximate inference algorithm must therefore be applied, and a Sequential Monte Carlo Method is developed here for its feasiblity of onboard implementation for a sUAS and its ability to produce online state estimates for complex non-Gaussian uncertainties [3] [4]. A particle filter is used to perform state estimation for the augmented state with position, velocity and the latent emitter association variable. Consider a single target where the state of the $i^{th}$ particle at time $k$ is given by

$$X_k^i = \begin{bmatrix} r_k^i \\ \dot{r}_k^i \\ L \end{bmatrix} \tag{10}$$

The particle filter is a recursive Bayesian filter implementation based on Monte Carlo importance sampling simulations. The idea is to represent the posterior probability density function by a set of random samples with associated importance sampling weights, which allows for computation of state estimates based on these samples and weights. The posterior density at time $k$ can be approximated in particle form as:

$$p(X_{0:k} \mid Z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(X_{0:k} - X_{0:k}^i), \tag{11}$$

where $\delta(.)$ is the Dirac delta measure. It can be shown that as $N_s \to \infty$, the approximation allows us to approach the true posterior density $p(X_k \mid Z_{1:k})$ [2]

In this work, we assume a prior of the form $p(X_0) = p(r_0)p(\dot{r}_0)p(L_0)$, where $p(r_0) = \mathcal{U}(\mathbb{A})$, $p(\dot{r}_0) = \mathcal{U}[-V_{pred}, V_{pred}]$ and $p(L_0) = \mathcal{U}[0, n_k + 1]$. For $k \geq 1$,as RF and OOI detection observations become available, the samples are drawn from an importance density, $q(X_k \mid X_{0:k-1}, Z_{1:k})$. The particles are then weighted by the ratio of the importance density to the posterior distribution, which simplifies to,

$$w_k^i = w_{k-1}^i \times \frac{p(Z_k \mid X_k^i) p(X_k^i \mid X_{k-1}^i)}{q(X_k^i \mid X_{0:k-1}^i, Z_{1:k})} \tag{12}$$

where $Z_k = \{z_k; O_k\}$ is the set of all available observations, and $z_k$ and $O_k$ measurements arrive at the same rate. For the sequential importance resampling (SIR) 'bootstrap' particle filter implementation, the importance sampling distribution is picked such that $q(X_k^i \mid X_{0:k-1}^i, Z_{1:k}) = p(X_k^i \mid X_{k-1}^i)$ [2], and so the weight update becomes

$$w_k^i = w_{k-1}^i \times p(Z_k \mid X_k^i) \tag{13}$$

The full derivation of the RF sensor likelihood $p(z_k \mid x_k^i)$ is documented in detail in our past paper [1]. Since $O_k$ is comprised of $B_k^1 .... B_k^{N_k}$, each particle has to account for all of the detection boxes (observations) that come from the object detection algorithm. Since all detections are independent,

$$p(O_k \mid X_k^i) = \prod_{j=1}^{n_k} P(B_k^j \mid r_k, L_k). \tag{14}$$

The final particle weight update is thus

$$p(Z_k \mid X_k^i) = p(O_k \mid X_k^i) \times p(z_k \mid X_k^i) \tag{15}$$

The particle filter is modified by a sequential localize and track technique to maintain performance with low number of particles as per [1] Section IV, B, 3. Particles are resampled when the expected sample size goes below 65% [2].

## IV. SIMULATION RESULTS AND DISCUSSION

The following section compares the localization and tracking performance of the RF+vision fusion filter to a vision-only particle filter representative of current state-of-the-art technique on simulated data. This comparison demonstrates the advantage of the fusion filter for tracking targets when RF data are available. Performance is also assessed on the fusion filter for various $N_s$ values. This is used to inform the selection of the number of particles to balance solution

accuracy against computational load. Fig. 3 depicts the RF and visual OOI detection sensor models using a snapshot of a typical particle filter run. The simulation was run in a python environment, with models described below.

To generate model-based ground truth simulations, the RF data were simulated according to the sensing model developed in [1]. OOI data with visual clutter were generated according to a Poisson process with a random uniform distribution over the image frame. The measurements from the vision sensor and the RF are synchronized. The target moves from the top to the bottom of the area of interest, with a velocity of 10 m/s, with process noise described in Section II B. The aircraft moves in a lawnmower pattern over the target.

The metrics used for validation and comparison are: 1) Root Mean Square Error (RMSE) for the target states; 2) Association Correctness for the latent emitter association variable (i.e. classification correctness for inferred types of OOI box data); 3) Particle health (effective sample size); and 4) Convergence of filter error estimates, where convergence occurs if (position and velocity) states are within 5 m and 1 m/s of the ground truth value. Each type of filter is run for $N_{MC} = 10$ Monte Carlo (MC) simulations, and the metrics above are calculated for those simulations. The RMSE for time step $k$ is calculated by:

$$RMSE_k = \sqrt{\frac{1}{N_{MC}} \sum_{c=1}^{N_{MC}} R_{k,c}^T R_{k,c}}. \tag{16}$$

Here, $R_{k,c}$ is the error at time step $k$ for MC run $c$. The average association correctness percent (ACP) at time step $k$ is:

$$ACP_k = \frac{1}{N_{MC}} \sum_{c=1}^{N_{MC}} \frac{\sum_{s=1}^{N_s} \delta(L_{k,c}^s, L_{true,k,c})}{N_s} \times 100 \tag{17}$$

The ACP for all time over all MC runs is:

$$ACP = \frac{\sum_{k=1}^{N_t} ACP_k}{N_t} \tag{18}$$

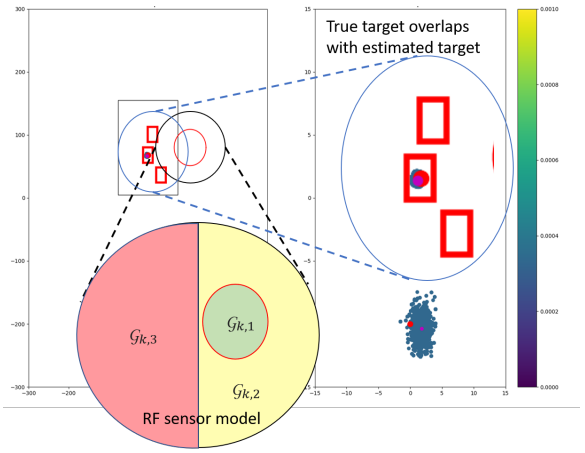


Fig. 3. Particle filter snapshot showing RF and camera data models.

where $L_{true,k}$ is the true association at time step $k$, $N_t$ is the total number of time steps, and $L_k^i$ is the association assigned to the $i^{th}$ particle at time step $k$. The particle health is measured by the effective sample size ($N_{eff}$). The average $N_{eff}$ at time $k$ is:

$$N_{eff,k} = \frac{1}{N_{MC}} \sum_{c=1}^{NMC} \frac{1}{1 + \text{var}(w_{k,c})}, \tag{19}$$

The $N_{eff}$ for all time over all MC runs is:

$$N_{eff} = \frac{\sum_{k=1}^{N_t} N_{eff,k}}{N_t} \tag{20}$$

where $w_k$ is the unnormalized particle weight at time step k.

### A. *Comparison of vision-only and fused vision + RF tracking*

In this section, we compare the RF+vision fusion filter, to the vision-only filter, emulating the current state-of-the-art. Both types of tracking filters are also individually run with perfect association. Perfect association implies that the Emitter Association Variable $L_k$ has the value associated with the true target, at all times. The Perfect Association (PA) implementations of the vision-only filter and fusion (vision+RF) filter shows the degree to which the data association problem impacts filter performance, since the PA filters provide high baselines for the best tracking results obtainable for either sensing modality. All of the filters are run with 1500 particles (the reason for this number of particles is discussed in the next subsection).

Fig 4 shows the RMSE errors in position and velocity for the 4 types of filters in this section. The filter exhibits two types of behavior, converging and diverging, and Table III outlines the convergence rates for these filters. The RMSE position errors for the vision only filter are high and

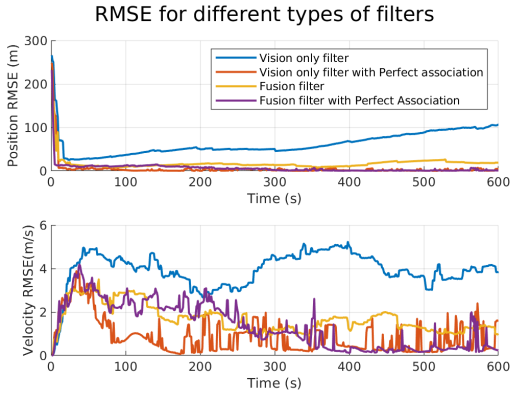


Fig. 4. RMSE error for vision only filters and fusion filters

TABLE III
CONVERGENCE RATE FOR DIFFERENT TYPES OF FILTERS FOR MC SIMULATIONS

| ***Type of filter*** | Vis-only | PA Vis-only | Fusion | PA Fusion |
|---|---|---|---|---|
| ***Percent of sims that converged*** | 40 % | 90% | 90% | 100% |

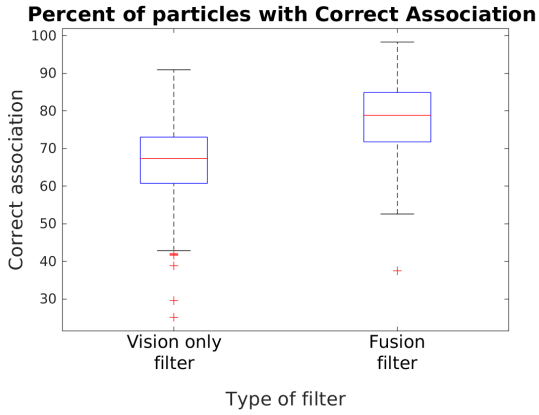


Fig. 5. Correct association comparison for different types of filters

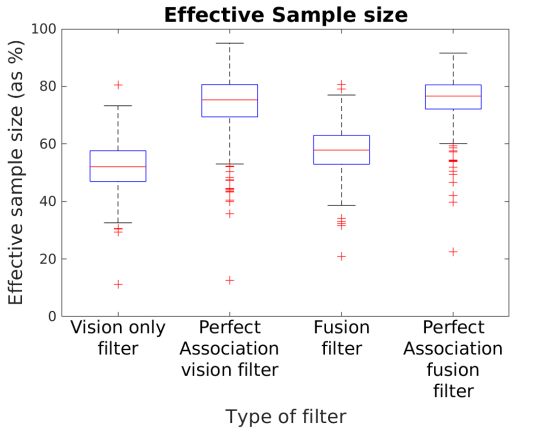


Fig. 6. Particle health for different types of filters

increasing with time, whereas the RMSE position errors for the fusion filter are lower and within 15 m of the baseline. The RMSE velocity error is the highest, consistently in the vision only data filter case. The fusion filter's RMSE velocity errors are comparable and similar to the baseline PA cases. Fig 5 shows that the fusion filter performs better association than the vision only filter, with respect to the median ACP. The vision only filter also has outliers with lower ACPs, further indicating that the fusion filter associates better than the vision only filter. Table III also shows that 90 % of the simulations converged vs. diverged for the fusion filter as compared to 40% for the vision only filter. Fig. 6 shows consistently large $N_{eff}$ values, which demonstrates a healthy, non-collapsing particle filter. The PA cases show some outliers, but 95% of the outliers are above 35 %, which implies low $N_{eff}$ but not degeneracy.

The results show that fusion of RF and vision-based sensor data leads to significant improvements in the ability of the aircraft to precisely localize and thus track moving RF emitter ground targets. The RF sensor provides coarse but consistent positive and negative information that helps narrow down the uncertainty in visual OOI association hypotheses over time. Likewise, the visual OOI detection data provides more precise but sporadic localization information for tracking target states once association hypothesis uncertainties are addressed.

### *B. Assessment of number of particles*

One of the challenges using a particle filter is the number of particles, $N_s$. High $N_s$ values lead to intractable and huge computation loads. In this section, we identify the sensitivity of the filter to particle size, as computation tends to scale with particle size. Fig. 7 shows the RMSE position and velocity errors over time for a fusion filter for different number of particles. Table IV outlines the convergence rates for different $N_s$ values for the fusion filter. We see that the position error for 500 and 1000 particle filters are high and increase over time. However, for a 1500 particle filter the RMSE position error drops over time, and stays within 15 m. The final RMSE error for 1000 particles is higher than the error for 500 particles, but this might be because a larger number of MC simulations are needed to fully capture the filter's behavior. The RMSE velocity plots show that the error is high for both 500 and 1000 particles, but reduces to under 1 m/s in 600 seconds for 1500 particles. We also see that as we increase the number of particles, the filter converges more often, as seen in IV thus reducing the error seen in RMSE plots.

From Fig. 8 we see that the ACP increases with higher $N_s$ values, implying better association with higher number of particles. Fig 9 shows that the median of $N_{eff}$, is about the same for all 3 cases, and greater than 50% implying good particle health.

The higher the number of particles in a particle filter, the better the representation of the posterior distribution. [2]. Multi-dimensional state particle filters need enough particles to sufficiently encompass the different behaviors of each dimension and combinations of state variables. In this case, 1500 particles provided adequate estimation performance for

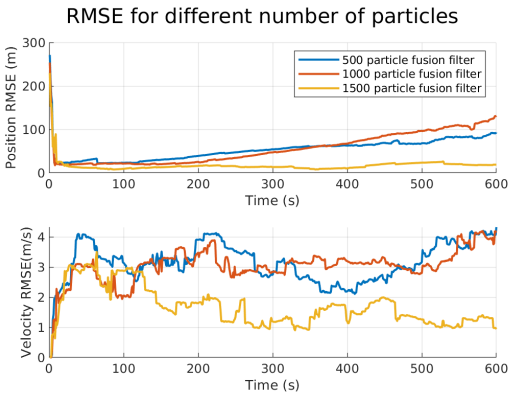


Fig. 7. Comparison of RMSE for the fusion filter with different $N_s$.

TABLE IV
CONVERGENCE RATE FOR MC SIMULATIONS

| ***Number of particles*** | 500 | 1000 | 1500 |
|---|---|---|---|
| ***Percent of sims that converged*** | 50 % | 70% | 90% |

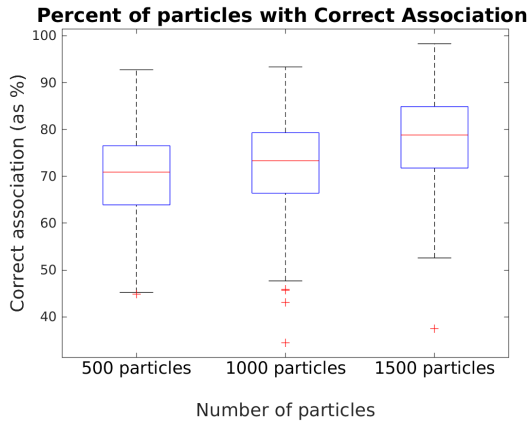


Fig. 8. Correct association comparison for different number of particles

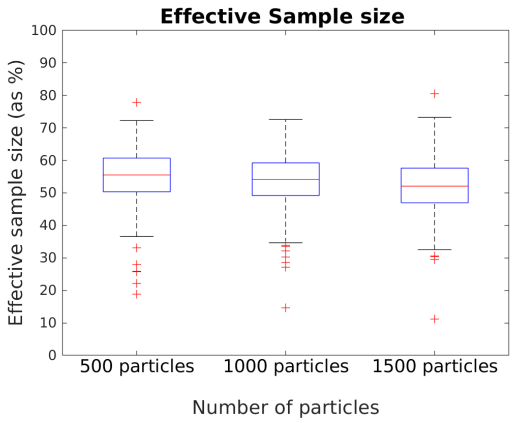


Fig. 9. Particle health for different number of particles

a single target. In order to limit computation costs, for this specific simulation, the number was not increased beyond this point.

## V. Conclusion and Future Work

This paper introduced and validated the concept of a novel RF and visual data fusion filter for aerial tracking of a ground target. Developing reliable and robust estimation architecture is important for target localization and tracking in uncertain scenarios. A Bayesian model and approximate state estimation techniques was developed to simultaneously solve the data association and tracking/localization problem for an RF-emitting target in the presence of unlabeled vision data. Simulations validated the data association algorithm and the utility of implementing a fusion filter as compared to state-of-the-art vision-only tracking filters, and also showed that a reasonable number of particles can be used for online sUAS implementation. Although the exact details for the number of samples to use will depend on the scenario, these results gives us an idea of the order of the number of particles needed to assess onboard sUAS computation costs.

Future work involves examining harder and more unpredictable trajectories and behaviors for the target, including periods of occlusion from sensors. We also are working on onboard sUAS deployment for live flight experiments.

## References


[1] R. K. Rajasekaran and E. Frew, "Assessing particle filter algorithms for tracking radio emitters using small unmanned aircraft," *AIAA Scitech 2019 Forum*, no. January, pp. 1–13, 2019.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/nongaussian bayesian tracking," in *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, 2007.

[3] W. Ng, J. Li, and S. Godsill, "Online multisensor-multitarget detection and tracking," *IEEE Aerospace Conference Proceedings*, vol. 2006, no. October, 2006.

[4] J. Vermaak, S. J. Godsill, and P. Pérez, "Monte Carlo filtering for multi-target tracking and data association," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 1, pp. 309–332, 2005.

[5] V. N. Dobrokhodov, I. I. Kaminer, K. D. Jones, and R. Ghabcheloo, "Vision-based tracking and motion estimation for moving targets using small UAVs," *Proceedings of the American Control Conference*, vol. 2006, pp. 1428–1433, 2006.

[6] R. Opromolla, G. Fasano, and D. Accardo, "A vision-based approach to uav detection and tracking in cooperative applications," *Sensors (Switzerland)*, vol. 18, no. 10, 2018.

[7] J. H. Lee, J. D. Millard, P. C. Lusk, and R. W. Beard, "Autonomous target following with monocular camera on UAS using Recursive-RANSAC tracker," *2018 International Conference on Unmanned Aircraft Systems, ICUAS 2018*, pp. 1070–1074, 2018.

[8] T. Kirubarajan and Y. Bar-Shalom, "Probabilistic data association techniques for target tracking in clutter," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 536–556, 2004.

[9] B. Zhou and N. K. 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.

[10] A. Al-Kaff, M. J. Gómez-Silva, F. Miguel Moreno, A. De La Escalera, and J. M. Armingol, "An appearance-based tracking algorithm for aerial search and rescue purposes," *Sensors (Switzerland)*, vol. 19, no. 3, 2019.

[11] M. H. Lee and S. Yeom, "Detection and tracking of multiple moving vehicles with a UAV," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 18, no. 3, pp. 182–189, 2018.

[12] K. Wyffels and M. Campbell, "Negative observations for multiple hypothesis tracking of dynamic extended objects," *Proceedings of the American Control Conference*, pp. 642–647, 2014.

[13] I. Miller, M. Campbell, and D. Huttenlocher, "Obstacles Under Large Viewpoint Changes," vol. 27, no. 1, pp. 29–46, 2011.

[14] N. Ahmed, D. Casbeer, Y. Cao, and D. Kingston, "Multitarget localization on road networks with hidden Markov Rao-Blackwellized particle filters," *Journal of Aerospace Information Systems*, vol. 14, no. 11, pp. 573–596, 2017.

[15] J. H. White, K. T. Salva, and R. W. Beard, "Extending Motion Detection to Track Stopped Objects in Visual Multi-Target Tracking," *Proceedings of the American Control Conference*, vol. 2018-June, pp. 5825–5830, 2018.

[16] P. C. Niedfeldt, K. Ingersoll, and R. W. Beard, "Comparison and Analysis of Recursive-RANSAC for Multiple Target Tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, pp. 461–476, 2017.

[17] T. Bailey, B. Upcroft, and H. Durrant-whyte, "Validation Gating for Non-Linear Non-Gaussian Target Tracking," *2006 Int'l Conf. on Information Fusion (FUSION 2006)*, pp. 1–6, 2006.

[18] C. C. Bidstrup, J. J. Moore, C. K. Peterson, and R. W. Beard, "Tracking multiple vehicles constrained to a road network from a UAV with sparse visual measurements," *Proceedings of the American Control Conference*, vol. 2019-July, pp. 3817–3822, 2019.

[19] T. Germa, F. Lerasle, N. Ouadah, and V. Cadenat, "Vision and RFID data fusion for tracking people in crowds by a mobile robot," *Computer Vision and Image Understanding*, vol. 114, no. 6, pp. 641–651, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2010.01.008

[20] T. Miyaki, T. Yamasaki, and K. Aizawa, "Multi-sensor fusion tracking using visual information and Wi-Fi location estimation," *2007 1st ACM/IEEE International Conference on Distributed Smart Cameras, ICDSC*, pp. 275–282, 2007.