Ground to air testing of a fused optical-radar aircraft detection and tracking system

Chester V. Dolph[1]
NASA Langley Research Center, Hampton, Virginia 23681

Thomas Lombaerts[2], Evan Kawamura[3], Corey Ippolito[4], Vahram Stepanyan[5]
NASA Ames Research Center, Moffett Field, CA, 94035

Khan M. Iftekharuddin[6]
Old Dominion University, Norfolk, Virginia 23529

George Szatkowski[7], Robert G. McSwain[8], Chris Morris[9], Mahyar R. Malekpour[10]
NASA Langley Research Center, Hampton, Virginia 23681

Cyrus Minwalla[11]
Scoped Systems, Ottawa, Ontario K2G 7A5, Canada

Onboard detection and tracking capability are integral to the sensing component in collision avoidance systems needed to safely operate autonomous urban air taxis and small Unmanned Aircraft Systems. Ground-based validation of detection and tracking systems is an important milestone towards the end goal of real-time collision avoidance using onboard sensors and algorithms. In this work, we evaluate three Extended Kalman Filter (EKF) based fusion trackers with radar and vision detection inputs, and compare them with baseline trackers for each sensor type. Performance is assessed using field collected data of ground to air test flights with the sensors co-located on a stable platform with an instrumented multirotor acting as the intruder performing a waypoint pattern at a distance of 1.1 km to 0.3 km to simulate a head-on collision geometry. Fusing an image-based morphological detector with a radar detector using an EKF covered 74% of the ground truth position updates logged by the flight controller on the multirotor within an error of 50 meters, after accounting for alignment offsets, while covering 15% more ground truth updates relative to radar only. Removing timestamps when the intruder aircraft is occluded by trees and only considering timestamps where the radar has an update, the EKF image-based morphological detector combined with the radar detector covered 90% of the ground truth position updates within an error of 50 meters and 97% within an error of 100 meters.

I. Introduction

Onboard autonomy is needed to enable safe, widespread operations for Urban Air Mobility and small Unmanned Aerial Systems (sUAS) [1]. Despite decades of research, an aviation authority has yet to approve an onboard collision avoidance system for widespread usage. The FAA and RTCA have proposed Minimum Operation Performance Standards (MOPs) to guide development of onboard collision avoidance systems. The MOPs include metrics on

[1] Aerospace Engineer, Aeronautics Systems Engineering Branch, AIAA Member
[2] Aerospace Research Engineer, KBR Wyle Services, Intelligent Systems Division, AIAA Associate Fellow
[3] Aerospace Scientist
[4] Aerospace Scientist
[5] ESOF40-Technical Advisor, KBR Wyle Services, Intelligent Systems Division
[6] Principal Scientist
[7] Associate Dean for Graduate Studies and Graduate Programs
[8] Electrical Engineer, Safety Critical Avionics Branch
[9] Engineer Level IV, Safety Critical Avionics Branch
[10] Computer Engineer, Safety Critical Avionics Branch
[11] Principal, Scoped Systems

evaluating detecting capability based on range. Most research on sensor-based detection and tracking of aircraft does not provide a baseline performance metric using the FAA/RTCA MOPs. Further, research often limits the analysis to one sensor, making it difficult to compare different modalities: infrared, vision, acoustic, or radar. Additionally, data is often not publicly available to support innovation by research and development industries. This work analyzes radar and vision sensors using a ground-to-air test with the contributions explained in detail in the following contribution sections.

While aviation applications need to sense aerial objects at greater distances to maintain proposed MOPs, both autonomous ground vehicles and aircraft rely on extracting data from sensors to avoid non-cooperative objects. A survey of different sensor modalities for automobile detection in on-road and off-road environments in [2] indicates that a single sensor may not be sufficient for diverse operational environments, low visibility or weather degraded conditions, or hazard detections, such as puddles in roads or pedestrians in crosswalks. Requirements for autonomous vehicle sensor systems and algorithms remain an active area of research, especially given the fatalities as a result of sensor fusion systems incorrectly perceiving their environments [3] [4]. Automobile and aviation autonomous systems share the need to understand their operational environments and to avoid objects and hazards.

In [5] a fused optical-radar system mounted to a General Aviation (GA) aircraft for detection and tracking of a GA intruder aircraft on a head-on collision geometry using an Extended Kalman Filter (EKF) showed an angular accuracy of ~2.2 degrees for azimuth and 0.6º elevation over ranges 2 to 4km. The work in [6] fused different types of lidar and ultrasonic sensors for detecting objects from an ownship multirotor at distances up to 4m with sub meter range error, however, this does not address the need for collision avoidance of moving aircraft, which requires greater detection distances to meet proposed MOPs. The work in [7] field tested a real-time fusion method using Millimeter wave (MMW) radar and vision on a multirotor with a collision avoidance algorithm combined with path planning, without specifying minimum separation distances for static objects in an uncluttered environment or reporting detailed system performance metrics. The survey in [8] compared different sensing modalities for ground-based detection of sUAS and found: 1) consensus that radar has strong ranging capability but suffers in the presence of clutter, especially at low altitudes, 2) vision provides perception capability with deep networks using rich intensity data, 3) thermal sensors provide low resolution data, and 4) acoustic sensors are subject to noise from propellers and surroundings.

This work compares a vision tracking system, a radar tracking system, a radar EKF tracking system, and three fused radar-vision EKF trackers for aerial objects at distances from 300 to 1100m. A technical analysis of the fused system is performed and contrasted with the vision and radar sensors individually. Comparing the fused system with the vision and radar sensors individually increases understanding of the benefits of the combined system with respect to the individual sensor capabilities. Vision sensors provide strong resolution accuracy of pixel coordinates (i.e. x,y), which can be converted to azimuth and elevation angles, and radar sensors provide strong ranging capabilities. This paper is organized as follows: Techniques, Tracker Evaluation, Experimental Methodology, Detection and Measurement Results, Tracker Analysis, and finally Conclusion.

II. Techniques

The vision detection pipeline employed here is described in [9]. For this work, data below the tree-line is masked using a custom image filter. Prior work has shown that the rich texture of the terrain in the imagery below the horizon causes substantial activation of the morphological detector that utilizes a crosshair 2D convolutional kernel to detect objects in the image and thus saturates the image detection system. After segmenting the above horizon area, two types of image-based detectors are employed: morphological and image differencing. The morphological detector utilizes the close-minus-open extraction technique with the convolutional kernel and is followed by a FAST feature extraction [10] technique to determine objects of interest. The image-differencing detector steps include contrast enhancement, image stabilization across frames using the Lucas Kanade optical flow technique [11] , subtraction of current frame from the previous frame, and FAST feature extraction on the differenced image. The vision detection system provides localizations of the aircraft in terms of pixel x and pixel y, which are converted to azimuth and elevation angles after performing a camera calibration and using the transformations detailed later in this section. A linear Kalman filter generates velocity and position estimates from the detection data for the vision tracker.

The radar generates a Range Doppler Map (RV Maps) for each beam angle. RV Maps consist of the target's amplitude in relation to its range distance and radial velocity. The Constant False Alarm Rate algorithm detects probable targets after processing each RV Map collected by the radar. The time, relative location, radial-velocity, signal-to-noise-ratio, and power of each detection are logged and can be reported to the user. The relative location is recorded as (azimuth, elevation, range) in radar coordinates as well as (X, Y, Z) antenna coordinates. Next, the radar software implements a linear Kalman filter that produces position and velocity estimates.

When available, the radar supplies object location information in 3D by means of range, azimuth, and elevation. These measurements are disturbed by radar noise. For vision, a single camera does not provide depth perception with high fidelity and only provides azimuth and elevation of the detected object, which are also perturbed by camera noise. In general, radar noise is significantly larger than camera noise, especially for elevation. On the other hand, the camera observations lack range information, so it cannot provide a full 3D localization. Therefore, sensor fusion aims to combine both measurements and eliminate, or at least reduce as much as possible, the sensor's inherent disturbances such as the aforementioned noise.

A. Vision System

This section describes the hardware used in the vision system, camera calibration technique and analysis, and camera model for establishing the camera reference frame.

The camera and lens configuration were chosen with consideration to meet proposed performance standards for aircraft detection requirements to minimize collision risk described in [12] and [13] for sUAS and GA aircraft while considering typical flight operations and performance envelopes of an ownship multirotor. Table 1 shows the vision system hardware specifications. The camera aligns with the radar and MEMS-based INS using an aluminum bracket with milled aluminum mounting holes within a tolerance of ± 0.005 in, and the camera's orientation is in the same direction as the radar. The camera calibration in this work was completed using the Bouget Toolbox [14] using a 7 by 10 square checkerboard pattern affixed to a milled aluminum sheet where each square's measured dimension is 88.2 by 88.2 mm. The calibration used seven images capturing head-on, rolled left and right, pitched up and down, and yawed left and right as shown by the grid planes in the extrinsic parameters in Figure 1. The theoretical principal point has the coordinates 2048, 1500 and is computed by dividing the sensor resolution by two, while the calibration revealed a pixel center of 2070.67, 1458.31 ± 15.098, 10.72. The distortion model is shown in Figure 2 such that each arrow shows the displacement of the pixel from radial and tangential lens distortion. The reprojection error in pixels is shown in Figure 3 and is bounded by -1.5 to 1.5 pixel in both dimensions. The standard deviation of the pixel error is 0.550, 0.520 pixels.

Table 1: Camera and Lens Specification

Camera Feature	Specification
Pixel resolution	4096 x 3000 pixels
Lens	16 mm
Horizontal Field of View x Vertical Field of View	32.9 x 24.8°
Pixel Size Horizontal x Vertical	3.45 x 3.45 µm
Frames per second	~30 Hz

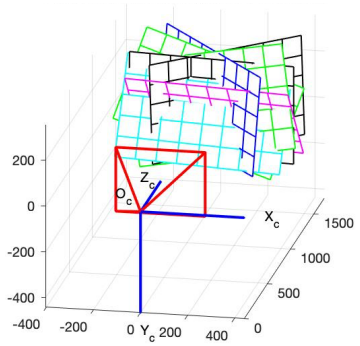


Figure 1: Extrinsic Parameters (camera-centered)

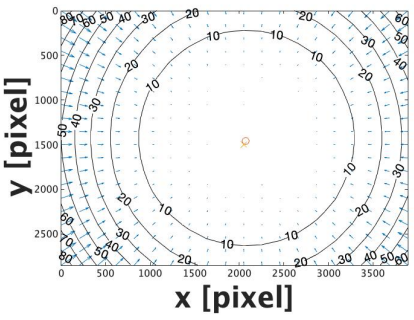


Figure 2: Complete Distortion Model

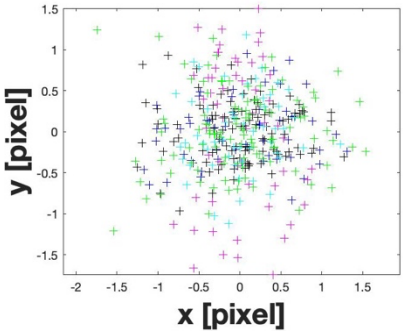


Figure 3: Reprojection Error

After camera calibration, the pixel coordinates are corrected for radial and tangential distortion:

$$\begin{bmatrix} x_{corrected} \\ y_{corrected} \end{bmatrix} = (1 + k_1 r^2 + k_2 r^2 + k_5 r^2) \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2k_1 xy + k_4(r^2 + 2x^2) \\ k_3(r^2 + 2y^2) + 2k_4 xy \end{bmatrix} \quad (1)$$

where k_i is the tangential and radial distortion coefficients with $i = 1,\ldots,5$, $x = (u - c_x)/f_1$, $y = (u - c_y)/f_2$, and $r^2 = x^2 + y^2$. The k_2 and k_5 distortion coefficients are found to be within their uncertainties during camera calibration and thus equation [1] is rewritten as:

$$\begin{bmatrix} x_{corrected} \\ y_{corrected} \end{bmatrix} = (1 + k_1 r^2) \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2k_1 xy + k_4(r^2 + 2x^2) \\ k_3(r^2 + 2y^2) + 2k_4 y \end{bmatrix} \quad (2)$$

After correcting for distortion, the perspective camera model is used for relating the pixel coordinate to the camera coordinate system as shown in Figure 4.

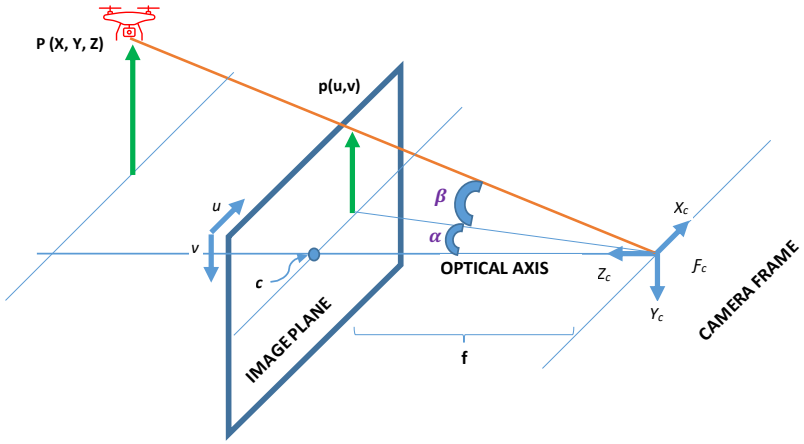


Figure 4: Perspective Camera Model

The 3D coordinates are defined with respect to the camera reference frame F_c. The optical axis is perpendicular to the camera reference frame. The distance between the camera reference and the image planes is the focal length, f. Thus the image of the point $\mathbf{P} = [\, X, Y, Z \,]^T$ is given by $\mathbf{p} = [\, x, y, z \,]^T$ in the camera frame.

The azimuth angle, α, and elevation angle, β, shown in Figure 4 are computed by converting the image frame to camera frame to sensor coordinates using physical sensor properties and image frame coordinates [15] as derived here:

$$d_x = \frac{length\ of\ sensor\ in\ x\ dimension}{sensor\ resolution\ x\ pixel} \quad (3)$$

$$d_y = \frac{length\ of\ sensor\ in\ y\ dimension}{sensor\ resolution\ \ y\ pixels} \quad (4)$$

$$x_{corrected} = \frac{x_p}{d_x} + c_x \quad (5)$$

$$y_{corrected} = \frac{y_p}{d_y} + c_y \quad (6)$$

$$x_p = (x_{corrected} - c_x) d_x \quad (7)$$

$$y_p = (y_{corrected} - c_y) d_y \quad (8)$$

$$\alpha = arctan\left(\frac{x_p}{F}\right) \quad (9)$$

$$\beta = arctan\left(\frac{y_p}{F}\right) \tag{10}$$

Using the focal length computed during camera calibration and the physical hardware specifications:

$$F = f\frac{W}{w} = 4741.74\ pixels\ * \frac{14.13mm}{4096\ pixels} = 16.45\ mm \tag{11}$$

For the purposes of understanding vision system performance when provided range and systematic offsets for transformation from sensor coordinates to the INS coordinate frame, the image coordinates are converted to **P** = [*X, Y, Z*]^T by using Z_{fc} after transforming the multirotor position using the GPS position recorded from the flight controller (*fc*) to the camera reference frame. The transformation of the multirotor's GPS coordinates recorded by the flight controller is derived in section II.E. Thus, X and Y in the camera frame for camera are:

$$x_p = F\frac{X}{Z_{fc}} \tag{12}$$

$$X = \frac{x_p * Z_{fc}}{F} \tag{13}$$

$$y_p = F\frac{Y}{Z_{fc}} \tag{14}$$

$$Y = \frac{y_p * Z_{fc}}{F} \tag{15}$$

B. Radar System

The radar selected for this research was a Frequency Modulated Continuous Wave (FMCW) radar possessing a field of view of 120 degrees in azimuth and 80 degrees in elevation and operating at 24.55GHz center frequency with a 45 MHz bandwidth. The radar transmits 2 watts of power through a Metamaterial Electronically Scanning Array antenna oriented in horizontal polarization. When the radar is transmitting, personnel must stay at least 1 meter away from the active array to limit nonionizing radiation exposure. The radar's physical dimensions are 18.7cm x 12.1cm x 4.1cm and it weighs less than 820 grams, making it well suited for installation on small UASs. The radar has independent transmit and receive antenna arrays separated by a small metal divider to reduce cross talk coupling.

The ground tests were conducted with the radar field of view configured to scan 60 degrees in azimuth and 30 degrees in elevation. To reduce the potential for false tracks, the masking commands were set to only monitor targets with radial velocities between +/- 0.91 meters per second. To ensure new tracks were identified as quickly as possible, the minimum reporting confidence level was set to 30 (out of 100). All other radar command settings were placed at the recommended default values to track airborne UAS.

The radar uses a right handed Cartesian coordinate system that is fixed to the face of the antenna to determine where antenna radiation is pointed relative to the radar aperture. The Z axis is the broadside direction normal to the face of the radar. The Y axis is the vertical direction in the plane of the radar aperture. The X axis is the horizontal direction and points from right to left in the plane of the antenna face. Azimuth is defined as the angle measured from the Z axis toward the X axis. The elevation angle is defined as the angle between the pointing vector and the XZ plane.

The radar is controlled using a command line interface API software through an Ethernet TCP protocol. Scripting commands are used to establish the radar's operational characteristics including the search and track fields of view, parameters to govern the tracking Kalman filter, range and velocity masking options to help reduce false tracks, and data logging selections. Radar detections are generated every time the radar is able to detect a Doppler target. Detections may or may not turn into radar tracks depending on the radar masking settings defined in the startup script command. The radar track and detection data packets are recorded in binary files. By default, detection packets are transmitted at a rate of every beam step period, 7-8 ms, while track data packets by default are transmitted at 4.98 Hz which gives about 210 ms between every update. The user cannot modify the detection transmit rate, but the track data transmit rate can be configured to transmit at the user specified track update rate in the range [0.0, 10.0] per second.

C. EKF Tracker Implementation

The object tracker is a conventional Extended Kalman Filter defined in detail in [16]. The filter gain calculation is modified to incorporate valid flags for the different sensor measurements, which makes this method adaptive and robust against invalid data, loss of signal, outliers and lower sampling rates. The current implementation uses measurements from a camera and a radar, but this modular architecture is simple to extend such that other sensors can be incorporated as well. Figure 5 illustrates the general overview of the Kalman Filter structure. The prediction step assumes constant velocity for calculating the predicted state $\hat{\boldsymbol{x}}_{\boldsymbol{k}}$ based on the previous state $\hat{\boldsymbol{x}}_{\boldsymbol{k-1}}$. The state consists of six elements, namely the three position coordinates and the three velocity coordinates, all in ENU coordinates (East-North-Up). A predicted measurement $\hat{\boldsymbol{z}}_{\boldsymbol{k}}$ is calculated, which is a function of the predicted state $\boldsymbol{h}(\hat{\boldsymbol{x}}_{\boldsymbol{k}})$. This predicted measurement consists of the measured variables that are provided by both sensors, namely range, azimuth and elevation from the radar as well as azimuth and elevation from the camera. By comparing the actual measurement $\boldsymbol{z}_{\boldsymbol{k}}$ with the predicted measurement $\hat{\boldsymbol{z}}_{\boldsymbol{k}}$, the innovation is computed as $\boldsymbol{\epsilon} = \boldsymbol{z}_{\boldsymbol{k}} - \hat{\boldsymbol{z}}_{\boldsymbol{k}} = \boldsymbol{z}_{\boldsymbol{k}} - \boldsymbol{h}(\hat{\boldsymbol{x}}_{\boldsymbol{k}})$. The main drivers for the innovation are errors in the predicted state (due to the constant speed assumption) and sensor disturbances. Finally, the Kalman gain matrix **W** is calculated by means of the predicted covariance $\hat{\boldsymbol{P}}_k$, measurement matrix $\boldsymbol{H} = \frac{\boldsymbol{dh(x)}}{\boldsymbol{dx}}$, and the innovation covariance matrix **S**. This Kalman gain matrix **W** is weighted with valid flags to deal appropriately with the sensor measurements as explained previously. The corrected state $\boldsymbol{x}_{\boldsymbol{k}}$ is calculated by means of predicted state $\hat{\boldsymbol{x}}_{\boldsymbol{k}}$, the innovation ϵ and the Kalman gain matrix **W** via the following equation: $x_k = \hat{\boldsymbol{x}}_{\boldsymbol{k}} + \mathbf{W}\,\boldsymbol{\epsilon}$. Finally, the corrected covariance $\boldsymbol{P}_{\boldsymbol{k}}$ is calculated.

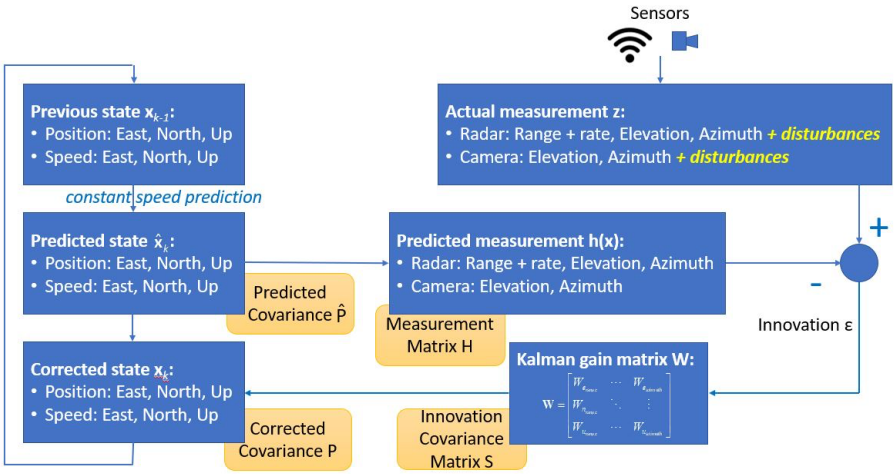


Figure 5: General Overview of the Kalman Filter setup

The information used by the Kalman filter is the perturbed sensor measurements from the radar and camera as well as the previous state. The output of the Kalman Filter is the estimated states of the tracked object, consisting of 3D position and speed in the ENU reference frame. The standard deviations of the estimates are calculated as well. Additional information needed by the Kalman Filter is the typical standard deviations for the sensors, which represent the expected level of disturbance for the sensors. For the sensor hardware used in this study, the corresponding standard deviations are given in Table 2. These values illustrate that the radar measurements are noisier when compared to the camera.

Table 2: Standard deviations for the different sensors

Sensor	Radar	Camera
Range	30 m	–
Azimuth	1 deg	0.8 deg
Elevation	1.5 deg	0.8 deg

D. Transformations to the INS Frame from Vision, radar, and multirotor coordinate system

The INS, radar, and camera each have a defined coordinate system. Mounting the INS at the center of gravity of an sUAS enables minimum transformations for extracting the Euler angles of the aircraft's position. As the long-term effort for this work is onboard detection and tracking for autonomous avoidance of other aircraft, the radar and vision sensors are transformed to the fixed INS coordinate frame. The INS reference frame is defined using North East Down (NED) coordinate system. The transformation from the Camera Reference Frame (CRF) to INS Reference Frame (IRF) NED, $M_{ccf2irf}$, may be written as:

$$INS_{NED} = M_{crf2irf} CRF \tag{16}$$

where INS_{NED} are vision detection or tracks coordinates in the INS reference frame and CRF is the corresponding vision detection and tracks. Expanding $M_{crf2irf}$ to show the axis rotation and translation between camera and the INS:

$$\begin{bmatrix} X_{IRF} \\ Y_{IRF} \\ Z_{IRF} \\ 1 \end{bmatrix}_{NED} = \begin{bmatrix} 0 & 0 & 1 & T_{x,crf2irf} \\ 1 & 0 & 0 & T_{y,crf2irf} \\ 0 & 1 & 0 & T_{z,crf2irf} \\ 0 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} X_{CRF} \\ Y_{CRF} \\ Z_{CRF} \\ 1 \end{bmatrix} \tag{17}$$

where $T_{crf2irf}$ is the translation between the camera to the INS. Similarly, the transformation from the radar reference frame (RRF) to the INS reference frame, $M_{rrf2ins}$, may be computed:

$$INS_{NED} = M_{rrf2ins} RRF \tag{18}$$

and expanded:

$$\begin{bmatrix} X_{INS} \\ Y_{INS} \\ Z_{INS} \\ 1 \end{bmatrix}_{NED} = \begin{bmatrix} 0 & 0 & 1 & T_{x,rrf2irf} \\ -1 & 0 & 0 & T_{y,rrf2irf} \\ 0 & -1 & 0 & T_{z,rrf2irf} \\ 0 & 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} X_{RRF} \\ Y_{RRF} \\ Z_{RRF} \\ 1 \end{bmatrix} \tag{19}$$

where $T_{rrf2irf}$ is the translation between the radar to the INS reference frame.

E. Flight controller WGS84 to INS

The steps to transform the GPS coordinates logged by the multirotor flight controller are shown in Figure 6.

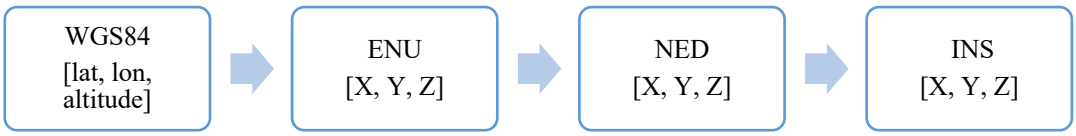


Figure 6: Transforming Multirotor from GPS to INS Reference Frame

The conversion from WGS84 to ENU as derived in [17] is shown in equations (20) and (21):

$$N = R_m(Multirotor_{lat} - INS_{lat}) \tag{20}$$

$$E = R_n \cos(INS_{lat}) \, (Multirotor_{lon} - INS_{lon}) \tag{21}$$

such that N is the North coordinate in meters, and E is the East coordinate in meters. The radius of curvature in the meridian is R_m and the radius of curvature in the prime meridian is R_n, which are defined as:

$$R_n = \frac{R_{EARTH}}{\sqrt{1-\chi\Gamma}} \tag{22}$$

$$R_m = \frac{R_n(1-\Gamma)}{1-\chi\Gamma} \tag{23}$$

with Γ, χ, Earth's inverse flattening (EF), Earth's radius (R_{EARTH}) defined as:

$$\Gamma = \sin^2(INS_{lon}) \tag{24}$$

$$\chi = 2EF - EF^2 \tag{25}$$

$$EF = \frac{1}{298.257223563}, R_{EARTH} = 20925646.3ft \tag{26}$$

After converting latitude and longitude into North and East and using the altitude as the Up coordinate, transforming them into INS (NED) yields:

$$INS_{NED} = M_{ned2ins} M_{enu2ned} WCS_{enu} \tag{27}$$

$$\begin{bmatrix} X_{INS} \\ Y_{INS} \\ Z_{INS} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta\cos\psi & \sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi & T_{x,ned2ins} \\ \cos\theta\sin\psi & \sin\phi\sin\theta\sin\psi & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi & T_{y,ned2ins} \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta & T_{z,ned2ins} \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_{wcs} \\ Y_{wcs} \\ Z_{wcs} \\ 1 \end{bmatrix}_{enu} \tag{28}$$

where $T_{ned2ins}$ is the translation between the ground marker beneath the tripod to the location of the INS, and ϕ, θ, ψ are the roll, pitch, and yaw angles, respectively. Roll and pitch are extracted from the INS rigidly mounted near the optical and radar sensors.

F. $M_{alignment}$

Visualizing the radar and vision results in the INS reference frame revealed similarities in positions, however, there was an approximate 25m altitude offset with a slight angular offset between the sensor data and the ground truth flight controller data. Prior work in [9] showed a ~10 pixel error from ground truth for the vision tracking algorithm, which would translate to ~0.01 degrees for the hardware used in this work. Future work may include examining the altitude offset from the experimental data. Since the radar and vision offsets are similar, the local avoidance framework would be able to track and plan an avoidance maneuver using local sensor information. Standard GPS estimates have a vertical error of $\pm$5m 95% of the time [18]. Finally, proposed MOPs for sUAS collision avoidance of other sUAS and GA aircraft have much larger separation requirements relative to this altitude offset. For instance, in [13] a vertical accuracy between 50 to 100 ft (15.2 to 30.5m) would receive a score of 3 out of 5 for avoidance capability. Thus, converting the local sensor tracking coordinates to GPS coordinates with an altitude offset and communicating to a larger distributed network would still be useful for identifying non-cooperative aircraft because this alignment offset is within the realm of feasibility for proposed MOPs. As this work aims to characterize the benefits of fusing radar and vision sensors, the individual sensors are aligned using the Coherent Point Drift (CPD) algorithm. CPD extracts the transformation matrix from the sensor tracking coordinates to the ground truth flight controller in the INS reference frame [19]. The registration procedure is shown in Figure 7.

Figure 7: Alignment Transformation

The first step to correct for alignment offset is to convert the sensor and ground truth data to the INS reference frame as described in the previous sections. Since the radar and vision systems produce tracks from birds, other air traffic from the adjacent airport, and returns from other objects, the sensor data is filtered to only include data within 125 meters of the ground truth track time update. Next, the rigid implementation of the CPD algorithm is used to register radar and vision data to the ground truth data. Finally, the detection and track data from radar and vision is corrected using the radar or vision transformation matrices.

The vision detection and tracking algorithms used in this work do not provide depth. Therefore, the corrected INS coordinates are transformed back to the camera coordinate systems by rewriting equations (12) to (15) to include the systematic correction:

$$X_{corrected} = \frac{x_{pcorrected} * Z_{corrected}}{F} \tag{29}$$

$$\frac{X_{corrected} * F}{Z_{corrected}} = x_{pcorrected} \tag{30}$$

$$Y_{corrected} = \frac{y_{pcorrected} * Z_{corrected}}{F} \tag{31}$$

$$\frac{Y_{corrected} * F}{Z_{corrected}} = y_{pcorrected} \tag{32}$$

In this way the impact of Z_{fc} is negated and the azimuth and elevation angles may be computed using equations (9) and (10).

G. INS to WGS84

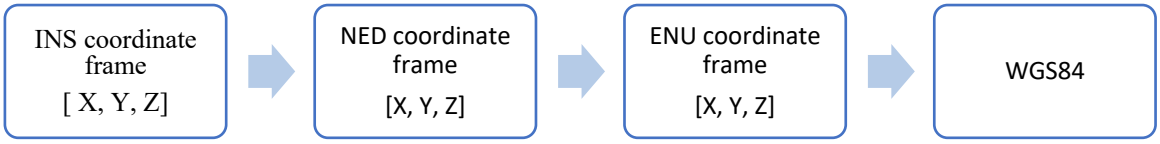


Figure 8: INS Reference Frame to WGS84 Transformation Steps.

After transforming the radar measurements to WCS (ENU), they are converted into latitude and longitude based on some known origin's latitude and longitude in radians, i.e., $latOr_{rad}$ & $lonOr_{rad}$. The WGS84 procedure for converting WCS (ENU) into latitude and longitude starts by computing parameters Γ, χ, R_m, R_n, from Earth's inverse flattening (EF), Earth's radius in ft ($Earth_{rad,ft}$), and the origin's longitude (see Eqns. 22-26). Finally, the latitude and longitude are computed:

$$lat_{rad} = \frac{N_{ft}}{R_m} + latOr_{rad} \tag{33}$$

$$lon_{rad} = \frac{E_{ft}}{R_n \cos(lat_{rad})} + lonOr_{rad} \ . \quad (34)$$

III. Tracker Evaluation

The trackers are evaluated by two metrics: localization error and ground truth trajectory coverage. The localization error, L_{error}, is the Euclidean distance between multirotor position estimate and the tracker position estimate in the INS reference frame at flight controller time, $fc(t)$, and the closest tracker time, $tracker(t)$:

$$L_{error,fc}(t) \approx \sqrt{\left(X_{fc} - X_{tracker}\right)^2 + \left(Y_{fc} - Y_{tracker}\right)^2 + \left(Z_{fc(t)} - Z_{tracker}\right)^2}. \quad (35)$$

A system time offset constant is applied to each sensor modality to synchronize the data to the flight controller time series. For error computation, given the differing sample rates of the flight controller (5 Hz), radar (9 Hz), and cameras (~30 Hz), a match between timestamps is considered to be within 1/(2*sample rate of the sensor) such that the flight controller time update may never have more than one sensor measurement from the same detector associated with a given timestamp.

A tracker update is defined as hit if $L_{error,fc}(t) < t_{error}$, where t_{error} is the error threshold. This work utilizes t_{error} values of 50, 100, and 200 m. Ground coverage, $GT_{coverage,t_{error}}$, is defined as the number of hits, N_{Hits}, divided by the number of ground truth updates, $N_{fc(t)}$:

$$GT_{coverage,t_{error}} = \frac{N_{Hits}}{N_{fc(t)}} * 100 \quad (36)$$

IV. Experimental Methodology

Data collection occurred on October 28th, 2021 at NASA Langley in Hampton, Virginia. Figure 9 shows the flight pattern (blue line) of the intruder flying at altitudes of 107 m and 91 m for a single head-on flight. The sensor rig's approximate location is denoted by the red star on the map.

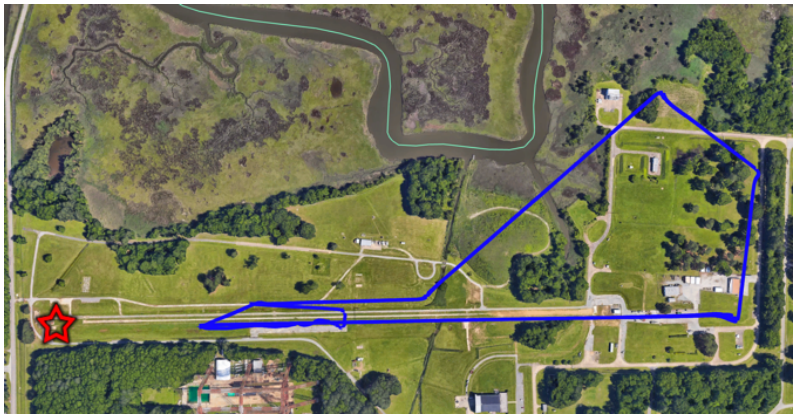


Figure 9: Ground to Air Flight Trajectories superimposed on Google Earth rendered image. Head-on style encounter with the sensor rig shown indicated by the red star and the multirotor by the blue path.

A Freefly Alta X is used as the target intruder aircraft as shown in Figure 10. The take-off weight was 20.4kg and provided up to 30 minutes of flight time for testing. The maximum diagonal length of the aircraft between rotor tips is 2.3 m. The entire mission was completed in autonomous mode with the Alta X flying waypoints pre-programmed from the ground control station with a safety pilot monitoring the entire mission.

Figure 10: Alta X.

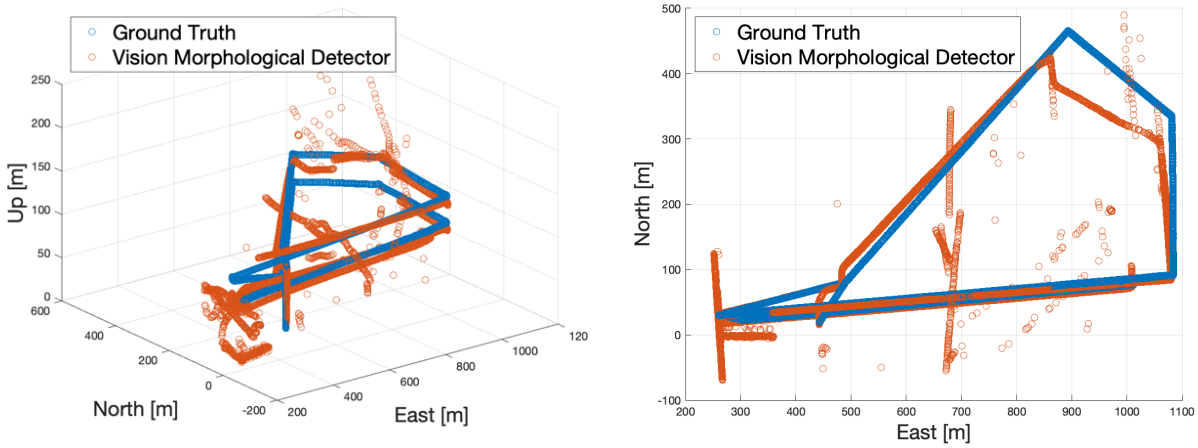
V.Detections and Measurements Results

Table 3 summarizes the dataset statistics. The camera frame rate is 30 Hz, which is approximately six times the update rate of the flight controller's position and three times the update rate of the radar. Congruently, the vision morphological detector generated nearly three times the samples as the flight controller and the difference detector fifteen times the flight controller.

Table 3: Dataset Summary

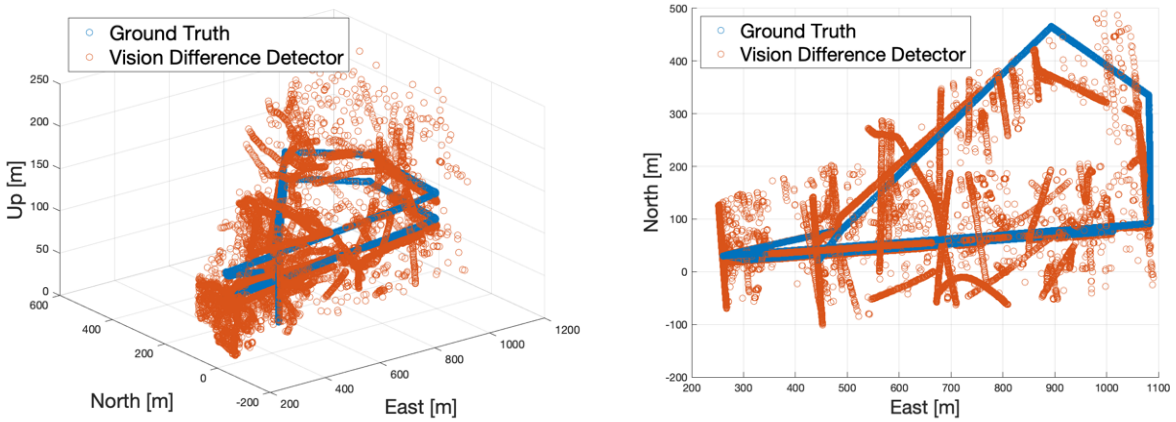
Sensor	Number of Updates
GPS position estimates from multirotor flight controller (ground truth)	7285
Vision Morphological Detector	19779
Vision Difference Detector	110802
Radar Detector (measurements)	9512

The vision detections and the radar measurements are shown in Figure 11 through Figure 13 where the blue circles are the ground truth trajectories transformed into the INS reference frame and red circles are the respective detection in the INS reference frame. Included in these figures are additional trajectories captured by the sensor from objects such as birds. Visual inspection shows that the vision morphological detector captures the aircraft with the least amount of additional detections. The vision image difference detector captures cloud movement in addition to other moving objects, i.e. birds. The radar detector 3D plot, Figure 13a, shows instability in Z direction for detections. Both sensors were subject to occlusion by trees for the 250 ft altitude in the Northeast corner and neither of them was successful in detecting the multirotor in that area as shown in the ENU detection plots.



(a) (b)

Figure 11: Vision Morphological Detector with X dimension from multirotor flight controller position transformed into INS coordinate frame (a) 3D (b) viewed from above.



(a) (b)

Figure 12: Vision Difference Detector with X dimension from multirotor flight controller position transformed into INS coordinate frame (a) 3D (b) viewed from above.

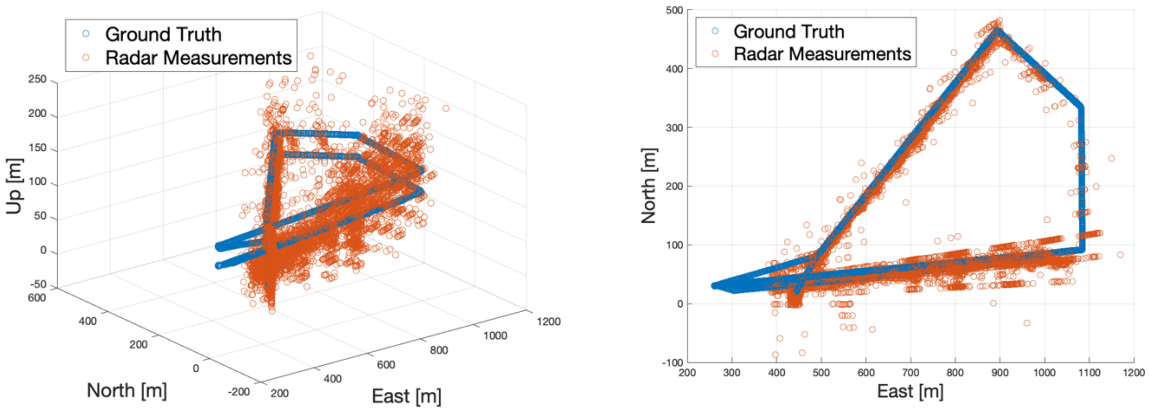


Figure 13: Radar Measurements (a) 3D (b) viewed from above.

VI. Tracker Analysis

This section first explores overall tracking performance and then presents visualizations of the tracking results. Six different trackers are evaluated in this work: 1) vision tracker augmented with range, 2) radar tracker from radar software as manufactured, 3) vision morphological detector with radar EKF tracker, 4) vision difference detector with radar EKF tracker, 5) fused vision where a blending of both aforementioned image detectors was combined with radar EKF tracker, and finally, (6) radar EKF tracker. The vision fusion method clustered morphological and difference detections within a Euclidean distance of 0.1 degrees and returns the centroid as elaborated in [9]. The vision data used in the EKF trackers does not include range data. For every time that the tracker provided a position estimate that matched a ground truth estimate, the Euclidian norm was calculated for the difference between the estimated and true position of the object as defined in Eqn. 35. The results in Table 4 show that the vision morphological detector with the radar EKF tracker provides the best $GT_{coverage,50m}$, followed by vision tracker augmented with range, EKF vision fusion with radar (within 1% of vision tracker augmented with range), radar EKF tracker, vision difference detector with radar EKF tracker, and finally the radar tracker as manufactured. Fusing the image morphological detector with the radar improved $GT_{coverage}$ by 15%. Visual inspection of the morphological detector in Figure 11 shows the most stable and least noisy detector source: combining the vision azimuth and elevation detections with radar for ranging information may provide the least noisy and most stable input for tracking purposes. The vision tracker augmented with range shows strong capabilities when combined with another source information for range.

For higher distance thresholds for $GT_{coverage,t_{error}}$, a different ranking exists, namely radar EKF tracker, followed by vision tracker augmented with range, vision morphological detector with radar EKF tracker, EKF vision fusion with radar, vision difference detector with radar EKF, and finally radar tracker as manufactured. At shorter range, the radar suffers loss of signal, resulting in significant measurement gaps. Consequently, there is a significant increase in fit error in these specific time spans. Therefore, Table 5 provides the same information as above only with EKF trackers but not considering the fitting errors during the radar measurement gaps. This results in higher percentage values but similar trends with the EKF vision morphological detector combined with radar, achieving the highest performance of 90% coverage for $GT_{coverage\ 50m}$. However, there are other important performance aspects such as the magnitude of the error or drift, which are not yet captured by the metrics as shown in Table 4 and Table 5.

Table 4: Ground truth coverage metric for the different Kalman Filter object tracker implementations sorted by $GT_{coverage,50m}$ performance.

Tracker	$GT_{coverage,50m}$	$GT_{coverage,100m}$	$GT_{coverage,200m}$
EKF Vision morphological detector with radar	74%	80%	87%
Vision Tracker Augmented with Range*	67%	85%	88%
EKF Vision fused morphological and difference detector with radar	66%	76%	85%
EKF Radar	59%	82%	89%
EKF Vision difference detector with radar	54%	70%	81%
Radar Tracker	27%	30%	38%

***Incorporates range in the INS reference frame of the multirotor flight controller position.**

Table 5: Ground truth coverage metric for the different EKF object tracker implementations constrained to timestamps where a radar detection exists sorted by $GT_{coverage,50m}$ performance.

Tracker	$GT_{coverage,50m}$	$GT_{coverage,100m}$	$GT_{coverage,200m}$
EKF Vision morphological detector with radar	90 %	97 %	99 %
EKF Vision fusion detector with radar	79 %	92 %	96 %
EKF Radar only	72 %	96 %	98 %
EKF Vision difference detector with radar	64 %	85 %	92 %

The $GT_{coverage,50m}$ in Table 5 shows 18% improvement from EKF radar tracker to vision morphological detector with radar EKF tracker, which demonstrates a strong benefit for using radar and vision sensors for multirotor tracking.

Visual inspection of the vision tracker in Figure 14 shows strong tracking of ground truth along with the flight trajectories of a few birds. The radar tracker missed large portions of the GT trajectory as shown in Figure 15. The radar tracker as manufactured may generalize well in a variety of different conditions and tracking a variety of different aircraft types whereas EKF trackers in this work, notably the EKF radar tracker, have only been tested in this experiment. Between 600 and 1100m East and 0 and 125m North, clusters of possible ground echoes are generated from the radar, which degrades radar tracking in these regions.

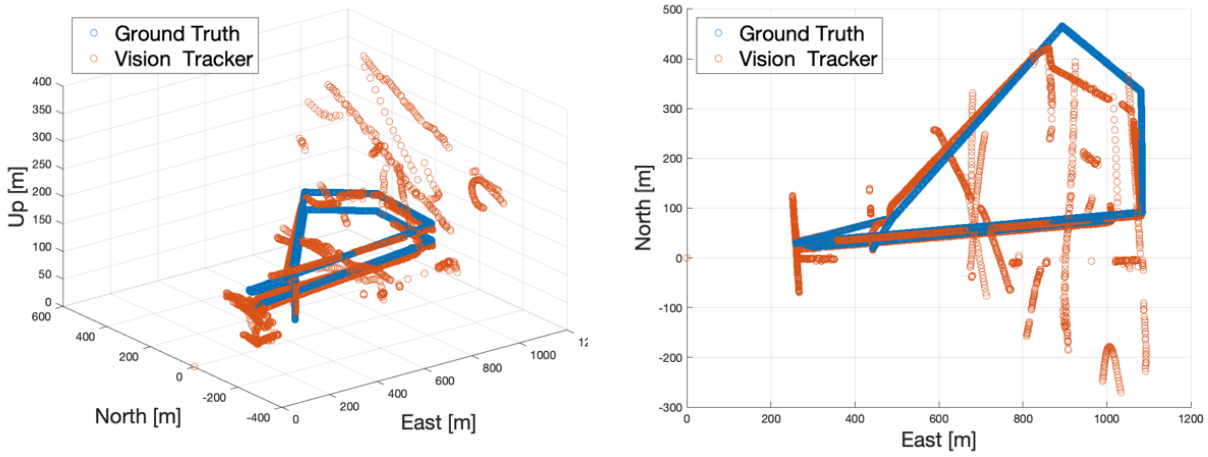


Figure 14: Vision Tracker with X dimension from multirotor flight controller position transformed into INS coordinate frame (a) 3D (b) viewed from above.

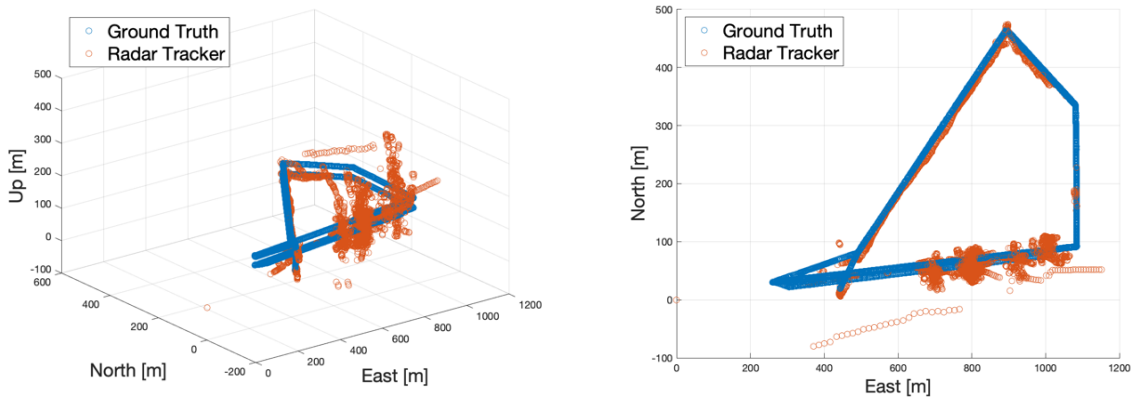


Figure 15: Radar Tracker (a) 3D (b) viewed from above.

Figure 16 through Figure 19 show detailed graphs that illustrate the performance of the different EKF implementations. The subfigures in the top row compare range, elevation, and azimuth based on the estimated track with the raw sensor measurements, radar on the left (Fig. 16a, 17a, 18a, 19a), and vision on the right (Fig. 16b, 17b, 18b). Note the significant difference in noise between the morphological detector and the difference and fused vision detectors.

The second row shows the estimated track in East – North – Up reference frame and compares against the ground truth in the left figures (Fig. 16c, 17c, 18c, 19b). These graphs show a good fit for each implementation except in the time spans where there is a loss of radar signal. The vertical estimate drifts more than the others, but this is over a smaller magnitude scale. Radar measurements for elevation are noisier than the ones for azimuth, which has an impact here as well. Figures 16d, 17d, 18d, 19c show the Euclidian norm of the fit error in Fig. 16c, 17c, 18c, 19b. Based on this signal, and the aforementioned thresholds of 50 m, 100 m and 200 m, the metrics are calculated for Table 4. The loss of radar signal has a major impact on performance, marked by the dashed vertical lines. The deterioration of the estimate in that time span is primarily caused by the lack of a third dimension in the vision measurements, which does not provide a range measurement. Also, the vision system provides ambiguous measurements at two different elevation levels in that timespan, making the estimation problem more challenging for the filter.

Finally, the last row shows the estimated track compared to ground truth in a 3D plot on the left (Fig. 16e, 17e, 18e, 19d), with the corresponding top view on the right (Fig. 16f, 17f, 18f, 19e). Comparing performance between all filter implementations shows that the estimate for the fused vision-based filter is noisier than for the morphological detector. Moreover, the estimate for the difference detector-based filter is even noisier than the previous two, which is consistent with the differences between the raw measurements. Using radar measurements only, the higher standard deviation compensates efficiently for the radar noise, but a loss of radar signal yields a loss of all sensor information (since there are no camera measurements). Consequently, this leads to a significant drift of the state estimate, which relies exclusively on prediction steps only with a constant speed assumption. This is highly undesirable and is not reflected in the numbers shown in Table 5. Convergence of the estimated state during initialization or after radar measurements becomes available again very quickly. The robustness thresholds incorporated into the radar tracker as manufactured likely provide higher stability of tracking than the EKF based radar tracker.

Overall, sensor fusion by combining measurements from different sensors results in higher performance than single sensor trackers in terms of $GT_{coverage,50m}$. Within the class of sensor fusion-based EKF implementations, the vision morphological detector combined with radar results in the best performance because ground truth coverage within 50 m is the highest, the fit error is generally the smallest, and convergence of the estimate is the fastest. To a large extent, this is due to the less noisy nature of the vision measurements. However, loss of radar measurements causes a loss of one dimension in the raw measurements (loss of range measurement), which results in significant short-term drift of the estimate. This may only be solved by including another independent sensor or source of data that provides range information. The radar EKF tracker in its current form is subject to divergence from the GT trajectory substantially as shown in Figure 19d. Adding thresholds to only assign measurements relevant to the current tracker would suppress erratic behavior in the EKF and should reduce overall localization error of all EKF iterations presented in this work.

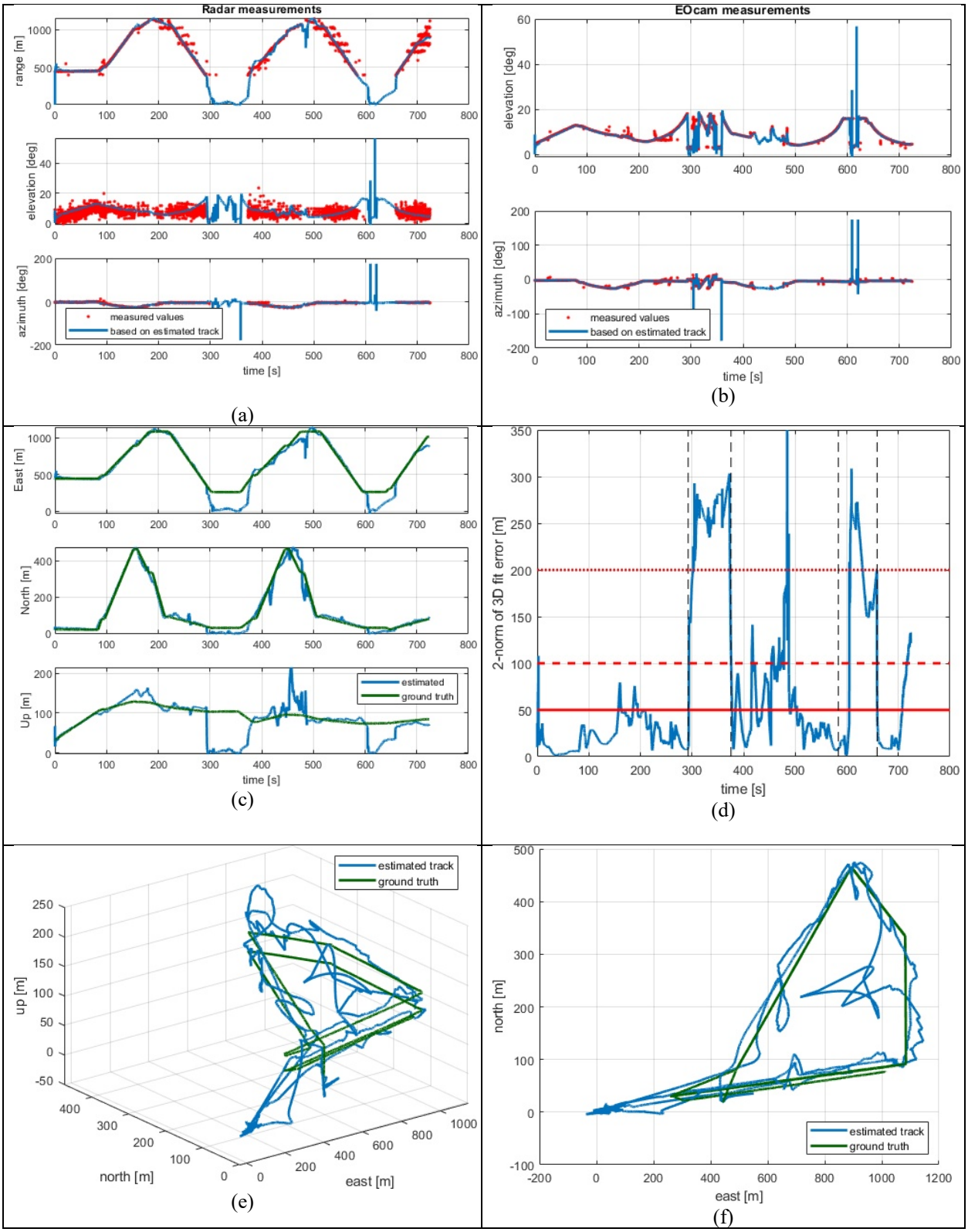


Figure 16: Performance of the EKF using vision morphological detector and radar (a) radar measurements, (b) vision detections, (c) tracking results juxtaposed against ground truth by dimension, (d) error vs time (e) tracker vs ground truth in ENU (f) viewed from above.

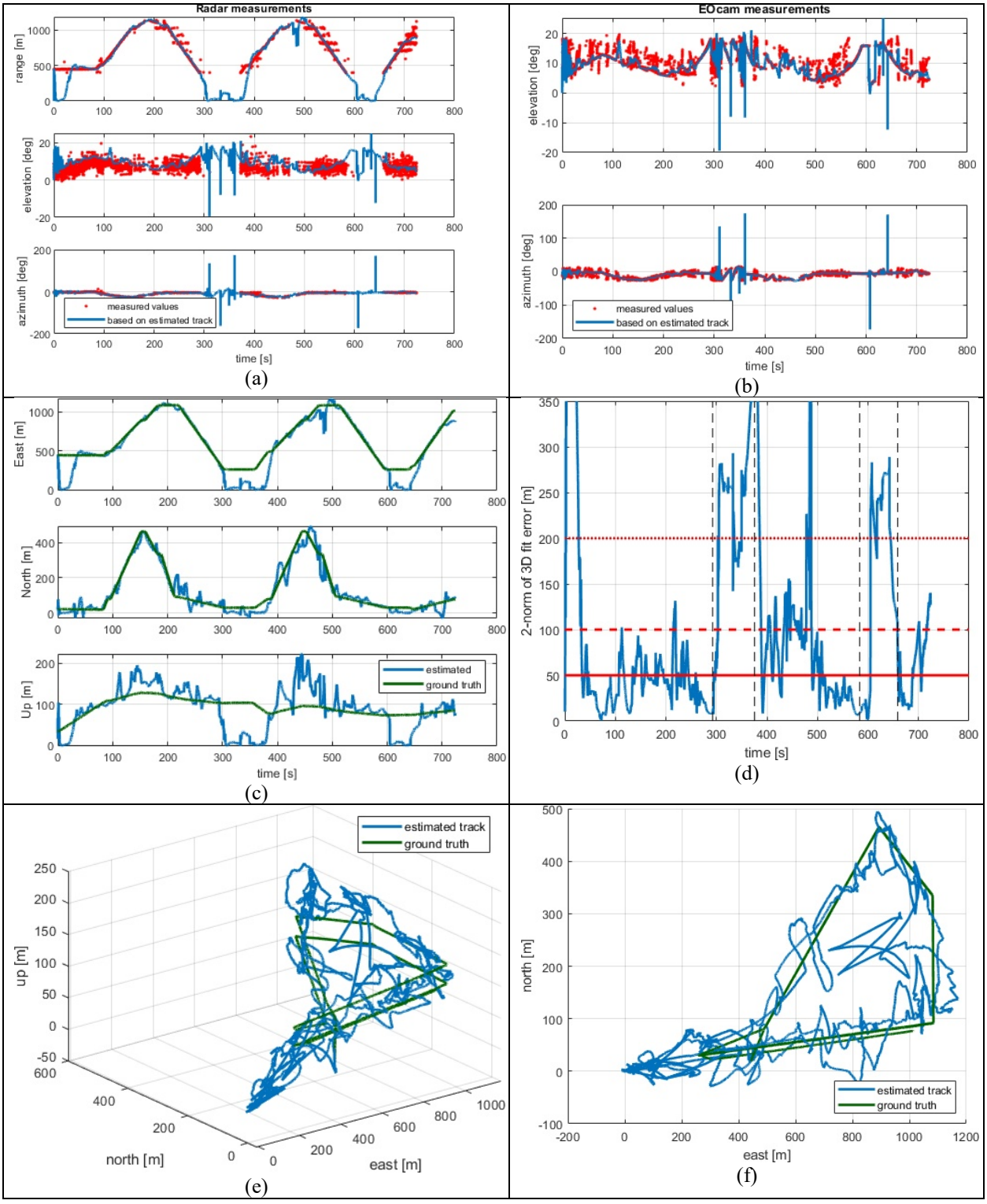


Figure 17: Performance of the KF using difference detector and radar (a) radar measurements, (b) vision detections, (c) tracking results juxtaposed against ground truth by dimension, (d) error vs time (e) tracker vs ground truth in ENU (f) viewed from above.

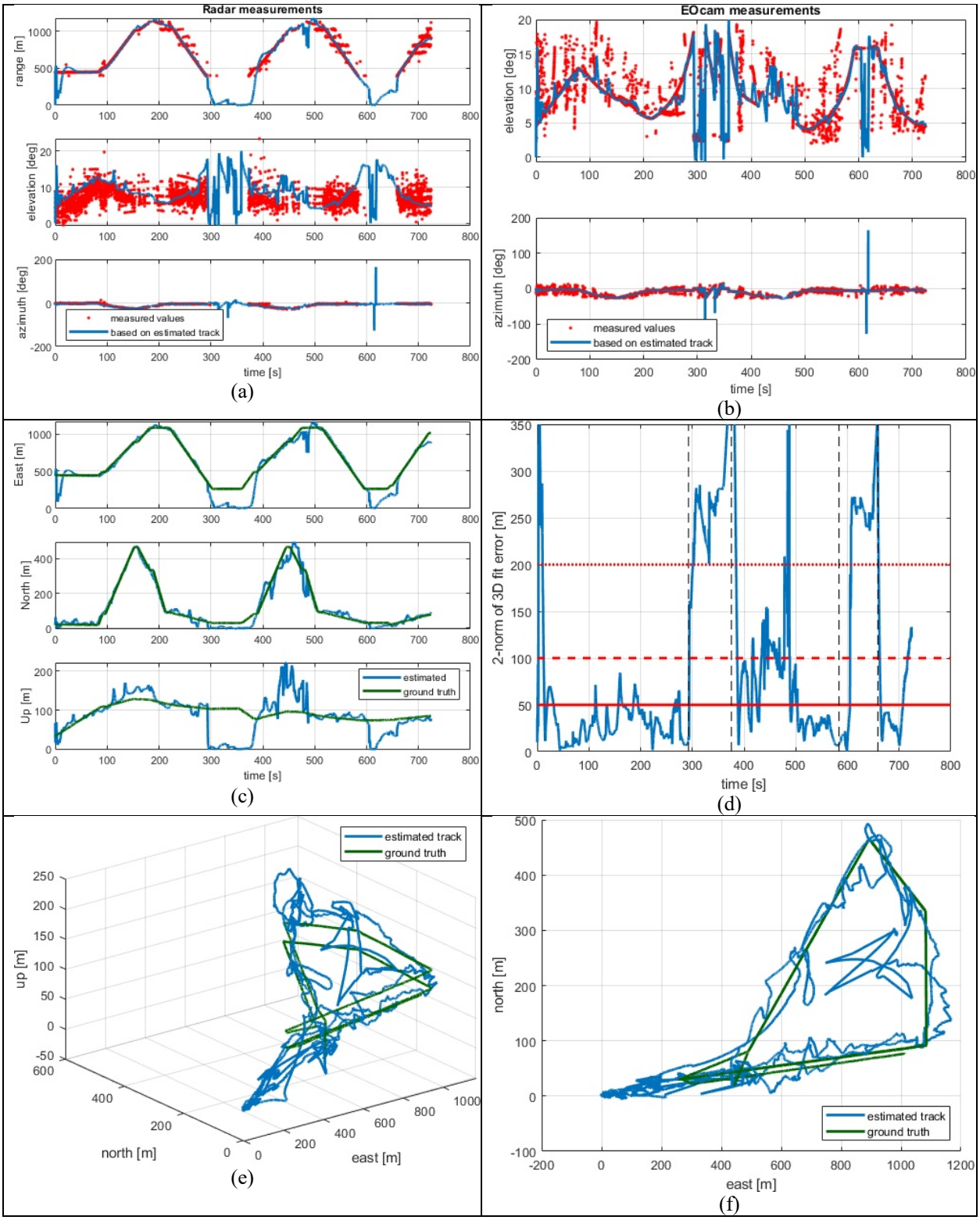


Figure 18: Performance of the KF using vision fusion detector and radar (a) radar measurements, (b) vision detections, (c) tracking results juxtaposed against ground truth by dimension, (d) error vs time (e) tracker vs ground truth in ENU (f) viewed from above.

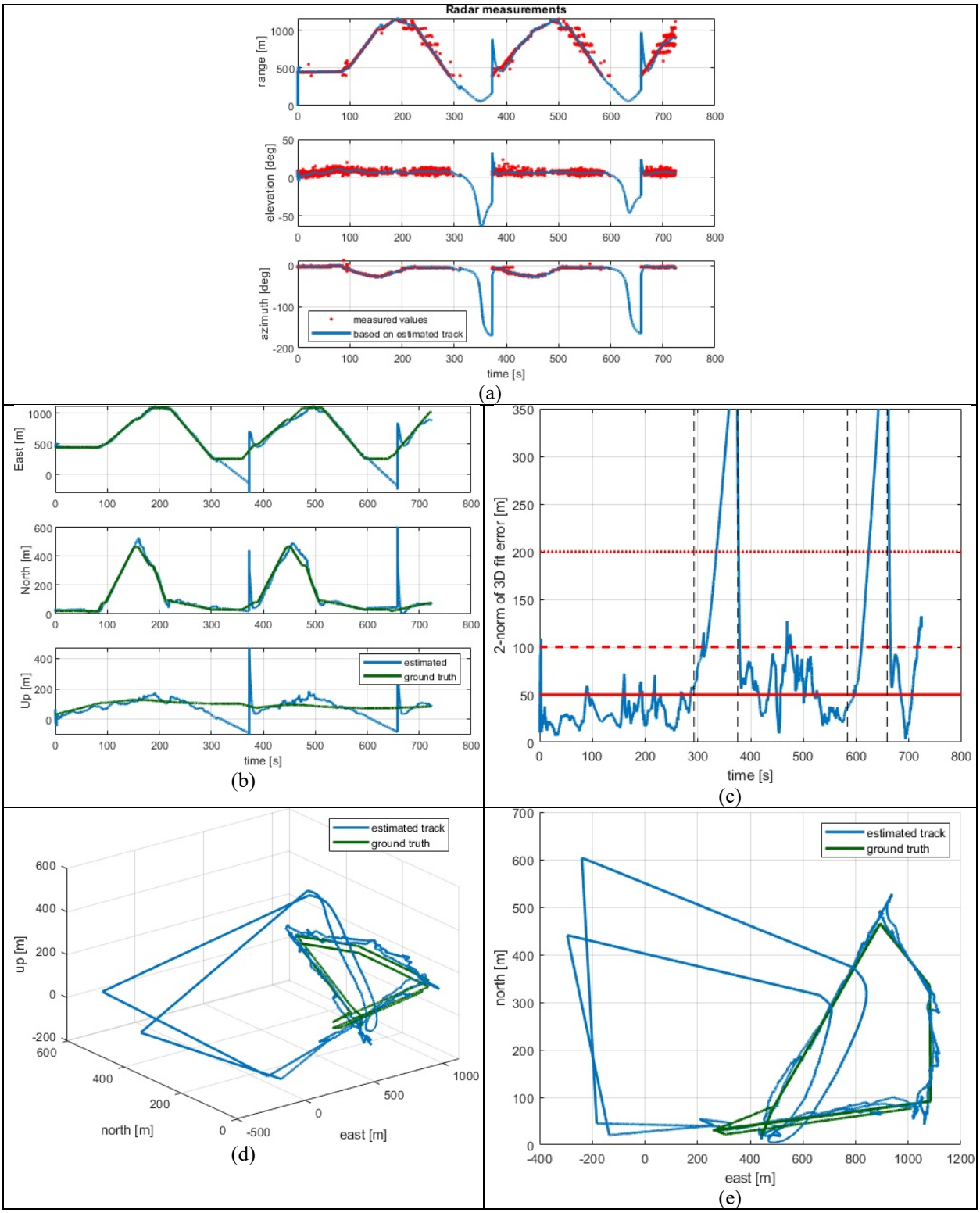


Figure 19: Performance of the KF using radar measurements only (a) radar measurements, (b) tracking results juxtaposed against ground truth by dimension, (c) error vs time (d) tracker vs ground truth in ENU (e) viewed from above.

VII. Conclusion

This work field tested an offline optical-radar fusion system designed for onboard sUAS applications for detection and tracking of aircraft at ranges relevant to intruder avoidance applications. Results indicated strong tracking performance for ranges 0.4 to 1.1 km and improved performance when fusing the radar and vision detectors in the EKF tracker. Future work will include implementing thresholds for the EKF tracker to improve stability, additional aircraft, and weather conditions. Future goals include onboard data collection and the development of a real-time optical-radar detection and tracking systems.

Acknowledgements

The authors would like to thank Greg Howland for integration and CAD work (AMA), Bryan Petty (AMA) for serving as Ground Control Station Operator, Mark Frye Range Safety Officer, and Jennifer Fowler Range Safety Officer.

References

[1] K. Shish, N. Cramer, G. Gorospe, T. Lombaerts, V. Stepanyan and K. Kannan, "Survey of Capabilities and Gaps in External Perception Sensors for Autonomous Urban Air Mobility Applications," in *AIAA SciTech*, Virtual, 2021.

[2] J. Hu, B. Zheng, C. Wang, C. Zhao, X. Hou, Q. Pan and Z. Xu, "A survey on multi-sensor fusion based obstacle detection for intelligent ground vehicles in off-road environments," *Frontiers of Information Technology & Electronic Engineering,* vol. 21, no. 5, pp. 675-692, 2020.

[3] N. Boudette, "Inside a Fatal Tesla Autopilot Accident: 'It Happened So Fast'," *New York Times,* pp. 1-7, 1 Sept 2021.

[4] U.S. Department of Transportation: National Highway Traffic Saffety Administration, "PE21-020," 31 Aug. 2021. [Online]. Available: https://static.nhtsa.gov/odi/inv/2021/INIM-PE21020-84913P.pdf.

[5] G. Fasano, D. Accardo, A. Tirri, A. Moccia and E. De Lellis, "Radar/electro-optical data fusion for non-cooperative UAS sense and avoid," *Aerospace Science and Technology,* vol. 46, pp. 436-450, 2015.

[6] M. Kramer and K. Kuhnert, "Multi-Sensor Fusion for UAV Collision Avoidance," in *2nd International Conference on Mechatronics Systems and Control* , New York City, 2018.

[7] H. Yu, F. Zhang, P. Huang, C. Wang and L. Yuanhao, "Autonomous Obstacle Avoidance for UAV based on Fusion of Radar and Monocular Camera," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, 2020.

[8] S. Samaras, E. Diamantidou, D. Ataloglou, N. Sakellariou, A. Vafeiadis, V. Magoulianitis, K. Votis, P. Daras and D. Tzovaras, "Deep Learning on Multi Sensor Data for Counter UAV Applications—A Systematic Review," *Sensors,* vol. 19, no. 22, pp. 1-35, 2019.

[9] C. Dolph, C. Minwalla, G. Glaab, M. Logan, D. Allen and K. Iftekharuddin, "Detection and Tracking of Aircraft from Small Unmanned Aerial Systems," *Journal of Aerospace Information Systems,* pp. 1-14, 2021.

[10] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration (2009)," in *VISAPP International Conference on Computer Vision Theory and Applications*, Lisboa, 2009.

[11] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence*, British Columbia, 1981.

[12] C. Serres, B. Gill and G. Wu, "EOIR Safety and Operational," 18 Feb. 2020. [Online]. Available: https://ntrs.nasa.gov/api/citations/20200002097/downloads/20200002097.pdf. [Accessed 12 Nov. 2020].

[13] M. Askelson and H. Cathey, "Small UAS Detect and Avoid Requirements Necessary for Limited Beyond Visual Line of Sight (BVLOS) Operations," U.S. Department of Transportation, Federal Aviation Administration, Washington, DC , 2017.

[14] J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," 14 Oct 2015. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/.

[15] J. Cai, P. Huang, B. Zhang and D. Wang, "A TSR Visual Servoing System Based on a Novel Dynamic Template Matching Method," *Sensor,* vol. 15, no. 12, pp. 32152-32167, 2015.

[16] T. Lombaerts, K. Shish, G. Keller, V. Stepanyan, N. Cramer and C. Ippolito, " Adaptive Multi-Sensor Fusion Based Object Tracking for Autonomous Urban Air Mobility Operations," in *AIAA SciTech* , San Diego , 2022.

[17] H. Schaub and J. Junkins, Analytical Mechanics of Space Systems, 4th Edition, Reston: American Institute of Aeronautics and Astronautics , 2018.

[18] Department of Defense, "GLOBAL POSITIONING SYSTEM STANDARD POSITIONING SERVICE PERFORMANCE STANDARD," United States of America, Washington DC, 2020.

[19] A. Myronenko and X. Song, "Point Set Registration: Coherent Point Drift," *IEEE Transactions on Pattern Analysis and Machine Intelligenc,* vol. 32, no. 12, pp. 2262-2275, 2010.