REGULAR PAPER

Camera-Based Tracking of Floating Objects using Fixed-wing UAVs

Håkon Hagen Helgesen[1] · Torleiv H. Bryne[1] · Erik F. Wilthil[1] · Tor Arne Johansen[1]

Abstract

This article concerns tracking of floating objects using fixed-wing UAVs with a monocular thermal camera. Target tracking from an agile aerial vehicle is challenging because uncertainty in the UAV pose negatively affects the accuracy of the measurements obtained through thermal images. Consequently, the accuracy of the tracking estimates is degraded if navigation uncertainty is neglected. This is especially relevant for the estimated target covariance since inconsistency is a likely consequence. A tracking system based on the Schmidt-Kalman filter is proposed to mitigate navigation uncertainty. Images gathered with an uncertain UAV pose are weighted less than images captured with a reliable pose. The UAV pose is estimated independently in a multiplicative extended Kalman filter where the estimated covariance matrix is a measure of the uncertainty. The method is compared experimentally with two traditional alternatives based on the extended Kalman filter. The results show that the proposed method performs better with respect to consistency and accuracy.

Keywords Unmanned aerial vehicle · Thermal vision · Target tracking · Navigation

1 Introduction

Target tracking is key in mapping and surveillance of the sea surface, and in collision avoidance for autonomous vehicles. It is also important for floating debris detection in seismic operations at sea, and for search and rescue operations. A common requirement is the need for a robust system that can estimate the target states precisely and deliver consistent covariance information.

Tracking systems use a single or several tracking sensors to observe targets. These observations are fused with prior information about how objects are expected to behave. The fusion typically involves filtering using variations of the Kalman filter where a prior target model is combined with a measurement model to estimate the posterior target states. The measurement model depends on the sensor, but a recurrent requirement is the need for accurate pose information (sensor position and orientation). This is particularly challenging when the sensor is mounted on an agile platform, such as fixed-wing unmanned aerial vehicles (UAVs) [12, 14].

This research investigates tracking of objects on the sea surface using a fixed-wing UAV equipped with a monocular thermal camera. The architecture presented in this work is also applicable for visual spectrum cameras. The sensor choice depends on the target characteristics. Much research have been conducted on target tracking using optical sensors where the pose of the sensor is stationary, assumed perfectly known or can be estimated with high precision [1, 15, 23]. However, tracking of objects in an Earth-fixed coordinate frame using an airborne thermal camera with navigation uncertainty is a different challenge. The camera pose depends on the position and attitude of the UAV. Therefore, errors in the navigation system of the UAV negatively affect the tracking accuracy and can cause a significant loss in performance. This is especially troublesome in multi target tracking because accurate covariance estimation is needed for measurement association.

All authors are with the NTNU Centre for Autonomous Marine Operations and Systems

✉ Håkon Hagen Helgesen
  hakon.helgesen@ntnu.no

  Torleiv H. Bryne
  torleiv.h.bryne@ntnu.no

  Erik F. Wilthil
  erik.wilthil@ntnu.no

  Tor Arne Johansen
  tor.arne.johansen@ntnu.no

[1] Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway

Target tracking using UAVs equipped with optical sensors has been studied previously [12, 13, 16, 17, 22, 30]. It is desirable to obtain high accuracy while maintaining filter consistency. Reduced tracking accuracy is not necessarily critical, but only if the covariance estimates represent the true uncertainty. In practice, this means that the tracking system should automatically adapt to the uncertainty of the UAV navigation estimates. This is closely related to the idea in simultaneous localization and mapping (SLAM) [20, 26, 28]. However, SLAM is not preferred in this research because the navigation states can be estimated precisely with inertial sensors and a global navigation satellite system (GNSS). Moreover, erroneous measurement association is a challenge in visual SLAM and can potentially cause divergence in the estimated UAV pose. Consequently, observations requiring data association are a threat to the airworthiness and are not preferred in the navigation system.

Target tracking in the presence of navigation uncertainty has been studied for autonomous ships [6, 31, 32]. However, these articles have focused on a subset of the navigation space and used a radar as the tracking sensor. Moreover, experimental results are not presented. Fixed-wing UAVs have much faster dynamics than water surface vessels and can change attitude quickly. Moreover, studying the effects of navigation uncertainty in all degrees of freedom for a UAV is to the authors best knowledge not studied previously.

This article proposes a tracking system based on the Schmidt-Kalman filter [32]. A multiplicative extended Kalman filter (MEKF) [19, 27] is used as UAV navigation system. Objects are detected automatically in thermal images using image-processing techniques [15]. The pixel position of detected objects is used as measurements in the tracking system where the uncertainty of the navigation estimates is incorporated. Several flight experiments have been carried out at sea to collect experimental data, and the results illustrate the benefit of incorporating navigation uncertainty in the tracking system. The Schmidt-Kalman filter is a variation of the extended Kalman filter and used as a proof of concept in this paper. Other filtering techniques such as the unscented Kalman filter could also benefit from the same inclusion of navigation uncertainty during tracking from an agile platform.

The rest of this article is organized as follows. Section 2 describes the navigation system and relevant coordinate frames. Section 3 describes tracking in the presence of navigation uncertainty and how the effect of navigation uncertainty is mitigated. Section 4 presents target motion and measurement models. Experimental results are described in Section 5 before the work is concluded in Section 6.

2 UAV Navigation using the Multiplicative Extended Kalman Filter

A navigation system estimates the position, velocity and attitude of a vehicle. Additional states such as acceleration, angular velocity and inertial sensor bias [8] can also be included. This section looks into UAV navigation since navigation uncertainty is incorporated in the tracking system in Section 3.

2.1 Preliminaries

The relevant coordinate frames are:

–　The north-east-down (NED) coordinate frame [9], assumed to be inertial locally and denoted {n}.
–　The body-fixed frame of the UAV [5], which is fixed to the UAV and denoted {b}.
–　A camera-fixed coordinate system {c}. The rotation matrix between {c} and {b} is known and constant.
–　The image plane, which is a two-dimensional coordinate frame. The horizontal component is denoted u and the vertical component v. Pixels in the image plane are transformed to {c} through a camera model (see Section 4).

The vector $\mathbf{p}_{nb}^n$ represents the position of {b} relative to {n} decomposed in {n}, and is the NED positions of the UAV. The target operates on the sea surface and the position is described through NED coordinates $\mathbf{p}_{nt}^n$. The coordinate systems are illustrated in Fig. 1.

The UAV states that are estimated in the navigation system are:

–　The UAV position in NED ($\mathbf{p}_{nb}^n$)
–　The UAV velocity in NED ($\mathbf{v}_{nb}^n$)
–　The UAV attitude (rotation between NED and body) parameterized by the unit quaternion ($\mathbf{q}_b^n$)
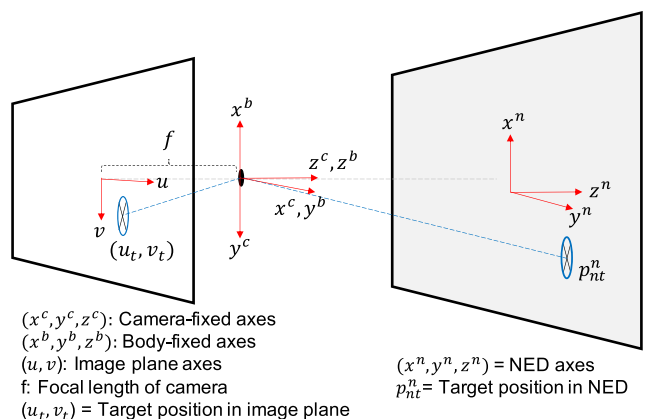

(x^c, y^c, z^c): Camera-fixed axes
(x^b, y^b, z^b): Body-fixed axes
(u, v): Image plane axes
f: Focal length of camera
(u_t, v_t) = Target position in image plane

(x^n, y^n, z^n) = NED axes
p_{nt}^n = Target position in NED

Fig. 1 Illustration of coordinate frames

- Accelerometer bias decomposed in body ($\mathbf{b}_{acc}^b$)
- Angular rate bias decomposed in body ($\mathbf{b}_{ars}^b$)

The attitude $\mathbf{q}_b^n$ is parameterized by the unit quaternion (Hamiltonian representation) and is a global and singularity-free representation of attitude [27]. The relationship between the quaternion and the corresponding rotation matrix is

$$\mathbf{R}(\mathbf{q}) = \left(q_w^2 - \mathbf{q}_v^\top \mathbf{q}_v \right) \mathbf{I} + 2\mathbf{q}_v \mathbf{q}_v^\top + 2q_w \mathbf{S}(\mathbf{q}_v) \qquad (1)$$

where q_w is the real part of the quaternion, $\mathbf{q}_v$ is the vector part of the quaternion, $\mathbf{I}$ is the identity matrix and $\mathbf{S}(\cdot)$ is the skew-symmetric matrix representing a cross-product [9]. The rotation matrix must be reconstructed from the unit quaternion when target detections are integrated in the tracking system in Section 4. The quaternion product, used for composition of rotations and in the kinematic equations in Section 2.4, is defined as [27]

$$\mathbf{q}_1 \otimes \mathbf{q}_2 = \begin{bmatrix} q_{1,w}q_{2,w} - \mathbf{q}_{1,v}^\top \mathbf{q}_{2,v} \\ q_{1,w}\mathbf{q}_{2,v} + q_{2,w}\mathbf{q}_{1,v} + \mathbf{S}(\mathbf{q}_{1,v})\mathbf{q}_{2,v} \end{bmatrix} \qquad (2)$$

where $\mathbf{q}_1$ and $\mathbf{q}_2$ are unit quaternions. In the inertial navigation problem (Section 2.5), the true attitude $\mathbf{q}$ can be expressed as $\mathbf{q} = \hat{\mathbf{q}} \otimes \delta\mathbf{q}$ where $\delta\mathbf{q}$ is the attitude error and $\hat{\mathbf{q}}$ is the nominal estimated attitude. The attitude error is parameterized using the four times Modified Rodrigues parameters (MRP) $\delta\mathbf{a}$ [19]:

$$\delta\mathbf{a} = 4\delta\mathbf{a}_{\text{mrp}} = 4 \frac{\delta\mathbf{q}_v}{1 + \delta q_w} \qquad (3)$$

The modified Rodriguez parameters represents the attitude error with the minimal number of parameters and is used to avoid potential singularities in the covariance matrix. Euler angles is also a feasible parameterization for the attitude error since the singularity at a pitch of 90 degrees [5] is an unlikely issue for fixed-wing UAVs.

The UAV position is also a combination of a nominal state estimate $\hat{\mathbf{p}}$ and an error state $\delta\mathbf{p}$. The true position is $\mathbf{p} = \hat{\mathbf{p}} + \delta\mathbf{p}$ where the reference frame is neglected for simplicity. The relationship between the true states, nominal states and error states is also a sum for the velocity, accelerometer bias and angular rate bias.

2.2 Navigation Sensors

It is assumed that the UAV is equipped with the following sensors for navigation:

- An inertial measurement unit (IMU), providing measurements of the acceleration (specific force) and angular rate at high frequency.
- Two GNSS receivers with two antennas providing independent position measurements. One antenna is

mounted in the front of the fuselage and one in the back of the fuselage.

The augmented real-time kinematic (RTK) GNSS service is used to obtain position measurements at centimeter level for both receivers using corrections from a local base station [10]. RTK functionality may not be required, but increases the estimation accuracy, particularly when the baseline between the antennas is small as on small UAVs. Using two receivers is beneficial since heading can be estimated based on GNSS position measurements [10]. Thus, it is not necessary to measure heading with other sensors such as a magnetic compass, which has significant weaknesses on-board small vehicles [10]. Heading estimation using a single GNSS receiver is challenging since non-zero angular and linear velocities are needed for observability.

2.3 Sensor Models

IMUs measure specific force ($\mathbf{f}_{\text{imu}}^b$), and angular rate ($\boldsymbol{\omega}_{\text{imu}}^b$). The sensor models follow the definitions in [9]:

$$\begin{aligned}
\mathbf{f}_{\text{imu}}^b &= \mathbf{R}_n^b \left(\mathbf{q}_n^b \right) \left(\dot{\mathbf{v}}_{nb}^n - \mathbf{g}^n \right) + \mathbf{b}_{\text{acc}}^b + \mathbf{w}_{\text{acc}}^b \\
\boldsymbol{\omega}_{\text{imu}}^b &= \boldsymbol{\omega}_{nb}^b + \mathbf{b}_{\text{ars}}^b + \mathbf{w}_{\text{ars}}^b
\end{aligned} \qquad (4)$$

where $\mathbf{b}_{\text{acc}}^b$ and $\mathbf{b}_{\text{ars}}^b$ are biases on the accelerometers and angular rate sensors, respectively. $\mathbf{w}_{\text{acc}}^b$ and $\mathbf{w}_{\text{ars}}^b$ are zero-mean noise terms. The position measurements obtained with GNSS are modeled as:

$$\mathbf{p}_{\text{gnss}}^n = \mathbf{p}_{nb}^n + \mathbf{R} \left(\mathbf{q}_b^n \right) \mathbf{r}_{bi}^b + \mathbf{w}_{\text{gnss}}^n \qquad (5)$$

where $\mathbf{w}_{\text{gnss}}^n$ is a zero-mean noise term and $\mathbf{r}_{bi}^b$ is the lever arm between the origin of {b} and the position of antenna i.

2.4 Kinematics

The navigation system is defined by the geometrical relationship between the states (kinematics). The kinematic equations are based on the strapdown equations [27]:

$$\begin{aligned}
\dot{\mathbf{p}}_{nb}^n &= \mathbf{v}_{nb}^n \\
\dot{\mathbf{v}}_{nb}^n &= \mathbf{R}_b^n \left(\mathbf{q}_b^n \right) \mathbf{f}_{nb}^b + \mathbf{g}^n \\
\dot{\mathbf{q}}_b^n &= \frac{1}{2}\mathbf{q}_b^n \otimes \begin{bmatrix} 0 \\ \boldsymbol{\omega}_{nb}^b \end{bmatrix} = \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega}_{nb}^b)\mathbf{q}_b^n
\end{aligned} \qquad (6)$$

where $\mathbf{f}_{nb}^b$ is the specific force acting on the UAV and $\mathbf{g}^n = [0, 0, g]^\top$ is the gravity vector in NED [10]. Moreover, the angular velocity of {b} relative to {n} is $\boldsymbol{\omega}_{nb}^b$. $\boldsymbol{\Omega}(\boldsymbol{\omega}_{nb}^b)$ is a 4x4 matrix used to replace the quaternion product in Eq. 6:

$$\boldsymbol{\Omega}(\boldsymbol{\omega}_{nb}^b) = \begin{bmatrix} 0 & -(\boldsymbol{\omega}_{nb}^b)^\top \\ \boldsymbol{\omega}_{nb}^b & -\mathbf{S}(\boldsymbol{\omega}_{nb}^b) \end{bmatrix} \qquad (7)$$

2.5 Error-State Navigation System

The error-state formulation of the Kalman filter is used as navigation system in this research. The fundamental idea of the error-state formulation is to use the IMU to estimate high-frequency nominal state estimates and use low-frequency aiding measurements (GNSS) to estimate the error states. The error-state Kalman filter is described in detail in [27]. The nominal state vector is

$$\hat{\mathbf{x}} = \left[\hat{\mathbf{p}}_{nb}^n, \ \hat{\mathbf{v}}_{nb}^n, \ \hat{\mathbf{q}}_b^n, \ \hat{\mathbf{b}}_{\text{acc}}^b, \ \hat{\mathbf{b}}_{\text{ars}}^b \right]^\top \tag{8}$$

and includes the position, velocity, and attitude of the UAV. Moreover, the IMU biases are also estimated. The error-state vector is

$$\delta\mathbf{x} = \left[\delta\mathbf{p}_{nb}^n, \ \delta\mathbf{v}_{nb}^n, \ \delta\mathbf{a}, \ \delta\mathbf{b}_{\text{acc}}^b, \ \delta\mathbf{b}_{\text{ars}}^b \right]^\top \tag{9}$$

where each element in the nominal state vector $\hat{\mathbf{x}}$ has a corresponding error state. Note that $\delta\mathbf{a}$ is the attitude error represented by the four times modified Rodrigues parameters as described in Section 2.1. The true state $\mathbf{x}$ is expressed as $\mathbf{x} = \hat{\mathbf{x}} \otimes \delta\mathbf{x}$ where $\otimes$ is a generic composition used to describe the relationship between the true, nominal and error states. The nominal states are estimated through Eqs. (6) and (4):

$$
\begin{aligned}
\dot{\hat{\mathbf{p}}}_{nb}^n &= \hat{\mathbf{v}}_{nb}^n \\
\dot{\hat{\mathbf{v}}}_{nb}^n &= \mathbf{R}_b^n(\hat{\mathbf{q}}_b^n) \left(\mathbf{f}_{\text{imu}}^b - \hat{\mathbf{b}}_{\text{acc}}^b \right) + \mathbf{g}^n \\
\dot{\hat{\mathbf{q}}}_b^n &= \frac{1}{2} \boldsymbol{\Omega} \left(\boldsymbol{\omega}_{\text{imu}}^b - \hat{\mathbf{b}}_{\text{ars}}^b \right) \hat{\mathbf{q}}_b^n \\
\dot{\hat{\mathbf{b}}}_{\text{acc}}^b &= -\mathbf{T}_{\text{acc}}^{-1} \hat{\mathbf{b}}_{\text{acc}}^b \\
\dot{\hat{\mathbf{b}}}_{\text{ars}}^b &= -\mathbf{T}_{\text{ars}}^{-1} \hat{\mathbf{b}}_{\text{ars}}^b
\end{aligned}
\tag{10}
$$

$\mathbf{T}_{\text{acc}}$ and $\mathbf{T}_{\text{ars}}$ are diagonal time constant matrices used to model the biases.

The error states are estimated using two GNSS position measurements. The Jacobian of Eq. 5 with respect to the error states is

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & -\mathbf{R}(\hat{\mathbf{q}}_b^n)\mathbf{S}(\mathbf{r}_{b,1}) & \mathbf{0}_{3\times6} \\ \mathbf{I}_3 & \mathbf{0}_{3\times3} & -\mathbf{R}(\hat{\mathbf{q}}_b^n)\mathbf{S}(\mathbf{r}_{b,2}) & \mathbf{0}_{3\times6} \end{bmatrix} \tag{11}$$

where $\mathbf{0}_{m\times n}$ is a matrix of zeros with m rows and n columns and $\mathbf{I}_3$ is the identity matrix of dimension 3. The Jacobian is used to update the error states before the error is propagated back into the nominal estimates through the reset operation described in [27].

3 Target Tracking using the Schmidt-Kalman Filter

This section describes target tracking using the Schmidt-Kalman filter (SKF). The goal is to mitigate the effects of uncertainty in the sensor pose. The main issue when neglecting navigation uncertainty is inconsistency in the tracking filter. This is normally observed as the filter being optimistic and estimating a covariance that is much lower than the corresponding mean square estimation error. It is possible to counteract this behavior by increasing the covariance matrices for the noise, but this is an ad-hoc technique that disguises the underlying problem. Moreover, the filter parameters must be tuned towards a particular scenario or set of data. Therefore, it is hard to generalize this strategy to fit new and unknown data. The SKF is a variation of the extended Kalman filter (EKF) where navigation uncertainty is incorporated as states in the filter and where correlations between the navigation system and tracking estimates is maintained. Other tracking filters such as the unscented Kalman filter or the particle filter are viable alternatives to the EKF, particularly when the measurement model is highly nonlinear. Nevertheless, the motivation in this paper is to investigate how navigation uncertainty can be included in a filtering approach and the EKF is used as a proof of concept. The methodology can be generalized to fit into other filter structures such as the unscented Kalman filter.

3.1 Introduction to the Schmidt-Kalman Filter

The sensor pose is normally assumed perfectly known. This is reasonable in some applications (e.g., when only the relative position is of interest), but limits the performance of the tracking system in applications where the sensor pose is somewhat uncertain. This is particularly relevant when targets are tracked in an Earth-fixed coordinate frame from an agile fixed-wing UAV.

Simultaneous localization and mapping [7] handles mapping in the presence of navigation uncertainty. A common assumption is that landmarks are static. Thus, the conventional SLAM architecture cannot cover a situation with moving targets. SLAM can be extended to consider moving landmarks as suggested in [29]. However, the fundamental idea in SLAM is that target (landmark) observations can influence the UAV pose. Consequently, erroneous data association, inaccurate measurements and the unknown target behavior degrade the accuracy of the UAV pose when using SLAM but might be necessary in certain situations such as in GNSS-denied applications. However, this is not desired in this research because measurements from optical sensors can be inaccurate. Moreover, the sparse set of distinct features on the sea surface makes SLAM architectures less viable for UAV operations at sea.

A more attractive option is to use a framework where the navigation system of the UAV operates independently, but where the uncertainty of the navigation estimates is allowed

to influence the tracking system. This can be achieved with the SKF [8, 21, 25, 32], where information can flow from the navigation system to the tracking system. The SKF is a sub-optimal approach, but more robust when considering model mismatch and erroneous data association.

3.2 Tracking System Model

Let $\mathbf{x}_k^t$ denote the target state vector at time k. A common way to estimate the target states is to use an EKF. The conventional EKF does not account for uncertainty in parameters affecting the motion and measurement models. Instead, all uncertainty is modeled as additive white noise. This is not a reliable way to model the uncertainty when navigation errors are present.

Consider a situation where the camera pose is defined as $\mathbf{x}^o = \hat{\mathbf{x}}^o \otimes \delta\mathbf{x}^o$ where $\hat{\mathbf{x}}^o$ represents a nominal known estimate and $\delta\mathbf{x}^o$ is the error state (follows the notation of the error-state Kalman filter). Moreover, assume that the uncertainty of the error state is known through an estimated covariance matrix $\mathbf{P}^o$. The motivation behind this formulation is to augment the tracking system with $\delta\mathbf{x}^o$ and let $\mathbf{P}^o$ influence the target estimates. Only the error states that directly affect the tracking system are included. The augmented state transition model is written as

$$\begin{bmatrix} \mathbf{x}_k^t \\ \delta\mathbf{x}_k^o \end{bmatrix} = \begin{bmatrix} \mathbf{F}^t & \mathbf{0} \\ \mathbf{0} & \mathbf{F}^o \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k-1}^t \\ \delta\mathbf{x}_{k-1}^o \end{bmatrix} + \begin{bmatrix} \mathbf{v}_k^t \\ \mathbf{v}_k^o \end{bmatrix} \tag{12}$$

where $\mathbf{F}^t$ is the target model, and $\mathbf{F}^o$ is the navigation error model (camera pose). $\mathbf{v}_k^t$ and $\mathbf{v}_k^o$ are additive white noise for the target and navigation error states, respectively. The state transition models of the error state and target are uncoupled because the target behavior should not affect the trajectory of the UAV and vice versa. The covariance matrix of the augmented system is

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}^t & \mathbf{C} \\ \mathbf{C}^\top & \mathbf{P}^o \end{bmatrix} \tag{13}$$

where $\mathbf{P}^t$ is the target covariance and $\mathbf{C}$ is the cross-covariance.

Assume that the augmented tracking system has a measurement model that depends on both the target states and navigation states of the UAV:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}^t, \hat{\mathbf{x}}^o, \delta\mathbf{x}^o) + \mathbf{w} \tag{14}$$

where $\mathbf{z}$ is the measurement vector and $\mathbf{w}$ is additive zero-mean white Gaussian noise. The nominal navigation estimate $\hat{\mathbf{x}}^o$ is known from the navigation system. When the measurement model is nonlinear, the Jacobian of Eq. 14 is used in the EKF. The idea in the SKF is to calculate the Jacobian with respect to both the target states $\mathbf{x}^t$ and the

error states $\delta\mathbf{x}^o$. Consequently, the linearized measurement model is

$$\mathbf{z}_k = \mathbf{h}\left(\hat{\mathbf{x}}_k^t, \hat{\mathbf{x}}_k^o, \delta\hat{\mathbf{x}}_k^o\right) + \begin{bmatrix} \mathbf{H}_k^t & \mathbf{H}_k^o \end{bmatrix} \begin{bmatrix} \mathbf{x}_k^t - \hat{\mathbf{x}}_k^t \\ \delta\mathbf{x}_k^o - \delta\hat{\mathbf{x}}_k^o \end{bmatrix}$$
$$+ \mathbf{w}_k + \Delta \tag{15}$$

where $(\hat{\mathbf{x}}_k^t, \hat{\mathbf{x}}_k^o, \delta\hat{\mathbf{x}}_k^o)$ is the linearization point and Δ represents higher order terms that are neglected in the first-order filter. $\mathbf{H}^t$ and $\mathbf{H}^o$ are measurement Jacobians differentiated with respect to the target states and the navigation error states, respectively.

3.3 Update Equations for the Schmidt-Kalman Filter

The SKF state transition equations are based on the augmented system in Eqs. 12-13, and the Kalman filter equations. The a priori update is

$$\begin{aligned} \mathbf{x}_{k|k-1}^t &= \mathbf{F}^t \mathbf{x}_{k-1|k-1}^t \\ \delta\mathbf{x}_{k|k-1}^o &= \mathbf{F}^o \delta\mathbf{x}_{k-1|k-1}^o \\ \mathbf{P}_{k|k-1}^t &= \mathbf{F}^t \mathbf{P}_{k-1|k-1}^t (\mathbf{F}^t)^\top + \mathbf{Q}^t \\ \mathbf{C}_{k|k-1} &= \mathbf{F}^t \mathbf{C}_{k-1|k-1} \left(\mathbf{F}^o\right)^\top \end{aligned} \tag{16}$$

where $\mathbf{Q}^t$ is the covariance matrix of the process noise in the target model. The covariance of the navigation error state, $\mathbf{P}^o$, is extracted from the navigation system at each time step. The Kalman gain is divided into two components, one that affects the target states and one that affects the navigation error:

$$\mathbf{K}_k = \begin{bmatrix} \mathbf{K}_k^t \\ \mathbf{K}_k^o \end{bmatrix} = \mathbf{P}_{k|k-1} \begin{bmatrix} \left(\mathbf{H}_k^t\right)^\top \\ \left(\mathbf{H}_k^o\right)^\top \end{bmatrix} \mathbf{S}_k^{-1} \tag{17}$$

where the innovation covariance is defined as

$$\begin{aligned} \mathbf{S}_k &= \begin{bmatrix} \mathbf{H}_k^t & \mathbf{H}_k^o \end{bmatrix} \mathbf{P}_{k|k-1} \begin{bmatrix} \left(\mathbf{H}_k^t\right)^\top \\ \left(\mathbf{H}_k^o\right)^\top \end{bmatrix} + \mathbf{R}_k \\ &= \begin{bmatrix} \mathbf{H}_k^t & \mathbf{H}_k^o \end{bmatrix} \begin{bmatrix} \mathbf{P}_{k|k-1}^t & \mathbf{C}_{k|k-1} \\ \mathbf{C}_{k|k-1}^\top & \mathbf{P}_{k|k-1}^o \end{bmatrix} \begin{bmatrix} \left(\mathbf{H}_k^t\right)^\top \\ \left(\mathbf{H}_k^o\right)^\top \end{bmatrix} + \mathbf{R}_k \\ &= \mathbf{H}_k^t \mathbf{P}_{k|k-1}^t \left(\mathbf{H}_k^t\right)^\top + \mathbf{H}_k^t \mathbf{C}_{k|k-1} \left(\mathbf{H}_k^o\right)^\top \\ &\quad + \mathbf{H}_k^o \mathbf{C}_{k|k-1}^\top \left(\mathbf{H}_k^t\right)^\top + \mathbf{H}_k^o \mathbf{P}_{k|k-1}^o \left(\mathbf{H}_k^o\right)^\top + \mathbf{R}_k \end{aligned} \tag{18}$$

and $\mathbf{R}_k$ is the covariance of the measurement noise. The SKF formulation appears by choosing a suboptimal gain where $\mathbf{K}_k^o$ is forced to be zero so that the error state, $\delta\mathbf{x}_{k|k}^o$, is constant over the measurement update. The target gain is

$$\mathbf{K}_k^t = \left(\mathbf{P}_{k|k-1}^t \left(\mathbf{H}_k^t\right)^\top + \mathbf{C}_{k|k-1} \left(\mathbf{H}_k^o\right)^\top\right) \mathbf{S}_k^{-1} \tag{19}$$

and the a posteriori state estimates are given as

$$\begin{aligned} \mathbf{x}_{k|k}^t &= \mathbf{x}_{k|k-1}^t + \mathbf{K}_k^t \left(\mathbf{z}_k - \mathbf{H}_k^t \mathbf{x}_{k|k-1}^t\right) \\ \delta\mathbf{x}_{k|k}^o &= \delta\mathbf{x}_{k|k-1}^o \end{aligned} \tag{20}$$

The a posteriori estimates of the covariance and the cross-covariance differ from the regular EKF equations because a suboptimal gain is chosen. Therefore, it is not possible to use the simplified formula for the a posteriori estimate. The Joseph form [3], which is valid for any gain, is used instead:

$$
\begin{aligned}
\mathbf{P}_{k|k} = {} & \left(\mathbf{I} - \begin{bmatrix} \mathbf{K}_k^t \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{H}_k^t & \mathbf{H}_k^o \end{bmatrix} \right) \begin{bmatrix} \mathbf{P}_{k|k-1}^t & \mathbf{C}_{k|k-1} \\ \mathbf{C}_{k|k-1}^\top & \mathbf{P}_{k|k-1}^o \end{bmatrix} \\
& \left(\mathbf{I} - \begin{bmatrix} \mathbf{K}_k^t \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{H}_k^t & \mathbf{H}_k^o \end{bmatrix} \right)^\top + \begin{bmatrix} \mathbf{K}_k^t \mathbf{R}_k \left(\mathbf{K}_k^t \right)^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}
\end{aligned} \tag{21}
$$

Moreover, the following a posteriori estimates for the target covariance and the cross-covariance are acquired

$$
\begin{aligned}
\mathbf{P}_{k|k}^t = {} & \left(\mathbf{I} - \mathbf{K}_k^t \mathbf{H}_k^t \right) \mathbf{P}_{k|k-1}^t \left(\mathbf{I} - \mathbf{K}_k^t \mathbf{H}_k^t \right)^\top \\
& - \mathbf{C}_{k|k-1} \left(\mathbf{H}_k^o \right)^\top \left(\mathbf{K}_k^t \right)^\top + \mathbf{K}_k^t \mathbf{H}_k^t \mathbf{C}_{k|k-1} \left(\mathbf{H}_k^o \right)^\top \left(\mathbf{K}_k^t \right)^\top \\
& - \mathbf{K}_k^t \mathbf{H}_k^o \mathbf{C}_{k-1}^\top + \mathbf{K}_k^t \mathbf{H}_k^o \mathbf{C}_{k|k-1}^\top \left(\mathbf{H}_k^t \right)^\top \left(\mathbf{K}_k^t \right)^\top \\
& + \mathbf{K}_k^t \mathbf{H}_k^o \mathbf{P}_{k|k-1}^o \left(\mathbf{H}_k^o \right)^\top \left(\mathbf{K}_k^t \right)^\top + \mathbf{K}_k^t \mathbf{R}_k \left(\mathbf{K}_k^t \right)^\top
\end{aligned}
$$

$$
\mathbf{C}_{k|k} = \mathbf{C}_{k|k-1} - \mathbf{K}_k^t \left(\mathbf{H}_k^t \mathbf{C}_{k|k-1} + \mathbf{H}_k^o \mathbf{P}_{k|k-1}^o \right) \tag{22}
$$

This concludes the main structure of the SKF. The key step is to use the suboptimal gain. The tracking system must maintain the target estimates and corresponding covariance in addition to the cross covariance. The navigation system maintains the covariance of the error state and the nominal state estimates. It is not influenced by the tracking system.

3.4 Navigation Error Models

When a monocular camera is used to track objects, only the position and the attitude of the camera influence the tracking system. Consequently, the error state $\delta\mathbf{x}^o$ in the augmented system includes the position and attitude of the UAV:

$$
\delta\mathbf{x}^o = \begin{bmatrix} \delta\mathbf{p}^o \\ \delta\mathbf{a}^o \end{bmatrix} \tag{23}
$$

$\delta\mathbf{p}^o$ is the error state in position. $\delta\mathbf{a}^o$ is the error state in attitude and represented by the four times modified Rodriguez parameters (see Section 2). $\mathbf{F}^o$, in Eq. 16, is chosen as the identity matrix since the navigation error is estimated outside of the tracking system. This is a fundamental difference from SLAM where the error states are estimated within the same filter as the tracking estimates. The next part discusses how the equations in Section 3.3 are affected by the properties of the error states.

3.4.1 Zero-Mean and Uncorrelated Navigation State Errors

Assume that the navigation system is consistent without a bias. This gives uncertainty in the UAV pose that is zero

mean and uncorrelated. Consequently, the measurement noise induced in the tracking filter is white. Moreover, the cross-correlation between the target estimates and the navigation error, in Eq. 13, is zero ($\mathbf{C} = 0$). Therefore, the SKF equations can be simplified and the only effect on the tracking system is a so-called covariance inflation (also called consider covariance), where the innovation covariance is reduced to

$$
\mathbf{S}_k = \mathbf{R}_k + \mathbf{H}_k^t \mathbf{P}_{k|k-1}^t \left(\mathbf{H}_k^t \right)^\top + \mathbf{H}_k^o \mathbf{P}_{k|k-1}^o \left(\mathbf{H}_k^o \right)^\top \tag{24}
$$

and the a posteriori target covariance estimate is

$$
\begin{aligned}
\mathbf{P}_{k|k}^t = {} & \left(\mathbf{I} - \mathbf{K}_k^t \mathbf{H}_k^t \right) \mathbf{P}_{k|k-1}^t \left(\mathbf{I} - \mathbf{K}_k^t \mathbf{H}_k^t \right)^\top \\
& + \mathbf{K}_k^t \mathbf{H}_k^o \mathbf{P}_{k|k-1}^o \left(\mathbf{H}_k^o \right)^\top \left(\mathbf{K}_k^t \right)^\top + \mathbf{K}_k^t \mathbf{R}_k \left(\mathbf{K}_k^t \right)^\top
\end{aligned} \tag{25}
$$

The second term represents the covariance inflation. This is a viable strategy to mitigate the effect of navigation uncertainty, but one that neglects the cross-covariance. This strategy is called detuned EKF in the experimental validation in Section 5.

3.4.2 Constant Navigation State Bias

A second and more likely situation in experimental data is navigation errors that introduce a constant bias in the target observations. This can be caused by an unknown mounting misalignment between the IMU and the camera. It can also be caused by a constant bias that is estimated poorly in the navigation system. Moreover, experimental data are not necessarily obeying known models so it is likely that estimation errors are correlated in time [12]. Consequently, a nonzero cross-covariance between the target estimates and the navigation errors appears. This mindset gives the equations in Section 3.3 where $\mathbf{F}^o$ is the identity matrix. This formulation is called SKF in the experimental validation in Section 5.

3.4.3 Other Navigation State Error Models

Error states that behave like slowly varying biases can be covered by the previous case, but higher-order models cannot. If it is assumed that the navigation system is initialized properly and has been running for a few minutes before the tracking objective starts, it is likely that sensor biases and other disturbances (in the navigation filter) have converged. Therefore, it is not considered to be beneficial to model the navigation error with higher-order models in this research. Higher-order models are more relevant in dead reckoning situations caused by GNSS dropout for example. In such a situation, SLAM architectures must be considered.

4 Camera-Based Tracking of Floating Objects

The general idea of tracking in the presence of navigation uncertainty was derived in Section 3. This section defines specific models for tracking of floating objects using monocular optical sensors.

4.1 Target State-Transition Model

Floating objects typically have slow variations in speed and course except for wave-induced motions that are not considered. The target dynamics can be covered by a state vector, $\mathbf{x}^t = [\mathbf{p}_{nt}^n, \mathbf{v}_{nt}^n]^\top$, where the north and east positions $(\mathbf{p}_{nt}^n)$ and velocities $(\mathbf{v}_{nt}^n)$, are included. It is common to choose the following uncoupled (near) constant-velocity target state transition model:

$$\mathbf{x}_k^t = \mathbf{F}^t \mathbf{x}_{k-1}^t + \Gamma \mathbf{v}_k$$

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \frac{1}{2}T^2 & 0 \\ 0 & \frac{1}{2}T^2 \\ T & 0 \\ 0 & T \end{bmatrix} \tag{26}$$

where the subscript k refers to time step k, T is the sampling interval, and the covariance of the process noise is

$$\begin{aligned} \mathbf{Q} &= E[\Gamma \mathbf{v}_k \mathbf{v}_k \Gamma^\top] \\ &= \Gamma \sigma_v^2 \Gamma^\top \\ &= \sigma_v^2 \begin{bmatrix} \frac{1}{4}T^4 & 0 & \frac{1}{2}T^3 & 0 \\ 0 & \frac{1}{4}T^4 & 0 & \frac{1}{2}T^3 \\ \frac{1}{2}T^3 & 0 & T^2 & 0 \\ 0 & \frac{1}{2}T^3 & 0 & T^2 \end{bmatrix} \end{aligned} \tag{27}$$

where $\sigma_v^2 \mathbf{I}$ is the covariance of $\mathbf{v}$, assumed to be equal in north and east. The constant velocity model is defined in discrete time and assumes that the motions in the north and east coordinates are uncoupled. The down coordinate is neglected in the target vector since targets remain on the sea surface at approximately the same altitude. More complex target state transition models are presented in [24]. Without prior knowledge about the target behavior, a simple model with a small amount of parameters is often the most reliable choice [14].

4.2 Measurement Model

Monocular cameras only provide bearing information. However, the altitude of the targets is known for objects on the sea surface. Target observations are acquired through object detection in thermal images [15, 17]. The target detection algorithm returns the centroid of the target in

the image plane. The target pixel position is related to the camera frame through the pinhole camera model [11]:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} u_t \\ v_t \end{bmatrix} + \mathbf{w}_{\text{cam}} \tag{28}$$

where z_1 and z_2 are measurements, (u_t, v_t) are the pixel coordinates of the detected target center and $\mathbf{w}_{\text{cam}}$ is Gaussian zero-mean measurement noise. The pixel position of the target can be related to the target position in the camera-fixed frame through the pinhole camera model:

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \frac{f}{z_t^c} \begin{bmatrix} x_t^c \\ y_t^c \end{bmatrix} \tag{29}$$

where f is the focal length of the camera lens and $(x_t^c, y_t^c, z_t^c) = \mathbf{p}_{ct}^c$ are the target coordinates in the camera-fixed coordinate frame. The camera-fixed position of a target is related to the camera pose and target position in NED through the following equation:

$$\mathbf{p}_{ct}^c = \mathbf{R}_n^c \left(\mathbf{p}_{nt}^n - \mathbf{p}_{nb}^n - \mathbf{p}_{bc}^n \right) = \mathbf{R}_n^c \left(\mathbf{p}_{nt}^n - \mathbf{p}_{nb}^n \right) \tag{30}$$

where $\mathbf{R}_n^c$ is the rotation matrix between {c} and {n} and depends on the camera orientation. $\mathbf{p}_{bc}^n$ is the lever arm between the origin of {b} and the origin of {c}. These frames are assumed to coincide so $\mathbf{p}_{bc}^n = \mathbf{0}$. Equation 30 can also be expressed with the navigation error states:

$$\mathbf{p}_{ct}^c = \mathbf{R}_b^c \mathbf{R}_{\hat{b}}^b \mathbf{R}_n^{\hat{b}} \left(\mathbf{p}_{nt}^n - \hat{\mathbf{p}}_{nb}^n - \delta \mathbf{p}_{nb}^n \right) \tag{31}$$

where $\mathbf{R}_b^c$ is a known and constant rotation matrix between the camera-fixed frame and the body-fixed frame. $\mathbf{R}_{\hat{b}}^b$ is a rotation matrix given by the error state of the attitude, and $\mathbf{R}_n^{\hat{b}}$ is the (nominal) estimated rotation matrix between the body frame and NED in the navigation system. $\hat{\mathbf{p}}_{nb}^n$ is the nominal NED positions of the UAV estimated by the navigation system, and $\delta \mathbf{p}_{nb}^n$ is the error state in the UAV position. $\mathbf{R}_{\hat{b}}^b$ represents the difference between the estimated body frame and the true body frame.

To simplify the Jacobian calculation in the SKF, it is beneficial to rewrite (31) as:

$$\begin{aligned} \mathbf{p}_{ct}^c &= \mathbf{R}_b^c \mathbf{R}_{\hat{b}}^b \mathbf{R}_n^{\hat{b}} \left(\mathbf{p}_{nt}^n - \hat{\mathbf{p}}_{nb}^n - \delta \mathbf{p}_{nb}^n \right) \\ &= \mathbf{R}_b^c \underbrace{(\mathbf{I} + \mathbf{S}(\delta \mathbf{a}))^\top}_{\mathbf{R}_{\hat{b}}^b} \left(\mathbf{R}_{\hat{b}}^n \right)^\top \left(\mathbf{p}_{nt}^n - \hat{\mathbf{p}}_{nb}^n - \delta \mathbf{p}_{nb}^n \right) \\ &= \mathbf{R}_b^c \left(\mathbf{R}_{\hat{b}}^n \right)^\top \left(\mathbf{p}_{nt}^n - \hat{\mathbf{p}}_{nb}^n - \delta \mathbf{p}_{nb}^n \right) \\ &\quad - \mathbf{R}_b^c \mathbf{S}(\delta \mathbf{a}) \left(\mathbf{R}_{\hat{b}}^n \right)^\top \left(\mathbf{p}_{nt}^n - \hat{\mathbf{p}}_{nb}^n - \delta \mathbf{p}_{nb}^n \right) \\ &= \mathbf{R}_b^c \left(\mathbf{R}_{\hat{b}}^n \right)^\top \left(\mathbf{p}_{nt}^n - \hat{\mathbf{p}}_{nb}^n - \delta \mathbf{p}_{nb}^n \right) \\ &\quad + \mathbf{R}_b^c \mathbf{S} \left(\left(\mathbf{R}_{\hat{b}}^n \right)^\top \left(\mathbf{p}_{nt}^n - \hat{\mathbf{p}}_{nb}^n - \delta \mathbf{p}_{nb}^n \right) \right) \delta \mathbf{a} \end{aligned} \tag{32}$$

where the final two expressions are equal. The last expression is beneficial when differentiating the measurement model with respect to $\delta\mathbf{a}$, and the first expression is beneficial when differentiating with respect to $\delta\mathbf{p}_{\text{nb}}^n$. Note that Eq. 32 needs to be inserted in Eq. 29 before the Jacobian is calculated. Finding the Jacobian theoretically is challenging, but is achievable with a computer program. Nevertheless, for real-time calculation, it is typically more efficient to find the Jacobian numerically.

In the SKF, the Jacobian is calculated with respect to the target position ($\mathbf{p}_{nt}^n$), the target velocity ($\mathbf{v}_{nt}^n$) and the UAV error states ($\delta\mathbf{p}_{nb}^n$, $\delta\mathbf{a}$). Hence, the Jacobian has two rows and ten columns where column three and four only have zeros because the measurement model does not depend on target velocity.

5 Field Experiments

The methods presented in Sections 2, 3 and 4 are evaluated using experimental data from several field tests. This section describes the UAV platform used to collect data, how the methods are evaluated, and several case studies presenting experimental results.

5.1 Experimental UAV Platform

A fixed-wing UAV with electrical propulsion (displayed in Fig. 2) was equipped with the following list of sensors for collecting relevant data:

– FLIR Tau2 thermal camera with a resolution of 640x512 pixels, focal length of 19 mm and frame rate of 7.5 Hz. It is a longwave infrared camera sensitive to wavelengths from 7.5 μm to 13.5 μm.
– A ThermalGrabber used to extract the digital image from the Tau2 thermal camera.
– Pixhawk autopilot running Arduplane software.
– SenTiBoard to synchronize the camera and navigation data [2].
– Odroid-XU4 as on-board computer for storing data.
– Analog Devices ADIS 16490 IMU measuring specific force and angular rate at 250 Hz
– Dual-antenna RTK-GNSS based on uBlox NEO-M8T receivers measuring position at 5 Hz.

Three independent field tests were carried out where the goal was to gather data for tracking of the vessel displayed in Fig. 3. Parts of the data from these tests are used in three independent case studies presented in Sections 5.3 to 5.5. The field experiments were carried out on the same day at Raudstein in Norway. The weather was windy and sunny during all flights. The flight paths were predefined without customizing the path based on the weather conditions or

Fig. 2 The NTNU Cruiser-Mini fixed-wing UAV with cruise speed of approximately 20 m/s

the target behavior. Therefore, the data reflect a surveillance operation at sea without prior information about targets being present.

5.2 Experimental Evaluation

The purpose of the case studies is to evaluate the methods presented in this work. Sensor data from IMU and GNSS were used to estimate the pose of the UAV as described in Section 2. Navigation results are not presented in this article but have been compared and validated with navigation data from the autopilot. Target detection [15, 17] was used to detect the target in thermal images using image processing. Target detections were included as measurements in the tracking system (Section 4) together with the camera pose. The entire pipeline is illustrated in Fig. 4. The green boxes

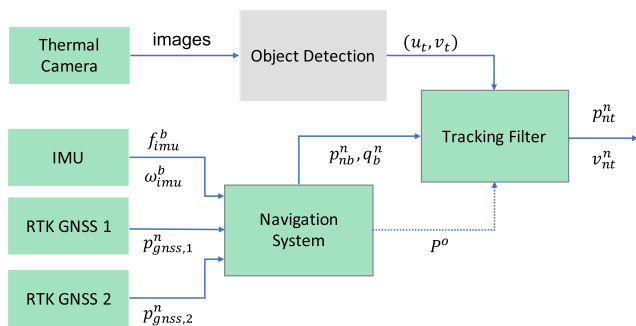**Fig. 3** The boat used as target in the experiments

Fig. 4 Illustration of subsystems and signal flow. Green boxes are covered in this research

illustrate the parts of the system that are covered in this article. Object detection is covered in [17]. Figure 5 shows a typical normalized thermal image of the target. The dark background shows the sea surface, and the bright spot is the target vessel. Figure 6 shows another raw normalized thermal image with both the sea surface and the shore.

The thermal images used in the forthcoming case studies only included one target in each image. Consequently, measurement association is not of importance for the case studies, and the filtering part of target tracking is key. The methodology can be extended to include multiple targets by using the Nearest Neighbor Standard Filter [4] for data association due to the strong signal-to-noise ratio in the images. Moreover, false positives are rare [17]. For increased robustness, the SKF combined with the Probabilistic Data Association Filter [4] is a viable strategy.

The methods have been analyzed in post processing, but real-time processing is also feasible on the on-board computer. The object detection algorithm has been executed in real-time previously [17]. Moreover, the execution time
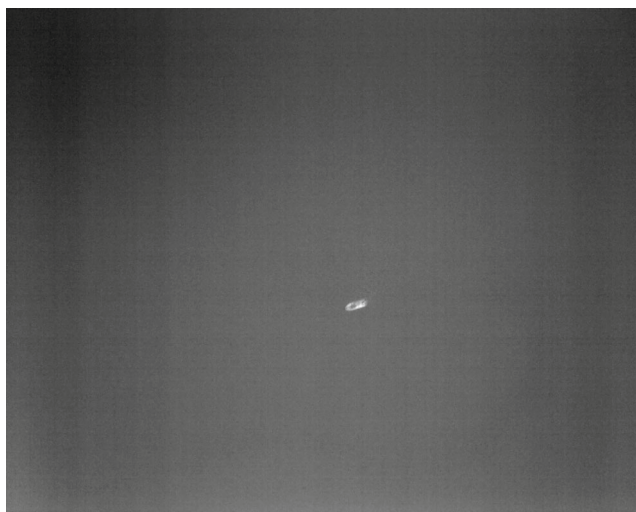
Fig. 5 Normalized raw thermal image with target (bright spot) and sea surface (background)
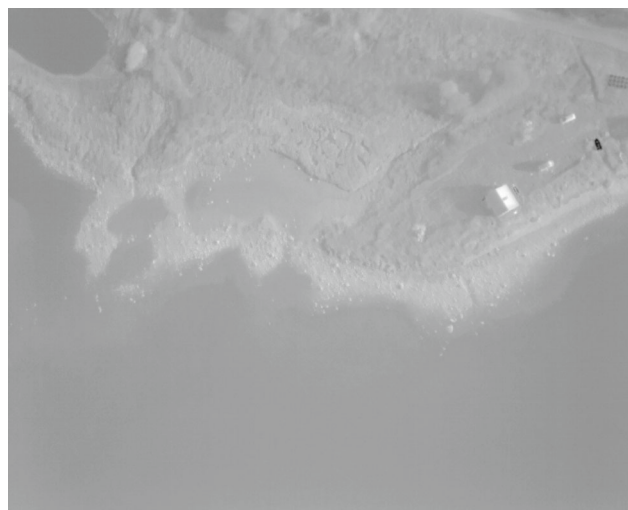
Fig. 6 Normalized raw thermal image with the sea surface (lower half) and the shore (upper half)

of the SKF is comparable with the extended Kalman filter (EKF). The only increase in processing time (compared to the EKF) is the calculation of the cross covariance, and the inclusion of the cross-covariance in the gain and target covariance update. In addition, the Jacobian is extended with additional elements, but is not producing a significant increase in execution time.

The SKF is compared with two alternative tracking filters in the forthcoming case studies. The first alternative is an EKF where navigation uncertainty is neglected. The Jacobian of Eq. 32 is only calculated with respect to the target states and not the navigation states. The second alternative is an EKF where navigation uncertainty is used to inflate the covariance of the target. This solution is based on Eqs. 24 and 25 where the Jacobian is calculated both with respect to the target and navigation states as specified in Eq. 15. It is denoted detuned EKF where detuned refers to the covariance inflation in Eq. 24. The main difference between the SKF and detuned EKF is the cross-covariance, which is maintained in the SKF. All filters have been initialized by using direct georeferencing [12] to find the north and east positions of the first available detection. The true states of the target were measured with a single-frequency GNSS receiver using standard GNSS service and used as reference. RTK corrections were not available for the target reference so the horizontal positioning accuracy is expected to be 2 m.

5.3 Case Study 1 - Tracking of High-Speed Vessel

The objective in the first case study is to detect and track a high-speed target. This case study is based on data from a field test where more than 8000 thermal images were captured in total together with navigation data from the

same period. The target was visible and have been detected successfully in 264 images within a chosen tracking period of three minutes. A subset of 1350 images were captured in the tracking period. The target was outside the field of view of the camera in the periods without detections. The UAV moved past the target and had to turn to observe the target again as shown in Fig. 7. Consequently, major parts of the estimates are solely based on prediction.

The covariance of the measurement noise in Eq. 28 was chosen as $(30\,\text{pixels})^2$ along both image frame axes. These values were chosen quite large to also fit the data in case studies 2 and 3. The measurement noise covariance could have been reduced significantly here, but a generic tuning was desired to prove that the system works in several scenarios, and for a fair comparison between the SKF and the alternatives. Therefore, it was not specially optimized for any of the case studies. All filters used the state transition model from Section 4.1 with σ_v from Eq. 27 equal to $(0.05\,\text{m/s}^2)^2$. The initial covariance in the north and east positions was chosen as $(10\,\text{m})^2$ and the corresponding velocities with covariance $(10\,\text{m/s})^2$. The initial covariance in velocity is not very far from the values proposed in [18] $\left(\frac{v_{\max}^2}{3}\right)$ for initialization based on a single position measurement. The true paths of the UAV and the target are displayed in Fig. 7. The UAV operated at an altitude of 350 m.

An interesting factor in the SKF and detuned EKF is the estimated covariance in the navigation states of the UAV. The mean estimated standard deviations for the north, east and down positions are 8 cm, 8 cm and 7 cm, respectively. The mean estimated standard deviations for the roll, pitch and yaw angles are 1.35°, 1.30° and 2.65°, respectively. The uncertainty in the position is obviously less influential than the attitude [12]. The estimated standard deviation is small
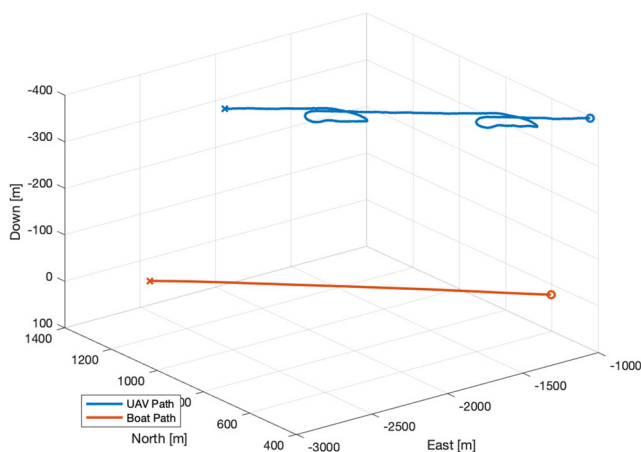
in position because position measurements were provided with RTK capability on the UAV.

The estimation errors in target position are shown in Fig. 8, and the main results of this case study are summarized in Table 1. The accuracy of the estimated position is comparable for the EKF and the SKF, but the SKF is more accurate overall. The mean estimation error is closer to zero and the average absolute estimation error is also smaller.

The estimated speed and course are displayed in Fig. 9. The SKF is more accurate initially, and this explains why the estimated positions of the EKF and detuned EKF drift more in the beginning (the first segment without measurements in Fig. 8).

The largest difference is seen in Fig. 10, which shows the normalized innovation squared (NIS) and normalized estimation error squared (NEES) in position [3]. NEES describes the estimation error squared normalized by the estimated covariance. In the same manner, the NIS describes the measurement innovation squared normalized by the innovation covariance. The NEES is above the upper confidence bound for the EKF in several time periods. This means that the EKF is too optimistic with a covariance estimate that is too small compared to the true estimation error. The SKF on the other hand has the opposite behavior. In fact, the SKF is too pessimistic, which means that the estimated covariance is larger than the corresponding estimation error. Therefore, the measurement noise covariance could have been reduced for the SKF in this case study. The detuned EKF falls somewhere in between but has a covariance that is too small in extended periods.

It is obviously possible to reduce the NEES in general by increasing the measurement noise covariance. However, the
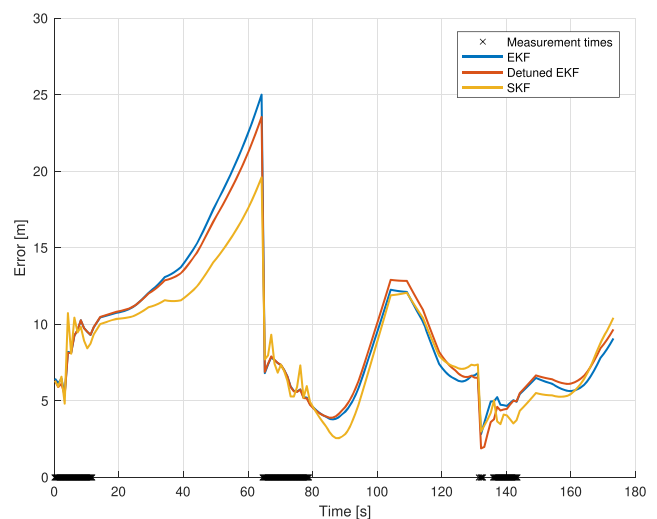

Fig. 7 UAV and target paths in case study 1

Fig. 8 Estimation error in target position

Table 1 Average performance metrics in case study 1

Parameter	EKF	Detuned EKF	SKF
ME position [north, east]	−[4.1 m, 1.2 m]	−[3.9 m, 1.2 m]	−[2.9 m, 1.3 m]
MAE estimated position	9.4 m	9.5 m	8.7 m
Mean NEES	7.90	5.09	0.40
Mean NIS	0.42	0.22	0.37

M(A)E = mean (absolute) error. NEES = normalized estimation error squared. NIS = normalized innovation squared

measurement noise standard deviation must be increased to 60 pixels (from 30) to keep the NEES within the confidence bounds for the EKF. This is much larger than the expected measurement noise from the object detection algorithm, and a solution that must be tailored to every new set of data. Consequently, it is not robust in the same sense as the SKF even though such a solution does not necessarily affect the accuracy negatively. The results presented in this case study are further discussed in Section 5.6.

5.4 Case Study 2 - Tracking of Slowly-Moving Vessel

The objective in the second case study is to track a slowly moving boat. This case study is based on data from another independent field test. The target was visible and have been detected successfully in 600 images (corresponds to 80 seconds of continuous detections) spread over a tracking period of 750 seconds (5625 images). The detections are spread into 11 segments as shown in Fig. 12. The target was outside the field of view of the camera in the periods without detections. All filters were tuned in the same manner as in the previous case study. The true paths of the UAV and the
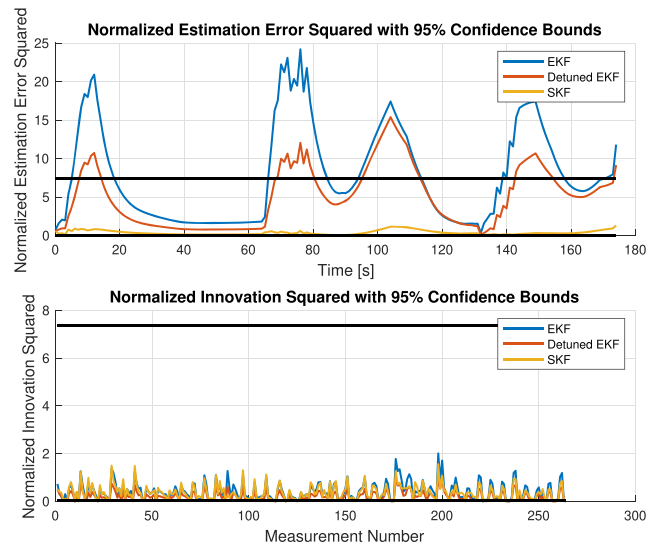
Fig. 10 Normalized estimation error squared and normalized innovation squared

vessel are displayed in Fig. 11. The UAV operated at an altitude of 300 m in a figure-eight pattern. The attitude of the UAV varied more in this field test.

The mean estimated standard deviations for the UAV north, east and down positions are 43 cm, 43 cm and 68 cm, respectively. The mean estimated standard deviations for the roll, pitch and yaw angles are 1.22°, 1.24° and 2.63°, respectively. The standard deviation in position is larger than in the previous case study because the IMU was weighted more in the navigation filter. This is further explained by the flight pattern in Fig. 11, which was dominated by turns.

The estimation errors in target position are shown in Fig. 12. The main results of this case study are summarized in Table 2. The SKF is the most accurate filter. The EKF and detuned EKF drift in periods without measurements while the SKF predicts the position more accurately.
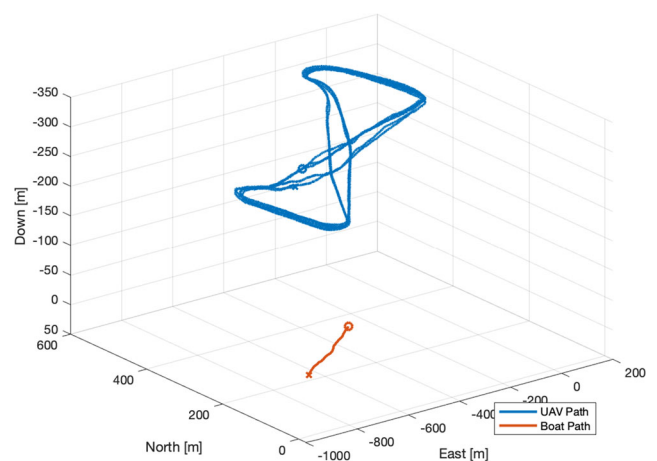

Fig. 9 Estimated speed and course

Fig. 11 The paths of the UAV and target in case study 2
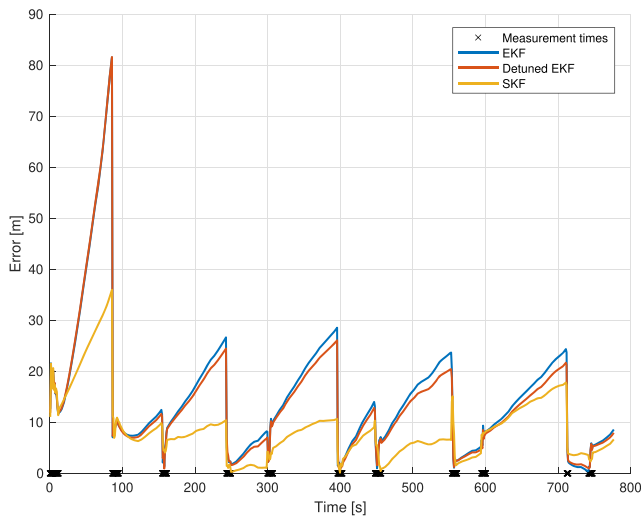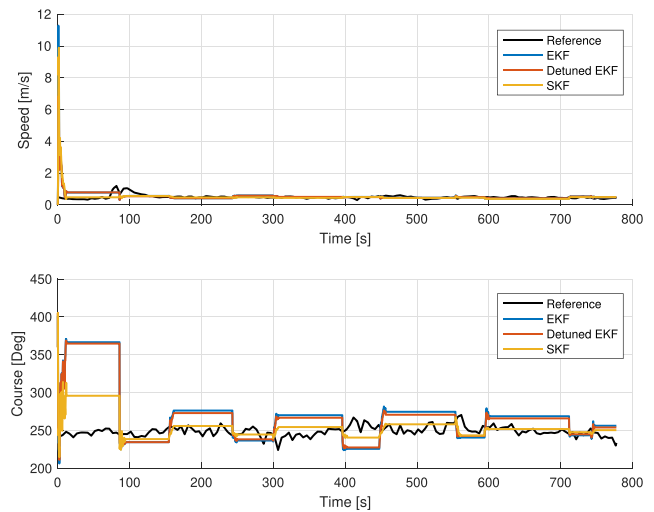
Fig. 12 Estimation error in target position

Fig. 13 Estimated speed and course

The estimated speed and course are displayed in Fig. 13. The EKF and detuned EKF manage to estimate the total speed accurately, but the estimated course is wrong and never converges in the same sense as for the SKF. This explains why the position estimates drift more for the EKF and detuned EKF, and similar behavior was also observed in the previous case study. Since all filters have the same state transition model and the same set of measurements, one can conclude that the SKF is more accurate in this case study.

The NIS and NEES (in position) are displayed in Fig. 14. The NIS is similar for both the SKF and EKF, but smaller for the detuned EKF. A few measurements are outside of the confidence bounds, but that is not necessarily an issue for consistency. The largest NIS values appear when it is a significant period between two consecutive measurements, which is the expected behavior.

The NEES is not similar. The EKF and detuned EKF have NEES values that are much larger than the upper confidence bound for major parts of the tracking period. This means that they are too optimistic, and that the estimated covariance in position is much smaller than the corresponding

estimation error. The SKF on the other hand is within the confidence bounds except for a short initial period. This supports the findings in the previous case study and shows that the SKF can estimate the covariance more precisely when the other solutions struggle.

5.5 Case Study 3 - Tracking Of Drifting Vessel

The third and final case study concerns tracking of a drifting vessel. This is based on data from a third independent field test. The target was visible and successfully detected in 2400 images (corresponds to 320 seconds with continuous detection) over tracking period of 1500 seconds (11250 images). The detections are spread into several segments as shown in Fig. 16. The target was outside the field of view

Table 2 Average performance metrics in case study 2

Parameter	EKF	Detuned EKF	SKF
ME position [north, east]	[10.5 m, 1.8 m]	[10.0 m, 1.6 m]	[4.4 m, 0.1 m]
MAE estimated position	15.0 m	14.2 m	8.8 m
Mean NEES	17.8	11.5	1.07
Mean NIS	1.84	1.08	1.84

M(A)E = mean (absolute) error. NEES = normalized estimation error squared. NIS = normalized innovation squared
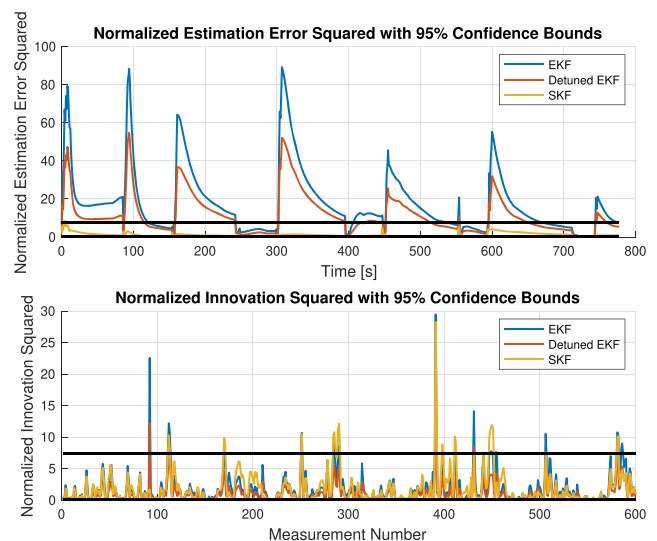

Fig. 14 Normalized estimation error squared and normalized innovation squared

of the camera in the periods without detections. All filters were tuned in the same way as previously. The true paths of the UAV and the target are displayed in Fig. 15. The UAV operated at an altitude of 200 m to 300 m in a figure-eight pattern.

The mean estimated standard deviations for the estimated UAV north, east and down positions are 87 cm, 87 cm and 1.42 m, respectively. The mean estimated standard deviations for the UAV roll, pitch and yaw angles are 1.73°, 1.78° and 2.48°, respectively.

The estimation errors in target position are shown in Fig. 16. The main results of this case study are summarized in Table 3. The performance, with respect to tracking accuracy, is comparable for all filters. The EKF and detuned EKF have mean estimation error closer to zero, but the SKF has the smallest average absolute estimation error. The average absolute estimation error is smaller in this case study overall. That can both be because of the reduced altitude of the UAV and the vessel speed, which was smaller during this field test. The large initial estimation error is due to the large initial covariance in velocity and the poor accuracy in the first few measurements. Moreover, only a small set of measurements was available initially, so the filters were not converging until new measurements arrived. A small set of initial measurements is particularly challenging when estimating the velocities which causes drift in position rapidly. Note that the vessel only moved approximately 200 m during the tracking period.

Figure 17 shows the estimated speed and course. The estimated speed is a fraction smaller for the SKF and closer to the reference. All filters struggle to estimate the correct course. The SKF converges somewhat halfway into the tracking period and is more accurate than the EKF and detuned EKF. Note that the speed of the target was so small that the course reference from the GNSS receiver on the target could be inaccurate at times and not properly defined.
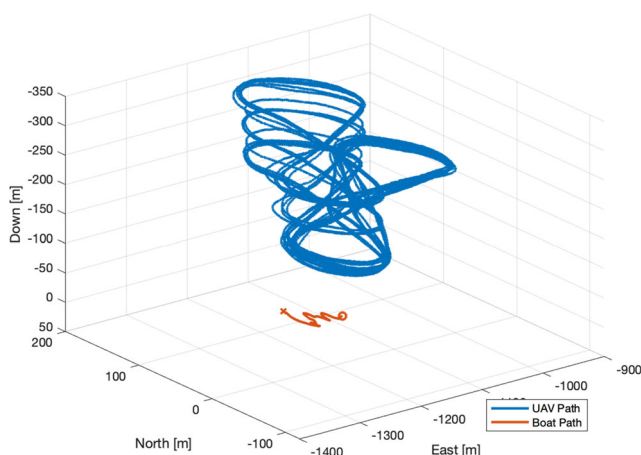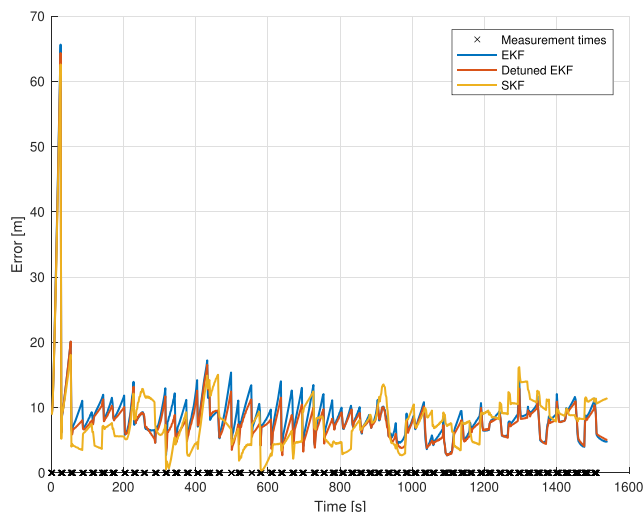

Fig. 16 Estimation error in target position

The course reference could have been smoothed, but the raw data were chosen because it has been used in the previous case studies. Nevertheless, these results show that the SKF estimates the speed and course more accurately, but that it does not affect the accuracy of the estimated position significantly because of the total number of measurements available in this case study.

Figure 18 shows NIS and NEES (in position) for all filters. The NEES is almost never within the confidence bounds for the EKF and detuned EKF, which means that the estimated covariance is too small. This is a weakness observed in the previous case studies. The SKF on the other hand, has NEES within the confidence bounds for most of the time. This case study also shows a significant difference in NIS. The NIS is too large for the EKF for several measurements, and must be explained by a small, estimated innovation covariance. Thus, the covariance of the measurement noise should have been increased for the EKF. The SKF also has a few measurements outside of the confidence bound. The NIS and NEES are interesting for the detuned EKF. The small NIS values indicate that

Fig. 15 The paths of the UAV and target in case study 3

Table 3 Average performance metrics in case study 3

Parameter	EKF	Detuned EKF	SKF
ME position estimates [north, east]	[−3.3 m, 3.6 m]	[−3.4 m, 3.5 m]	[−4.8 m, 3.1 m]
MAE estimated position	8.8 m	8.1 m	8.0 m
Mean NEES	40.4	19.1	2.99
Mean NIS	1.75	0.80	1.09

M(A)E = mean (absolute) error. NEES = normalized estimation error squared. NIS = normalized innovation squared
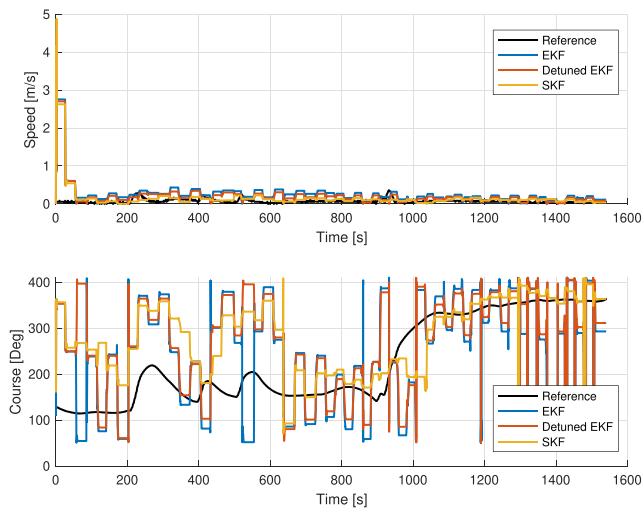
Fig. 17 Estimated speed and course

the innovation covariance is too large. However, the NEES values indicate that the covariance is too small compared to the estimation error. The same effect is not observed for the SKF. Consequently, the cross-covariance between the navigation and tracking estimates, which is maintained in the SKF, has a positive effect for consistency and shows the advantage of the SKF. Similar behavior is observed in case studies 1 and 2, but in a less significant manner.

5.6 Discussion

A comprehensive amount of data has been used to evaluate the SKF in case studies 1-3. A few conclusions can be drawn from the results:
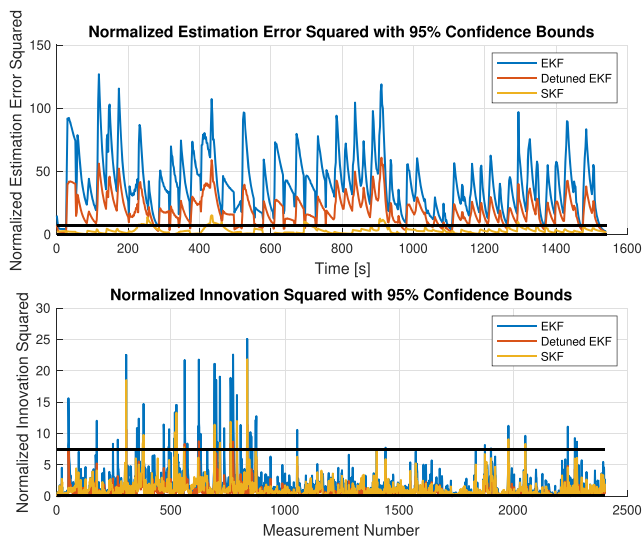
Fig. 18 Normalized estimation error squared and normalized innovation squared

– The SKF is more accurate and is especially better when it comes to estimating the speed and course.
– The position estimates with the EKF and detuned EKF are accurate, but are weak for long periods solely based on prediction.
– The most significant difference is observed for the NEES where the EKF and detuned EKF are overconfident. The estimated covariance is much smaller than the corresponding estimation error, which is an issue for data association in multi-target tracking. It is possible to increase covariance of the noise in the motion and measurement models, but it is hard to generalize such a solution to fit new data. Moreover, this solution does not fit well with the NIS values. The SKF works for a fixed set of parameters, and is more robust and reliable with respect to NEES.
– Even though the SKF is superior in some means, the necessary covariance of the measurement noise is still larger than the expected uncertainty in the object detection algorithm [15]. Consequently, the errors affecting the tracking system are not mitigated perfectly. This can be caused by other error sources than navigation errors and effects not mitigated by the SKF. That can for instance be time-varying biases with a small time constant or unknown timing errors in sensor data.
– The same tuning was used in all case studies and different results were achieved. The tuning fitted the data in the final case study better and it highlights the challenge when tracking unknown targets. The behavior is unknown beforehand, so it is impossible to find parameters that are ideal in many different scenarios unless multiple model approaches are considered. Nevertheless, it is possible to identify values that are acceptable in several scenarios as shown in this research.
– The covariance of the navigation estimates significantly affects the performance. This is particularly observable in the NEES where the major difference is caused by the navigation uncertainty and how the cross covariance is handled. Moreover, the attitude is the critical factor and a mean estimated standard deviation of about 1° to 2° in roll, pitch and yaw seems to have the desired effect. The influence of the position uncertainty is negligible unless the standard deviation is large (5 meters or greater).

6 Conclusions

This research article has investigated tracking of floating objects using fixed-wing UAVs with a monocular thermal camera. Uncertainty in the UAV position and attitude has

been incorporated in the tracking system with the Schmidt-Kalman filter. This design has been used to mitigate issues related to consistency that often occur in this type of UAV mission. A large amount of experimental data has been gathered and analyzed to compare several methods. The results demonstrate that the SKF performs better with respect to consistency and estimation accuracy. Moreover, the position of floating objects is estimated with a mean error below 10 m with the SKF when the UAV operated at altitudes of 200 m to 350 m. Similar performance is also demonstrated in extended periods without target measurements and shows the reliability of the system.

Author Contributions This article has four authors, and their names and contributions are: Dr. Håkon Hagen Helgesen wrote the first version of the manuscript, planned, prepared and participated in the experiments (data collection), implemented and derived the tracking system, and performed the analysis of the results. Associate prof. Torleiv H. Bryne prepared the navigation sensors, participated in the experiments, and implemented the navigation system. Dr. Erik F. Wilthil derived the theoretical foundation of the tracking system together with Håkon. Prof. Tor Arne Johansen led the research effort and provided valuable suggestions and discussions before conducting experiments and in the process of writing the manuscript. All authors have commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability Data available upon request.

Declarations

Ethics approval and consent to participate This paper does not report research that requires ethical approval. Consent to participate or consent to publish statements are accordingly also not required.

Competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Al-Isawi, M.M., Sasiadek, J.Z.: Guidance and control of a robot capturing an uncooperative space target. J. Intell. Robot. Syst. **93**(3), 713–721 (2019)

2. Albrektsen, S.M., Johansen, T.A.: Syncboard - a high accuracy sensor timing board for uav payloads. In: The International Conference on Unmanned Aircraft Systems, pp. 1706–1715 (2017). https://doi.org/10.1109/ICUAS.2017.7991410

3. Bar-Shalom, Y., Li, X.R., Kirubarajan, T.: Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. Wiley, New York (2004)

4. Bar-Shalom, Y., Willett, P.K., Tian, X.: Tracking and Data Fusion: A handbook of Algorithms. YBS Publishing Storrs, USA (2011)

5. Beard, R.W., McLain, T.W.: Small Unmanned Aircraft: Theory and Practice. Princeton University Press, Princeton (2012)

6. Brekke, E.F., Wilthil, E.F.: Suboptimal kalman filters for target tracking with navigation uncertainty in one dimension. In: 2017 IEEE Aerospace Conference, pp. 1–11 (2017). https://doi.org/10.1109/AERO.2017.7943601

7. Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J.J.: Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. IEEE Trans. Robot. **32**(6), 1309–1332 (2016). https://doi.org/10.1109/TRO.2016.2624754

8. Farrell, J. Aided Navigation: GPS with High Rate Sensors, 1st edn. McGraw-Hill Inc, New York (2008)

9. Fossen, T.: Handbook of Marine Craft Hydrodynamics and Motion Control. Wiley, New York (2011)

10. Fossen, T. Handbook of Marine Craft Hydrodynamics and Motion Control, 2nd edn. Wiley, New York (2021)

11. Heikkila, J., Silven, O., et al.: A four-step camera calibration procedure with implicit image correction. In: CVPR, vol. 97, p. 1106. Citeseer (1997)

12. Helgesen, H.H., Leira, F.S., Bryne, T.H., Albrektsen, S.M., Johansen, T.A.: Real-time georeferencing of thermal images using small fixed-wing uavs in maritime environments. ISPRS J. Photogramm. Remote. Sens. **154**, 84–97 (2019). https://doi.org/10.1016/j.isprsjprs.2019.05.009

13. Helgesen, H.H., Leira, F.S., Fossen, T.I., Johansen, T.A.: Tracking of ocean surface objects from unmanned aerial vehicles with a pan/tilt unit using a thermal camera. J. Intell. Robot. Syst. https://doi.org/10.1007/s10846-017-0722-3 (2017)

14. Helgesen, H.H., Leira, F.S., Johansen, T.A.: Colored-noise tracking of floating objects using uavs with thermal cameras. In: The International Conference on Unmanned Aircraft Systems (2019)

15. Helgesen, H.H., Leira, F.S., Johansen, T.A., Fossen, T.I., Pettersen, K.Y., Nijmeijer, H.: Detection and tracking of floating objects using a uav with thermal camera. In: Fossen, T.I. (ed.) Sensing and Control for Autonomous Vehicles: Applications to Land, Water and Air Vehicles, pp. 289–316. Springer International Publishing (2017). https://doi.org/10.1007/978-3-319-55372-6_14

16. Kanellakis, C., Nikolakopoulos, G.: Survey on computer vision for uavs: Current developments and trends. J. Intell. Robot. Syst. **87**(1), 141–168 (2017)

17. Leira, F.S., Johansen, T.A., Fossen, T.I.: Automatic detection, classification and tracking of objects in the ocean surface from uavs using a thermal camera. In: 2015 IEEE Aerospace Conference, pp. 1–10 (2015). https://doi.org/10.1109/AERO.2015.7119238

18. Mallick, M., La Scala, B.: Comparison of single-point and two-point difference track initiation algorithms using position measurements. Acta Autom. Sin. **34**(3), 258–265 (2008)

19. Markley, F.L.: Attitude error representations for kalman filtering. J. Guid. Control Dyn. **26**(2), 311–317 (2003). https://doi.org/10.2514/2.5048

20. Mullane, J., Vo, B., Adams, M.D., Vo, B.: A random-finite-set approach to bayesian slam. IEEE Trans. Robot. **27**(2), 268–282 (2011)

21. Novoselov, R.Y., Herman, S.M., Gadaleta, S.M., Poore, A.B.: Mitigating the effects of residual biases with schmidt-kalman filtering. In: 2005 7th International Conference on Information Fusion (2005). https://doi.org/10.1109/ICIF.2005.1591877

22. Pajares, G.: Overview and current status of remote sensing applications based on unmanned aerial vehicles (uavs). Photogramm. Eng. Remote Sens. **81**(4), 281–329 (2015). https://doi.org/10.14358/PERS.81.4.281

23. Prasad, D.K., Rajan, D., Rachmawati, L., Rajabally, E., Quek, C.: Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. IEEE Trans. Intell. Transp. Syst. **18**(8), 1993–2016 (2017). https://doi.org/10.1109/TITS.2016.2634580

24. Li, R.X., Jilkov, V.P.: Survey of maneuvering target tracking. part i. dynamic models. IEEE Trans. Aerosp. Electron. Syst. **39**(4), 1333–1364 (2003). https://doi.org/10.1109/TAES.2003.1261132

25. Schmidt, S.F.: Applications of state space methods to navigation problems. Adv. Control Syst. **3**, 293–340 (1966)

26. Silva, A., Mendonça, R., Santana, P.: Monocular trail detection and tracking aided by visual slam for small unmanned aerial vehicles. J. Intell. Robot. Syst. **97**(3), 531–551 (2020)

27. Sola, J., Quaternion kinematics for the error-state kf. http://www.iri.upc.edu/people/jsola/JoanSola/objectes/notes/kinematics.pdf. Accessed 02 Feb 2018 (2017)

28. Urzua, S., Munguía, R., Grau, A.: Monocular slam system for mavs aided with altitude and range measurements: a gps-free approach. J. Intell. Robot. Syst. **94**(1), 203–217 (2019)

29. Wang, C.C., Thorpe, C., Thrun, S., Hebert, M., Durrant-Whyte, H.: Simultaneous localization, mapping and moving object tracking. Int. J. Robot. Res. **26**(9), 889–916 (2007). https://doi.org/10.1177/0278364907081229

30. Wang, D., Xu, C., Yuan, P., Huang, D.: A revised monte carlo method for target location with uav. J. Intell. Robot. Syst. **97**(2), 373–386 (2020)

31. Wilthil, E.F., Brekke, E.F.: Compensation of navigation uncertainty for target tracking on a moving platform. In: 2016 19th International Conference on Information Fusion (FUSION), pp. 1616–1621 (2016)

32. Yang, C., Blasch, E., Douville, P.: Design of schmidt-kalman filter for target tracking with navigation errors. In: 2010 IEEE Aerospace Conference, pp. 1–12 (2010). https://doi.org/10.1109/AERO.2010.5446689

Håkon Hagen Helgesen received the MSc degree in 2015 and the PhD degree in 2019, both in Engineering Cybernetics, from the Norwegian University of Science and Technology (NTNU), Trondheim, Norway. He currently works as a researcher in Engineering Cybernetics at NTNU and is affiliated with the NTNU Centre for Autonomous Marine Operations and Systems. He works with topics related to inertial navigation, SLAM, target tracking and control systems for unmanned aerial vehicles and surface vessels.

Torleiv H. Bryne received the MSc and PhD degrees in Engineering Cybernetics from the Norwegian University of Science and Technology (NTNU), Department of Engineering Cybernetics, in 2013 and 2017. He worked as a research scientist at SINTEF before he became an associate professor at the Department of Engineering Cybernetics, NTNU, in 2020. His research interests are in the field estimation applied to navigation systems. Unmanned aerial vehicles and marine applications are the main focus areas of his research.

Erik F. Wilthil received his M.Sc. and PhD degrees in Engineering Cybernetics from NTNU, Norway in 2015 and 2019 on target tracking for autonomous surface vehicles. Since 2019, he has been a Postdoc at NTNU and is affiliated with the NTNU Centre for Autonomous Marine Operations and Systems. His research focus is on sensor fusion and situational awareness for urban passenger ferries.

Tor Arne Johansen received the MSc degree in 1989 and the PhD degree in 1994, both in electrical and computer engineering, from the Norwegian University of Science and Technology, Trondheim, Norway. From 1995 to 1997, he worked at SINTEF as a researcher before he was appointed Associated Professor at the Norwegian University of Science and Technology in Trondheim in 1997 and Professor in 2001. He has published several hundred articles in the areas of control, estimation and optimization with applications in the marine, aerospace, automotive, biomedical and process industries. In 2002 Johansen co-founded the company Marine Cybernetics AS where he was Vice President until 2008. Prof. Johansen received the 2006 Arch T. Colwell Merit Award of the SAE, and is currently a principal researcher within the Center of Excellence on Autonomous Marine Operations and Systems (NTNU-AMOS) and director of the Unmanned Aerial Vehicle Laboratory at NTNU.