Colored-Noise Tracking of Floating Objects using UAVs with Thermal Cameras

Håkon Hagen Helgesen[1] Frederik Stendahl Leira[1] Tor Arne Johansen[1]

***Abstract*— Tracking of floating objects using a fixed-wing UAV equipped with a thermal camera requires precise knowledge about the position and attitude of the UAV. Errors in the navigation estimates reduce the accuracy of the tracking system. Navigation errors are usually correlated in time and can propagate colored noise into the tracking filter. This work analyzes two approaches that seek to mitigate colored noise and they are compared experimentally with a third approach which assumes that the noise in the tracking system is purely white. Two independent flight experiments have been carried out where a small marine vessel was used as target. Thermal images of the target were captured and the position and velocity of the target have been estimated in an Earth-fixed coordinate system only using the images. The results show that the tracking filters perform well when considering the estimation error, but that colored noise should be accounted for to maintain filter consistency.**

I. Introduction

Monitoring the sea surface is important in several applications and industries. Restricting access for foreign objects to certain areas is crucial in operations with high demands for safety. This can for instance be in operations where autonomous underwater vehicles need to approach the sea surface without the risk of facing other vessels or obstacles on the sea surface. It is also important for autonomous ships which need situational awareness to operate safely, and seismic surveys where maneuvering should be minimized.

Covering large geographical areas is challenging with stationary sensors or sensors mounted on slowly moving vessels at sea. Satellite images can provide information about large areas and be used to detect ships [1], but have challenges related to delay in the transmitted data, bandwidth restrictions and low temporal and spatial resolution. A suitable option is to use aircraft equipped with optical sensors to monitor larger areas [2]. Manned aircraft have traditionally been used to monitor coastlines but is expensive. Fixed-wing unmanned aerial vehicles (UAVs) are a cheaper alternative for monitoring large areas in real-time with the required temporal and spatial resolution [3].

UAVs are used in many types of operations and much research is focused on how to utilize and operate UAVs in a safe manner. To meet the safety requirements by the authorities, research on robust and safe operations is needed. This includes topics such as navigation [4]–[6], navigation in GPS-denied environments [7], [8], collision avoidance [9] and autonomous take-off and landing [10]. From the application-based point of view, popular topics are target tracking [11], [12], inspection [13] and agriculture [14].

When a UAV is used for monitoring, a sensor with capability to image the area is the heart of the sensor suite. Optical sensors are the obvious choice when monitoring an area from the air, but active sensors such as Radars and Lidars might also be more common on UAVs in the future. The imaging sensor in this work is a thermal camera. A thermal camera is chosen because floating objects usually have a distinct thermal signature which can be distinguished from the sea surface [15]. Detecting floating structures in thermal images is a task that requires image processing. Object detection in visual spectrum images is covered thoroughly in the literature and algorithms based on deep learning are dominating [16], [17]. However, object detection using neural networks is computationally expensive and not necessarily applicable in real-time applications with limited processing power. Thus, simpler techniques that can be executed in real-time are necessary. Object detection in thermal images is used in this work, and covered in several articles [18]–[20].

Detecting an object in one image is not sufficient for situational awareness. It is necessary to monitor foreign objects when they are within the surveillance region and predicting their motion whenever they are not observed. This task falls within the field of target tracking. Data association, which is association of new measurements with existing tracks, is not covered in this work. In other words, tracking of a single object where it is known which measurements belong to the object is the focus here. Note that a situation where multiple targets are present, but not in close proximity, can also be covered with the same methodology.

Tracking of floating objects using optical sensors is challenging because optical sensors only provide bearing information. In general, it is not possible to compute 3-dimensional Earth-fixed coordinates from a single image where only two pixel coordinates are known. However, since the sea surface can be approximated as a planar flat surface at known altitude, it is possible to calculate both the north and east positions of objects. The third coordinate is known since it is assumed that objects operate on the sea surface.

Georeferencing is covered in detail in several articles [15], [21], and is used to transform pixel coordinates in images to coordinates in an Earth-fixed coordinate system (e.g. north-east). Georeferencing requires precise knowledge of the position and orientation of the camera, and, thus the position and attitude of the UAV. Errors in the navigation states of the UAV lead to significant errors in the georeferenced north

[1]Håkon Hagen Helgesen, Frederik Stendahl Leira and Tor Arne Johansen are with NTNU Centre for Autonomous Marine Operations and Systems, Department of Engineering Cybernetics, Norwegian University of Science and Technology, O. S. Bragstads plass 2D, 7491 Trondheim, Norway hakon.helgesen@ntnu.no

and east positions. The most common tracking filters are based on measurement models where the measurement error is modeled as Gaussian white noise. This makes sense if the error is not systematic. The uncertainty in the detected object position in a thermal image can normally be accepted as white. However, in georeferencing, the pixel coordinates are converted to an Earth-fixed frame based on the navigation states of the UAV. This is a converted measurement approach and the error is usually correlated for consecutive images because the error in the navigation estimates is often correlated [22]. Therefore, this article considers how colored or correlated measurement noise can be mitigated in a tracking filter based on georeferencing. This is to the authors best knowledge not studied experimentally previously.

A challenge in target tracking is to predict the motion of objects when they are outside the field of view of the sensor. Prediction depends on a predefined motion model, which is based on the expected motion or a priori information about the target behavior. Deviations from the motion model are modeled as Gaussian white noise, but colored noise might be present in the motion model as well, especially during target maneuvers. Therefore, how correlations in the process noise affect the tracking filter is also investigated in this work.

Correlated noise in the motion and measurement models might not be critical in terms of the root-mean-square estimation error. However, it is important when estimating the covariance of the state estimates and accurate covariance is vital for filter consistency [23]. Moreover, consistency is crucial for data association in multi-target tracking, but also for a single target since it says something about the reliability of the state estimates. Therefore, analyzing these challenges for a single object is of interest.

This article investigates tracking of a single target in the presence of colored noise. Clutter or false measurements are not considered and not experienced to be major issues when detecting objects in thermal imagery. The goal is to estimate the position and velocity of the target in an Earth-fixed coordinate system based on the pixel location in thermal images captured from an agile fixed-wing UAV. Motion and measurement models for this particular application are discussed and explained. Strategies for how to mitigate colored noise are highlighted. Moreover, the importance of these issues is demonstrated with results from two UAV experiments carried out with a fixed-wing UAV where a marine vessel was used as target.

The rest of this article is divided in six parts. The first part describes tracking filters in general and consistency criteria. Typical motion models for floating objects are covered in the second part. The third part describes the measurement model. The fourth part describes the experiments carried out to investigate this application. The fifth part presents and discusses the results before the work is concluded in the final part.

II. Single Target Tracking

Target tracking can be divided into three different subproblems. The first task involves extracting a measurement of where targets are located in the sensor frame. This is called target detection. The second task is to associate new measurements with existing tracks. This is called data or measurement association and is key in multi-target scenarios or in situations where clutter and false alarms are expected. The final part, which is the main focus in this work, is the filtering part where the goal is to estimate the position and velocity of the target using target detections.

The filtering part of target tracing can be solved with both deterministic and statistical estimators. The stochastic methods are dominating the tracking literature because the Bayesian framework is key in data association. Thus, the Kalman filter or variations such as the extended Kalman filter (EKF), the particle filter and other related methods are often used to estimate the target states [24].

A standard Kalman filter is studied in this work to keep the tracking filter simple. The goal is to investigate the filtering problem when the only imaging sensor is a thermal camera mounted in a fixed-wing UAV. This is not straightforward and the main challenges can be explored with a Kalman filter without considering more advanced state estimators. The Kalman filter is often divided into a predictor and a corrector. The predictor predicts the target motion based on a predefined motion model and the corrector is used whenever a new measurement is available to correct the prediction. Both estimates of the target states and the covariance of the estimation error are calculated by the Kalman filter. Accurate covariance estimation is especially important for filter consistency and key in multi-target tracking.

The Kalman filter is the optimal estimator for linear systems where the process and measurement noise are Gaussian white noise. Such ideal conditions cannot be expected in practice when maneuvering targets are tracked with optical sensors mounted in UAVs. Therefore, inconsistency might occur. A state estimator is said to be consistent if the following criteria are fulfilled [23]:

1) The state errors should be acceptable as zero mean and have magnitude commensurate with the state covariance as calculated by the filter.
2) The innovations should also have the same property.
3) The innovations should be acceptable as white.

The innovation is the difference between the actual measurement and the predicted measurement in the tracking filter.

Only the second and third criteria can be verified for experimental data because the state errors are not known perfectly. The Kalman filter assumes that consecutive measurement innovations are uncorrelated. However, this is not expected in practice because navigation errors are often correlated and affect consecutive images in the same manner. The third criterion can be checked experimentally through autocorrelation and the second criterion can be verified with the normalized innovation squared (NIS), which is defined as: [23]

$$\epsilon_\nu(k) = \boldsymbol{\nu}(k)^\top \mathbf{S}(k)^{-1} \boldsymbol{\nu}(k) \tag{1}$$

where $\boldsymbol{\nu}$ is the innovation and $\mathbf{S}$ is the innovation covariance matrix calculated in the Kalman filter. ϵ_ν should be chi-

square distributed with n_z degrees of freedom, where n_z denotes the number of elements in the measurement vector, to comply with the second criterion. The first criterion can be verified with the normalized estimation error squared (NEES), which is defined as

$$\epsilon(k) = \tilde{\mathbf{x}}(k|k)^\top \mathbf{P}(k|k)^{-1} \tilde{\mathbf{x}}(k|k) \tag{2}$$

where $\tilde{\mathbf{x}}(k|k)$ is the a posteriori estimation error and $\mathbf{P}(k|k)$ is the a posteriori covariance matrix. The NEES should be chi-square distributed with n_x degrees of freedom, where n_x is the dimension of the state vector, to comply with the first criterion. Note that this measure is only valid if the true state is known accurately so care must be taken when it is used for experimental data.

If the motion or measurement models are nonlinear, a nonlinear state estimator such as the EKF should be used. Consistency is equally important for the EKF and thus the discussion later is also valid for nonlinear tracking filters.

III. Motion Models

When predicting the next state in the Kalman filter, a motion model is used as foundation for prediction. The motion model describes how the target is expected to move and measurements are used to correct for deviations from the motion model. A survey of motion models for maneuvering targets is presented in [25].

Marine vessels are usually not maneuvering significantly. Therefore, it is common to represent the expected motion of a target with a decoupled constant velocity (also called white noise acceleration) model. This model is defined directly in discrete-time as

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{F}\mathbf{x}(k) + \Gamma\mathbf{v}(k) \\ \mathbf{F} &= \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \frac{1}{2}T^2 & 0 \\ 0 & \frac{1}{2}T^2 \\ T & 0 \\ 0 & T \end{bmatrix} \end{aligned} \tag{3}$$

where T is the sampling time and the covariance of the process noise $\mathbf{v}$ is calculated as

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

The first two states are the 2D horizontal position in an Earth-fixed frame (e.g. north and east coordinates) and the two latter are the velocities. The motion is assumed to be decoupled between the coordinate dimensions. σ_v is a design parameter and should be in the order of the maximum expected acceleration [23]. A larger value indicates that the constant velocity assumption is violated more significantly than what a smaller value indicates. It should be adjusted based on the target of interest. The target control input is neglected in this model and assumed to be white noise and this assumption is obviously violated during maneuvers. Moreover, decoupling between the coordinates is also a simplification, but works well in general. The advantage with this model is the simplicity and the main drawback is that maneuvers rarely can be modeled as white noise since they are correlated in time.

The Singer acceleration model [25] is an alternative to the constant velocity model where the acceleration is instead modeled as a zero-mean stationary Markov process with autocorrelation

$$E[a(t+\tau)a(t)] = \sigma^2 e^{-\alpha|\tau|} \tag{5}$$

Since acceleration is added to the state vector, it can account for correlations in the motion model. The Singer model is a continuous-time model and is defined as (only one coordinate dimension shown for simplicity):

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\alpha \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} w(t) \tag{6}$$

where the state vector $\mathbf{x}$ consists of the position, velocity and acceleration in the corresponding dimension. The model can obviously be extended to cover both of the horizontal coordinates. α is the inverse of the expected maneuver time constant. Low values indicate a long maneuver and vice versa. Marine vessels rarely change course so a small value is perhaps the most appropriate choice, but depends on the type of ship. σ is the variance of the acceleration and commonly modeled as

$$\sigma^2 = \frac{a_M^2}{3}(1 + 4p_M - p_0) \tag{7}$$

where a_M is the maximum expected acceleration, p_0 is the probability of zero acceleration and p_M is the probability of the target accelerating (or decelerate) at the maximum rate. The system must be discretized for implementation and the discrete-time process noise has covariance matrix equal to

$$\mathbf{Q} = 2\alpha\sigma^2 \begin{bmatrix} T^5/20 & T^4/8 & T^3/6 \\ T^4/8 & T^3/3 & T^2/2 \\ T^3/6 & T^2/2 & T \end{bmatrix} \tag{8}$$

Nonlinear motion models that account for turns or other maneuvers have been considered, but not included in this work because marine vessels mostly maintain a constant course. Therefore, if turns are also considered, multiple-model approaches must be used and that is not assessed to be necessary or beneficial for the purpose of this work.

IV. Measurement Model

The measurement model depends on the sensor observing the target. A survey of measurement models is presented in [26]. A thermal camera is used in this work and target detections are provided through the pixel position of the target in the image. Since the tracking filter estimates the target position and velocity in an Earth-fixed frame, the measurement model is responsible for relating the detected pixel position to the Earth-fixed frame.

As mentioned in the introduction, georeferencing is a tool for calculating the north and east positions of a pixel based on the navigation states of the UAV and the camera intrinsic matrix [21]. The major advantage with georeferencing is that the pixel position of a target is converted to a position measurement in north and east. This simplifies the structure of the tracking filter since the measurement model becomes linear and a regular Kalman filter can be used. The drawback is that correlated noise will be introduced through georeferencing since errors in georeferencing are dominated by errors in the navigation states of the UAV. Another drawback is that it is difficult to choose the covariance of the measurement noise since it cannot be connected directly to the uncertainty in the target detection process (in the thermal image), which is the "real" measurement. Therefore, it must be tuned experimentally.

One can choose to ignore the issue with correlated measurement noise (assume that the measurement noise is white) and this works well in terms of the accuracy [18], but is problematic for covariance estimation. Therefore, an alternative is to account for colored noise in the measurement model and this approach will be outlined here. The positions of the target in north and east are still used as measurement. However, the measurement noise is modeled as a Gauss Markov process. The main argument for this approach is to maintain the simplicity that georeferencing and a linear system provide, but account for some of the issues that georeferencing causes. Consider the system model

$$\begin{aligned}
&\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{v}(k) \\
&\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{w}_c(k) \\
&\mathbf{w}_c(k) = \mathbf{F}_c\mathbf{w}_c(k) + \mathbf{w}_w(k) \\
&E[\mathbf{w}_w(k)\mathbf{w}_w(j)^\top] = \mathbf{R}_w\delta_{kj} \\
&E[\mathbf{v}(k)\mathbf{v}(j)^\top] = \mathbf{Q}\delta_{kj} \\
&E[\mathbf{v}(k)\mathbf{w}_w(j)^\top] = \mathbf{0}
\end{aligned} \tag{9}$$

where δ_{jk} is 1 when $k = j$ and zero otherwise. $\mathbf{x}$ is the state, $\mathbf{z}$ is the measurement and $\mathbf{v}$ is process noise. The state transition matrix $\mathbf{F}$ is chosen as in (3). The measurement matrix $\mathbf{H}$ is

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{10}$$

when the first two elements in $\mathbf{x}$ are the north and east positions and the remaining two are the corresponding velocities. Colored noise in the measurement is modeled with $\mathbf{w}_c$. The transition matrix for the measurement noise, $\mathbf{F}_c$, is chosen as the identity matrix. The argument for choosing the identity matrix is that the navigation error for the UAV, which causes the correlated noise, is expected to be equal for two consecutive images captured closely in time.

One possibility for handling colored noise is to augment the state vector with $\mathbf{w}_c$ and estimate the noise process. However, this is troublesome because the estimate of $\mathbf{w}_c$ can be used to cancel the noise in the measurement $\mathbf{z}$. Consequently, the covariance of the corresponding states are pushed to zero and is problematic for measurement association in multi-target tracking. Therefore, a better approach is to use the so-called differenced measurement [23] and rewrite (9). A new measurement $\mathbf{y}$ is defined as

$$\begin{aligned}
\mathbf{y}(k) &= \mathbf{z}(k+1) - \mathbf{F}_c\mathbf{z}(k) \\
&= \mathbf{H}\mathbf{F}\mathbf{x}(k) + \mathbf{H}\mathbf{v}(k) + \mathbf{w}_w(k) - \mathbf{F}_c\mathbf{H}\mathbf{x}(k) \\
&= \underbrace{(\mathbf{H}\mathbf{F} - \mathbf{F}_c\mathbf{H})}_{\mathbf{H}^*}\mathbf{x}(k) + \underbrace{(\mathbf{H}\mathbf{v}(k) + \mathbf{w}_w(k))}_{\mathbf{w}(k)} \\
&= \mathbf{H}^*\mathbf{x}(k) + \mathbf{w}(k)
\end{aligned} \tag{11}$$

where the new measurement noise $\mathbf{w}(k)$ is white, but correlated with the process noise since it includes $\mathbf{v}(k)$. The covariance of the new measurement noise is

$$E[\mathbf{w}(k)\mathbf{w}(j)^\top] = \mathbf{R} = (\mathbf{H}\mathbf{Q}\mathbf{H}^\top + \mathbf{R}_w)\delta_{kj} \tag{12}$$

where $\mathbf{R}_w$ is the covariance of the expected error in the original measurement $\mathbf{z}$. To remove the correlation between the new measurement and the process noise, a new motion model is defined:

$$\begin{aligned}
\mathbf{x}(k+1) &= \mathbf{F}\mathbf{x}(k) + \mathbf{v}(k) + \mathbf{T}[\mathbf{y}(k) - \mathbf{H}^*\mathbf{x}(k) - \mathbf{w}(k)] \\
&= \underbrace{(\mathbf{F} - \mathbf{T}\mathbf{H}^*)}_{F^*}\mathbf{x}(k) + \underbrace{(\mathbf{v}(k) - \mathbf{T}\mathbf{w}(k))}_{\mathbf{v}^*} + \mathbf{T}\mathbf{y}(k) \\
&= \mathbf{F}^*\mathbf{x}(k) + \mathbf{v}^*(k) + \mathbf{T}\mathbf{y}(k)
\end{aligned} \tag{13}$$

where $\mathbf{T}$ is a transformation matrix used to remove the cross-correlation and $\mathbf{v}*$ is the new process noise. It is chosen so that

$$E[\mathbf{v}^*(k)\mathbf{w}(k)] = 0 \tag{14}$$

and it can be shown that $\mathbf{T} = \mathbf{Q}\mathbf{H}^\top\mathbf{R}^{-1}$ fulfills this requirement [23]. The new process noise covariance is:

$$\mathbf{Q}^* = E[\mathbf{v}^*(k)\mathbf{v}^*(k)^\top] = \mathbf{Q} - \mathbf{Q}\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}\mathbf{Q} \tag{15}$$

The new system system (based on $\mathbf{y}$, $\mathbf{F}^*$, $\mathbf{H}^*$, $\mathbf{R}$, $\mathbf{T}$ and $\mathbf{Q}^*$) is implemented as a regular Kalman filter, but note that the measurement model $\mathbf{y}(k)$ includes both $\mathbf{z}(k+1)$ and $\mathbf{z}(k)$. Thus, when a new measurement $\mathbf{z}(k+1)$ is obtained, the previous measurement $\mathbf{z}(k)$ must be subtracted and the time between them must be used as the sampling time in the Kalman filter. Note also that the new measurement $\mathbf{y}(k)$ is a measure of the velocity since the difference between two consecutive position measurements are used.

This approach is valid if any new measurement is received closely in time to the previous measurement. If the new measurement is far from the previous (e.g. more than a second), the approach outlined here is not used. In such a situation, the new position measurement is used directly to update the state estimate with the original system (9), where $\mathbf{w}_c$ is assumed to be white with covariance $\mathbf{R}_w$.

V. Description of Experiments and UAV

This section describes the UAV platform used to collect data and two independent flight experiments that have been conducted to gather experimental data. Three different tracking filters are evaluated experimentally and they are also described briefly in this section.

A. UAV Platform

A fixed-wing UAV with electrical propulsion was used to collect data and is shown in Figure 1. A tailor-made payload with georeferencing capabilities was used and consists of the following sensors:

- Flir Tau2 thermal camera with a resolution of 640x512 pixels and focal length of $19\,\mathrm{mm}$. The frame rate is $7.5\,\mathrm{Hz}$.
- A ThermalGrabber used to extract the digital image from the Tau2 thermal camera.
- Pixhawk autopilot running Arduplane software to acquire the navigation states of the UAV.
- SyncBoard to synchronize the camera and the autopilot [27].
- Odroid-XU4 as on-board computer.

Fig. 1. The NTNU Cruiser-Mini fixed-wing UAV.

The UAV is capable of operating for two hours with a cruise speed of approximately $20\,\mathrm{m/s}$. It covers large geographical areas in a relatively short amount of time. The UAV can operate at altitudes of several hundred meters, which gives significant ground coverage for the thermal camera. The altitude of the UAV is a trade-off between ground coverage and spatial resolution as well as safety and air space restrictions. Since thermal cameras have lower resolution than visual spectrum cameras, it is important to assess the operating altitude of the UAV. A single pixel in the thermal camera covers a square with sides of $0.36\,\mathrm{m}$ at an altitude of $400\,\mathrm{m}$.

B. Experiments

Two independent flight experiments have been carried out to gather relevant data. The same marine vessel, which is shown in Figure 2, was used as target in both experiments. A thermal image of the vessel captured during the first experiment is displayed in Figure 3. The vessel is visible in the bottom right corner of the image and illustrates the typical signal-to-noise ratio in the images. The target was recorded in 264 images during the first experiment and the detections are divided into three separate segments with a significant amount of time in between the segments. The target was moving at high speed trying to maintain a constant course. The speed varied somewhat because of varying water and wave resistance. The gathered data cover a time period of approximately 3 minutes and are used to illustrate how well the tracking filters are able to track a high speed target with a small amount of measurements.

Fig. 2. The surface vessel used as target in the experiments.

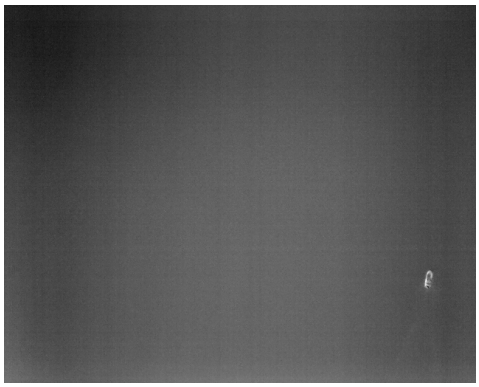

Fig. 3. A thermal image of the vessel, which is located in the bottom right corner, captured by the UAV in the first experiment.

The second flight experiment was conducted to collect thermal images of the same vessel. 2200 images of the target were recorded in many segments spread over approximately 25 minutes. The amount of gathered data is larger in this experiment and the vessel was only drifting in the waterline for most of the time leading to a small ground speed. This experiment is used to demonstrate how the system works for a longer time period with varying amount of measurements.

The UAV was steered manually during take-off and landing, but was controlled by the Pixhawk autopilot during the rest of the time in both experiments. The path was chosen manually based on a map where the position of the target was visible. Automatic path design using information from the tracking system could have been utilized [28], but choosing the flight pattern manually was preferred to ensure that the desired data were captured during different maneuvers to give more variety in the data.

C. Tracking Filters

Three different tracking filters are evaluated in this work. Measurements of the north and east target positions are acquired through georeferencing and all filters use the same measurements. The three tracking filters are:

TABLE I

FILTER PARAMETERS IN IN CASE STUDY 1

Parameter	Filter 1	Filter 2	Filter 3
Initial covariance	diag(10^2, 10^2 100^2, 100^2)	diag(10^2, 10^2, 100^2, 100^2, 0.01^2, 0.01^2)	diag(10^2, 10^2, 100^2, 100^2)
$\mathbf{R}$	diag (7^2, 7^2)	diag(7^2, 7^2)	diag(7^2,7^2)
$\mathbf{Q}$	σ_v = 0.05 (see (4))	a_M = 0.01, α = 1/10, p_M = 0.05, p_0 = 0.7 (see (7)-(8))	σ_v = 0.05 (see (4))

- Filter 1 - Kalman filter with the constant velocity motion model (3).
- Filter 2 - Kalman filter with the Singer motion model (5)-(7).
- Filter 3 - Kalman filter with the constant velocity motion model (3), and where colored noise are modeled in the measurement model as outlined in (11)-(15).

VI. RESULTS

The results are divided into two independent case studies, one for each flight experiment. The three tracking filters are evaluated on the same data and denoted Filter 1-3 as described in Section V-C. All results are based on real experimental data that have been post processed after the flight experiments.

A. Case study 1 - Tracking of high speed vessel

The goal of the first case study is to track a vessel moving at high speed and the data gathered at the first flight experiment are used. The filter parameters are summarized in Table I. The process noise is adjusted to have the same influence in all tracking filters, but filter 2 is different because the Singer motion model has a distinct set of adjustable parameters. The process noise covariance is chosen to be small because the target tries to maintain a constant course, which fits with the constant velocity motion model. An analysis of how the covariance of the process noise can be chosen for different vessels is presented in [29].

The initial velocity is chosen to be zero because no prior information about the velocity exist. The position is initialized with the first measurement. The initial covariance in position is chosen to be larger than the measurement noise to let the first few measurements adjust the initial position quickly. The covariance of the initial velocity is chosen much larger because the velocity is unknown initially. The standard deviation of the measurement noise is chosen to be 7 m in both north and east, which is a typical georeferencing error at the given altitude. The mean absolute georeferencing error in this case study is approximately 6 m.

The paths of the UAV and the vessel are displayed in Figure 4. The UAV operated at an altitude of 350 m. Figure 5 shows the target position measured with a single-frequency GPS receiver with update rate of 1 Hz and the distribution of georeferenced measurements. The measurements are clustered into three different groups because the target was outside the field of view of the camera for extended periods. The UAV flew over and past the target before it took a turn to do a new flyover. This was repeated and gave three segments with target observations as shown in Figures 4 and 5. Since the frame rate of the camera is 7.50 Hz, many measurements are available whenever the target is within the field of view of the camera.

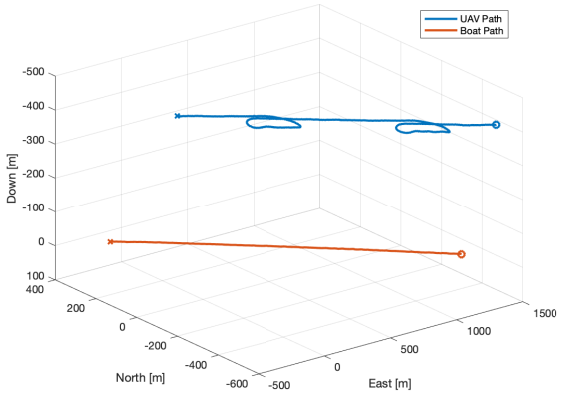

Fig. 4. The positions of the UAV and target in case study 1.

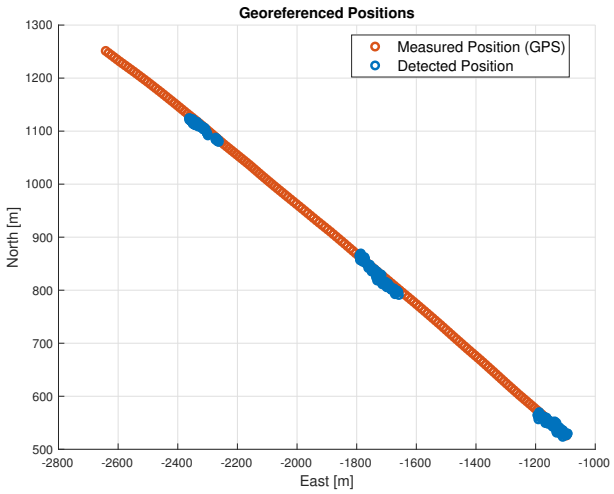

Fig. 5. Distribution of the georeferenced positions (blue) together with the position measured by a single-frequency GPS on-board the target in case study 1.

Figure 6 shows the root-mean-square estimation error in position for all tracking filters together with the measurement errors. The single-frequency GPS receiver on-board the target is used as a reference for the ground truth position. This is not necessarily a perfect reference so the results should be interpreted with that in mind (especially the estimation error and NEES). The measurement error is obviously equal for all tracking filters since the same measurements are used. Figure 6 also shows the periods where measurements are available (black crosses) and note that this only corresponds to a total time of approximately 35 s. Therefore, the target estimates are based on prediction for most of the time and

explains why the estimation error is larger initially. Filter 1 and 2 have the smallest estimation error overall, but filter 3 is closing in near the end.

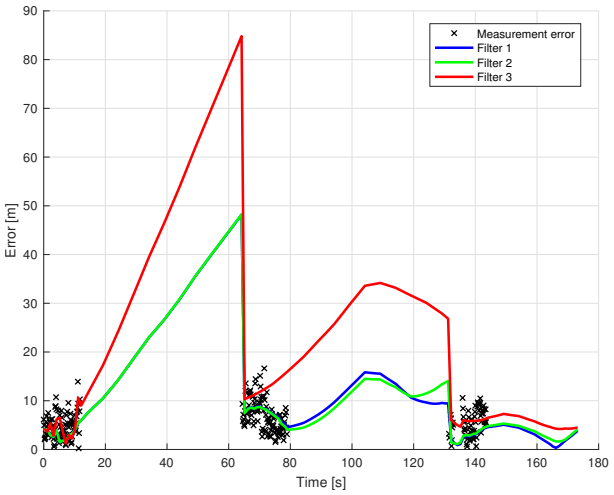

Fig. 6. The estimation error in position (absolute distance) and measurement errors in case study 1.

Figure 7 shows the estimated speed and heading together with the reference. All filters are able to estimate the speed and heading accurately after the initial period. The main difference is that filter 3 converges more slowly, which may be a drawback with this filter since it reduces the filter gain. This is also the reason for why the predicted position of the third filter drifts more than the other two initially.

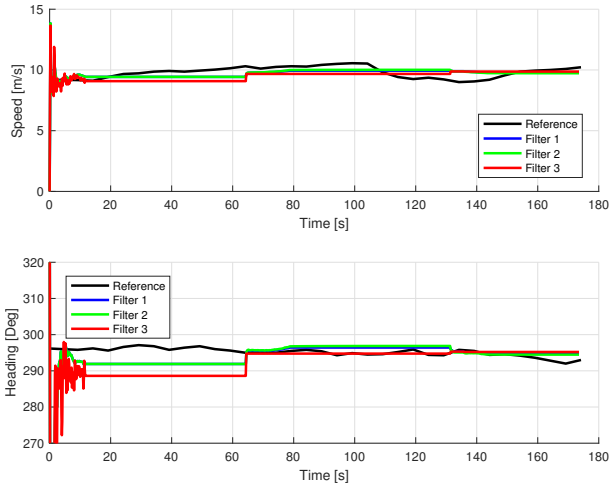

Fig. 7. Estimated target speed and heading in case study 1.

Figure 8 shows NEES (only in position since an accurate true reference for the north and east velocities is unavailable) and NIS for all filters (see Section II). The confidence bounds are also shown and the values should be within the bounds for a filter to be consistent. The NEES exceeds the upper bound for Filter 1 and Filter 2, and this means that the estimated covariance is smaller than the corresponding estimation error. Thus, the filter is too optimistic and assesses the estimates to be more trustworthy than they really are. The

TABLE II

AVERAGE PERFORMANCE METRICS IN CASE STUDY 1

Parameter	Filter 1	Filter 2	Filter 3
Average Measurement error [north, east]	$[-3.77\,\text{m}, -0.94\,\text{m}]$	$[-3.77\,\text{m}, -0.94\,\text{m}]$	$[-3.77\,\text{m}, -0.94\,\text{m}]$
Average absolute measurement error	$5.87\,\text{m}$	$5.87\,\text{m}$	$5.87\,\text{m}$
Average Estimation error [north, east]	$[-9.2\,\text{m}, 1.1\,\text{m}]$	$[-8.6\,\text{m}, 0.9\,\text{m}]$	$[-21.6\,\text{m}, 5.0\,\text{m}]$
Average absolute Estimation error	$12.81\,\text{m}$	$12.84\,\text{m}$	$25.12\,\text{m}$
Average NEES	15.65	13.08	1.88
Average NIS	0.46	0.44	0.21

third filter has estimates of the covariance which represent the true error more correctly since NEES is within the bounds. The NIS is similar for the different tracking filters and within the confidence bounds. The NIS values are closer to the lower limit and this suggests that the covariance of the measurement noise could have been reduced since NIS increases with a reduction in the measurement noise covariance. However, this would increase NEES even more and lead to covariance estimates that are further from the truth. Since NIS is similar and a significant difference is experienced in NEES, a bias is most likely present in the estimates since Filter 1 and 2 are too optimistic. This is supported by the values in Table II which summarizes the average results for the entire tracking period in this case study. The mean estimation error in north is nonzero and is most likely because of a bias in the measurements caused by colored noise in the navigation errors. Filter 3 is better equipped to handle such a situation as the NEES shows even though the root-mean-square estimation error is comparable.

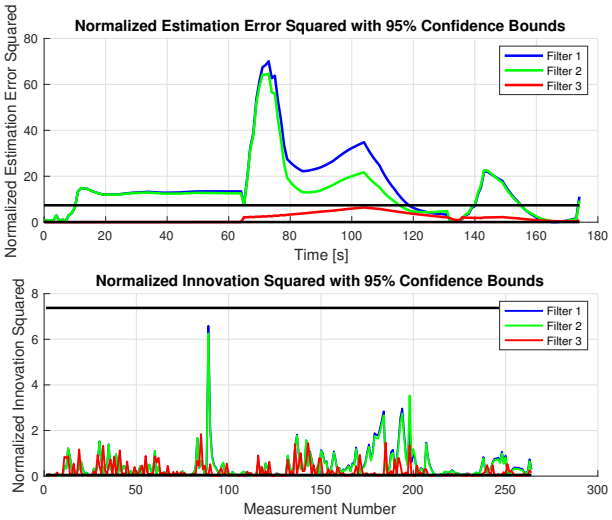

Fig. 8. The normalized estimation error squared and the normalized innovation squared for the different tracking filters in case study 1.

Figure 9 shows the correlation in the innovation sequences and the confidence bounds. The innovation sequence is correlated for all tracking filters, but in a much lower degree

TABLE III

FILTER PARAMETERS IN CASE STUDY 2

Parameter	Filter 1	Filter 2	Filter 3
Initial covariance	diag(13^2, 13^2, 1, 1)	diag(13^2, 13^2, 1, 1, 0.01, 0.01)	diag(13^2, 13^2, 1, 1)
$\mathbf{R}$	diag (13^2, 13^2)	diag(13^2, 13^2)	diag(13^2, 13^2)
$\mathbf{Q}$	σ_v = 0.01	$a_M = 0.002$, $\alpha = 1/10$, $p_M = 0.05$, $p_0 = 0.7$	$\sigma_v = 0.01$

for the third filter. This is expected since the main motivation with the third filter is to whiten the measurement noise and improve filter consistency. Therefore, considering NIS, NEES and the correlation in the innovation sequences, only filter 3 is close to fulfilling the criteria for consistency and the only filter that is reliable if multiple targets are considered.

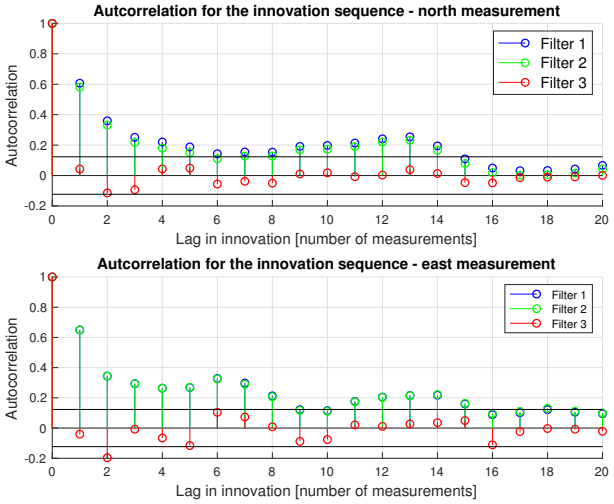

Fig. 9. The autocorrelation for the innovation sequences in case study 1.

B. Case study 2 - Tracking of drifting vessel at low speed

The goal of the second case study is to track a vessel drifting in the water at low speed and is based on the data gathered at the second flight experiment. The filter parameters are summarized in Table III and will not be discussed as thoroughly as in case study 1. The main difference is the reduction in the covariance of the process noise, which is explained by the reduced velocity of the target. This also explains the lower initial covariance in the velocity. The measurement noise is increased because the measurement errors are larger in this case study. This is most likely because the UAV followed a path similar to the figure eight (see Figure 10), which gives more variation in the attitude. The UAV followed straight-line segments in the first experiment, which normally reduce the georeferencing error.

The paths of the UAV and the vessel are displayed in Figure 10. The UAV operated at a varying altitude of 200 m to 300 m. The target was drifting in the water and moved only a short distance even though the experiment lasted for 25 min.

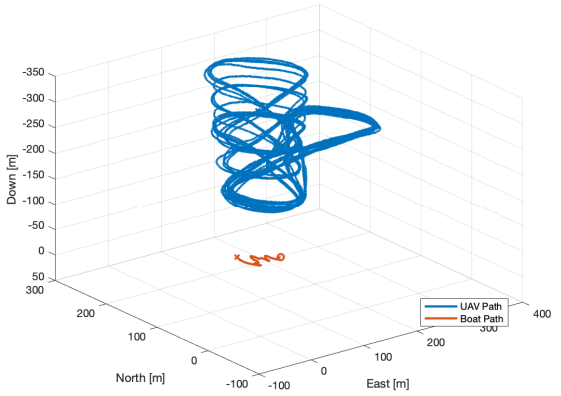

Fig. 10. The positions of the UAV and target in case study 2.

Figure 11 shows the path of the target measured with a single-frequency GPS receiver at 0.5 Hz and the position estimates from the tracking filters. The root-mean-square estimation error in position and the measurement errors are displayed in Figure 12. The accuracy is more similar in this case study. Filter 3 still has the largest estimation error. A total of 2200 measurements are available, but they are spread into segments. 2200 images correspond to a continuous observation of the target in 300 s. Thus, measurements are only available for 300 s in the tracking period which lasts for 1500 s. Therefore, the estimates are mostly based on prediction using the motion model, but works well since the speed of the target is low.

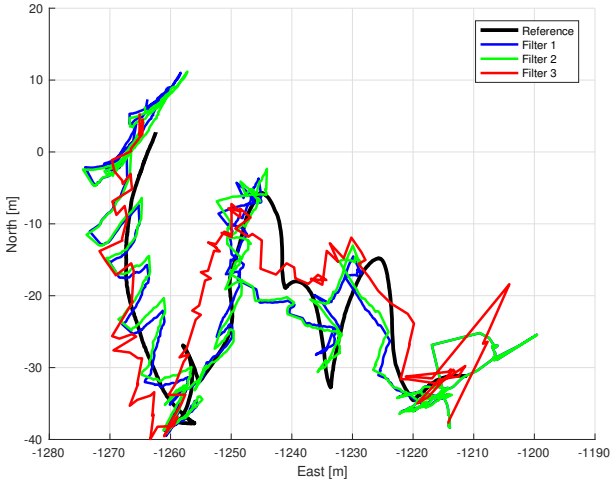

Fig. 11. Estimated position for all tracking filters in case study 2.

Figure 13 shows NEES and NIS for all filters. The NEES is larger than the upper confidence limit for filter 1 and filter 2, and implies that the estimated covariance is lower than the corresponding estimation error. Thus, filter 1 and 2 are too optimistic and this behavior was also experienced in case study 1. The NIS is similar for filter 1 and 2, but smaller for the third filter. This is supported by the values in Table IV which summarizes the average results in case study 2.

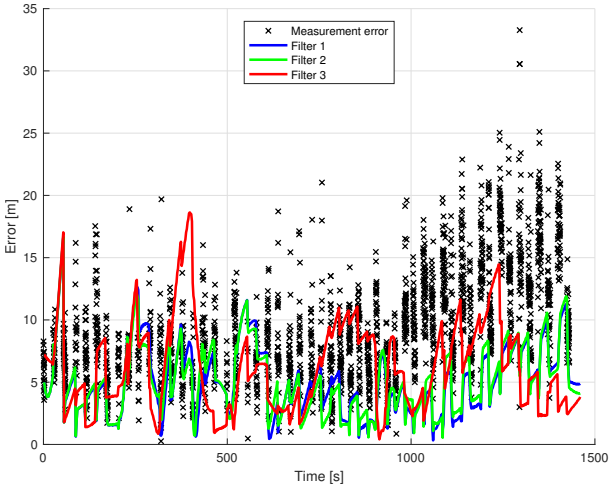

Fig. 12. The estimation and measurement errors in case study 2.

The estimation error is larger for filter 3, but the estimate of the covariance is better since NEES is within the confidence bounds for almost the entire tracking period.

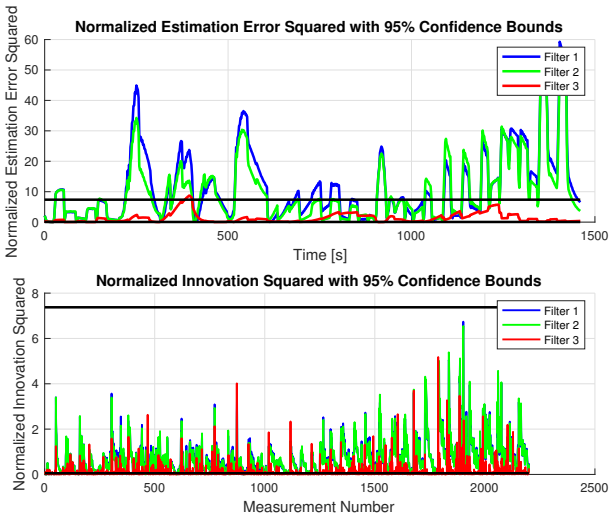

Fig. 13. The normalized estimation error squared and the normalized innovation squared in case study 2.

Figure 14 shows the correlation in the innovation sequences. The innovation sequence is correlated for all tracking filters, but in less degree for the third filter. The influence of colored measurement noise is obvious for filter 1 and 2, which is expected as they neglect it in the design. Considering correlation, NIS and NEES, only filter 3 is consistent.

Both cases studies have shown that colored noise is propagated into the tracking system through georeferencing and that filter 3 is better equipped to mitigate colored noise. The same difference in performance has not been illustrated between filter 1 and 2. Therefore, a clear benefit with the Singer motion model is not demonstrated. Nevertheless, it is expected because the target is not maneuvering significantly. In a situation with more maneuvers, the Singer model is expected to surpass the constant velocity motion model.

TABLE IV
AVERAGE PERFORMANCE METRICS IN CASE STUDY 2

Parameter	Filter 1	Filter 2	Filter 3
Average Measurement error [north, east]	[1.59 m, −0.32 m]	[1.59 m, −0.32 m]	[1.59 m, −0.32 m]
Average absolute measurement error	9.68 m	9.68 m	9.68 m
Average Estimation error [north, east]	[0.8 m, −0.4 m]	[0.9 m, −0.4 m]	[0.2 m, −0.5 m]
Average absolute Estimation error	5.08 m	5.12 m	6.22 m
Average NEES	11.53	10.12	1.34
Average NIS	0.69	0.67	0.12

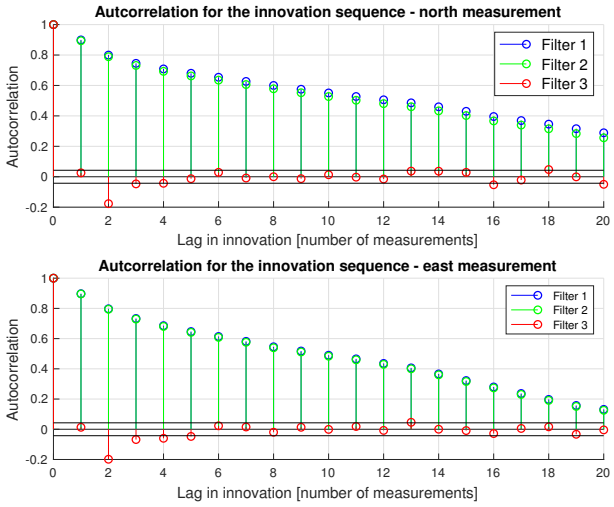

Fig. 14. The autocorrelation for the innovation sequences in case study 2.

VII. Conclusion

This work has investigated how tracking filters can be designed for floating objects when a thermal camera mounted in a fixed-wing UAV is used as the imaging sensor. Two independent flight experiments were carried out and it was demonstrated that a single target can be tracked with high accuracy. The results illustrate that it is important to compensate for colored noise in the measurement model to maintain consistency, but that colored noise does not reduce the estimation error. Thus, the application requirements decide if it is necessary to account for colored noise. If multi-target tracking is considered, accurate covariance estimation is vital and colored noise should be modeled in the Kalman filter. However, in terms of the root-mean-square estimation error, it is not of the same importance. Therefore, if the goal is to estimate the position of a single object through georeferencing without the need for accurate covariance information, a simpler approach is applicable.

Acknowledgment

This work has been carried out at the NTNU Centre for Autonomous Marine Operations and Systems (NTNU AMOS). This work was supported by the Research Council of Norway through the Centres of Excellence funding

scheme, Project number 223254. It has also been suppoorted by NFR/MAROFF with project number 269480 and NFR/FRINATEK projects 221666 and 250725. The authors are grateful for the flawless support from the excellent UAV operators Lars Semb and Pål Kvaløy, and with the assistance from Maritime Robotics during the experiments.

REFERENCES

[1] G. Yang, B. Li, S. Ji, F. Gao, and Q. Xu, "Ship detection from optical satellite images based on sea surface analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 3, pp. 641–645, March 2014.

[2] T. A. Johansen and T. Perez, "Unmanned aerial surveillance system for hazard collision avoidance in autonomous shipping," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2016, pp. 1056–1065.

[3] V. V. Klemas, "Coastal and environmental remote sensing from unmanned aerial vehicles: An overview," *Journal of Coastal Research*, pp. 1260–1267, 2015.

[4] L. Fusini, J. Hosen, H. H. Helgesen, T. A. Johansen, and T. I. Fossen, "Experimental validation of a uniformly semi-globally exponentially stable non-linear observer for gnss- and camera-aided inertial navigation for fixed-wing uavs," in *Proc. of the International Conference on Unmanned Aircraft Systems*, 2015, pp. 851–860.

[5] J. Hosen, H. H. Helgesen, L. Fusini, T. I. Fossen, and T. A. Johansen, "Vision-aided nonlinear observer for fixed-wing unmanned aerial vehicle navigation," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 8, pp. 1777–1789, 2016.

[6] H. F. Grip, T. I. Fossen, T. A. Johansen, and A. Saberi, "Nonlinear observer for gnss-aided inertial navigation with quaternion-based attitude estimation," in *2013 American Control Conference*, June 2013, pp. 272–279.

[7] G. Conte and P. Doherty, "An integrated uav navigation system based on aerial image matching," in *2008 IEEE Aerospace Conference*, March 2008, pp. 1–10.

[8] G. Chowdhary, E. N. Johnson, D. Magree, A. Wu, and A. Shein, "Gps-denied indoor and outdoor monocular vision aided navigation and control of unmanned aircraft," *Journal of Field Robotics*, vol. 30, no. 3, pp. 415–438, 2013.

[9] N. Gageik, P. Benz, and S. Montenegro, "Obstacle detection and collision avoidance for a uav with complementary low-cost sensors," *IEEE Access*, vol. 3, pp. 599–609, 2015.

[10] A. Cesetti, E. Frontoni, A. Mancini, P. Zingaretti, and S. Longhi, "A vision-based guidance system for uav navigation and safe landing using natural landmarks," *Journal of Intelligent and Robotic Systems*, vol. 57, no. 1, p. 233, Oct 2009.

[11] S. Xu, A. Savvaris, S. He, H. Shin, and A. Tsourdos, "Real-time implementation of yolo+jpda for small scale uav multiple object tracking," in *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2018, pp. 1336–1341.

[12] H. H. Helgesen, F. S. Leira, T. I. Fossen, and T. A. Johansen, "Tracking of ocean surface objects from unmanned aerial vehicles with a pan/tilt unit using a thermal camera," *Journal of Intelligent & Robotic Systems*, vol. 91, no. 3, pp. 775–793, Sep 2018.

[13] J. M. Vazquez-Nicolas, E. Zamora, I. Gonzalez-Hernandez, R. Lozano, and H. Sossa, "Towards automatic inspection: crack recognition based on quadrotor uav-taken images," in *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2018, pp. 654–659.

[14] S. Bhandari, A. Raheja, M. R. Chaichi, R. L. Green, D. Do, F. H. Pham, M. Ansari, J. G. Wolf, T. M. Sherman, and A. Espinas, "Lessons learned from uav-based remote sensing for precision agriculture*," in *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2018, pp. 458–467.

[15] H. H. Helgesen, F. S. Leira, T. A. Johansen, and T. I. Fossen, *Detection and Tracking of Floating Objects Using a UAV with Thermal Camera*. Cham: Springer International Publishing, 2017, pp. 289–316.

[16] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: A survey," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84–100, Jan 2018.

[17] C. D. Rodin, L. N. de Lima, F. A. de Alcantara Andrade, D. B. Haddad, T. A. Johansen, and R. Storvold, "Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems," in *2018 International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–8.

[18] F. S. Leira, T. A. Johansen, and T. I. Fossen, "Automatic detection, classification and tracking of objects in the ocean surface from uavs using a thermal camera," in *2015 IEEE Aerospace Conference*, March 2015, pp. 1–10.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 91–99.

[20] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11 – 28, 2016.

[21] E. M. Hemerly, "Automatic georeferencing of images acquired by uav's," *International Journal of Automation and Computing*, vol. 11, no. 4, pp. 347–352, Aug 2014.

[22] E. D. Kaplan and C. Hegarty, *Understanding GPS/GNSS: Principles and applications*. Artech House, 2017.

[23] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[24] S. Y. Chen, "Kalman filter for robot vision: A survey," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4409–4420, Nov 2012.

[25] X. Rong Li and V. P. Jilkov, "Survey of maneuvering target tracking. part i. dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, Oct 2003.

[26] ——, "Survey of maneuvering target tracking: Iii. measurement models," in *Proc.SPIE*, vol. 4473, 2001, pp. 4473 – 4473 – 24.

[27] S. M. Albrektsen and T. A. Johansen, "Syncboard - a high accuracy sensor timing board for uav payloads," in *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2017, pp. 1706–1715.

[28] E. Skjong, S. A. Nundal, F. S. Leira, and T. A. Johansen, "Autonomous search and tracking of objects using model predictive control of unmanned aerial vehicle and gimbal: Hardware-in-the-loop simulation of payload and avionics," in *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2015, pp. 904–913.

[29] E. F. Wilthil, A. L. Flåten, and E. F. Brekke, *A Target Tracking System for ASV Collision Avoidance Based on the PDAF*. Cham: Springer International Publishing, 2017, pp. 269–288.