# Real-time Georeferencing of Thermal Images using Small Fixed-Wing UAVs in Maritime Environments

Håkon Hagen Helgesen[a,*], Frederik Stendahl Leira[a], Torleiv H. Bryne[a], Sigurd M. Albrektsen[a], Tor Arne Johansen[a]

[a]*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*

**Abstract**

This article considers real-time georeferencing using a fixed-wing unmanned aerial vehicle (UAV) with a thermal camera. A flexible system for direct georeferencing is proposed without the need for ground reference points. Moreover, as the system is tailored for highly maneuverable and agile fixed-wing UAVs, no restrictions on the motion are assumed. The system is designed with a solution for accurate time synchronization between sensors. This feature enables tracking of objects with low uncertainty. Sensors for navigation, permitting estimation of the UAV pose with a nonlinear observer, are employed in addition to a thermal camera. The estimated UAV pose is utilized in georeferencing to acquire Earth-fixed coordinates of objects. The main examples studied in this research are georeferencing of a static object and of a moving marine vessel. To obtain the desired accuracy, thermal camera calibration and compensation of mounting misalignment errors are discussed. The entire system is validated in two independent field experiments with a thorough analysis of the results. Georeferencing of a static object is conducted with centimeter accuracy when the average position of all measurements is used. The position of a moving marine vessel is obtained with mean accuracy of two meters.

*Keywords:* Unmanned Aerial Vehicle, Real-time Georeferencing, Navigation, Target Tracking, Thermal Imaging

## 1. INTRODUCTION

Unmanned aerial vehicles (UAVs) are used in various remote sensing applications (Pajares, 2015; Colomina and Molina, 2014; Toth and Jóźków, 2016), such as inspection (Deng et al., 2014; Nikolic et al., 2013), disaster management (Yahyanejad and Rinner, 2015) and tracking of vessels and icebergs (Helgesen et al., 2017a,b; Sakamaki et al., 2017; Leira et al., 2017), because of their flexibility. Fixed-wing UAVs have the additional benefit of being able to carry numerous sensors, while covering larger areas. Therefore, UAVs are ideal for mapping and monitoring of marine areas, such as coastlines and the sea surface.

Optical sensors are suitable for mapping and surveillance of unknown territories to obtain situational awareness. Images can be used to detect undiscovered objects or gather information about the proximity of the sensor. A vital part of mapping and surveillance is the ability to relate findings in the optical sensor frame to Earth-fixed coordinates. This is called georeferencing and is a fundamental part of any mapping and surveillance system using a monocular optical sensor (Leira et al., 2015b; Hemerly, 2014). This research concerns real-time georeferencing using a fixed-wing UAV with a thermal camera. Conceptually, there is no difference in georeferencing of thermal and visual spectrum imagery. Thus, the algorithms presented in this work are relevant also for visual spectrum images.

Georeferencing is particularly useful when the optical sensor is mounted on a moving platform because pixel coordinates are less informative when the camera moves. In addition, the motion of the camera (UAV) affects the target pixel position more than the target motion itself in many applications. Georeferencing using a monocular camera is challenging since range information are missing. Consequently, it is hard to accurately reconstruct three unknown Earth-fixed coordinates from two pixel coordinates. Moreover, it is desirable to conduct georeferencing in real-time on a small embedded computer without needing to process data on a ground station.

The foremost challenge in real-time georeferencing is to obtain accurate coordinates from a single image without using ground reference points. Georeferencing rely on the camera pose (position and attitude) at the exact time when the image is captured. Small deviations from the true pose give large errors in the georeferenced position. Hence, conducting real-time georeferencing of thermal images on-board a UAV with significant roll and pitch motions is more difficult than georeferencing from a stationary or slowly-moving platform. Accurate time synchronization among the sensors is critical. In addition, thermal images have lower resolution than visual spectrum images and is consequently more fragile for pixel and sub-pixel errors that occur in object detection.

The most accurate existing solutions for real-time geo-

*Corresponding author

*Email address:* hakon.helgesen@ntnu.no (Håkon Hagen Helgesen )

referencing are based on utilizing reference points on the ground at known locations (Zhou et al., 2005; Xiang and Tian, 2011; Harwin and Lucieer, 2012; Hartmann et al., 2012), but is not feasible in unknown territories. The solution presented in (Choi and Lee, 2013) requires common features in consecutive images, which is hard to obtain in thermal imagery of marine areas.

A few systems for direct real-time georeferencing without the use of reference points exist. However, these solutions lack the desired accuracy, are not suitable for fixed-wing UAVs or are not validated experimentally with a significant amount of data. A georeferencing system for fixed-wing UAVs with a thermal camera is presented in (Leira et al., 2015b), but the variance in the results is large when considering the altitude of the UAV, which is below 100 meters. Capable accuracy is shown in georeferencing of the length and width of a runway (Hemerly, 2014) at an altitude exceeding 300 meters, but the results are based on a single image so the amount of data is limited. Moreover, reconstructing Earth-fixed coordinates is a different task because calibration errors in the principal point do not affect the size of an object. Challenges related to synchronization between the sensors and mounting misalignments are not addressed thoroughly in the aforementioned works. Time synchronization and camera calibration are addressed in (Rehak et al., 2013; Daakir et al., 2017), but these solutions are tailored towards multi-copters so it is arguably hard to verify how they apply to the fast dynamics and operating envelope of fixed-wing UAVs.

This research presents a novel complete solution for real-time georeferencing of thermal images captured with a fixed-wing UAV, and the system is not depending on ground reference points. The georeferencing system is fully integrated on the UAV and can be executed in real-time. The most common challenges and pitfalls are addressed, and state of the art methodology is utilized. The research focuses on a particular use case, namely georeferencing of floating objects where it is utilized that the sea forms a planar surface. Thermal cameras are ideal for detection of marine vessels and floating objects on the sea surface, but can be replaced by a visual spectrum camera if other objects are of interest. Georeferencing is used to track objects in (Helgesen et al., 2017a), so this work is relevant also for target tracking.

Since georeferencing requires precise knowledge about the camera pose, a state of the art nonlinear observer is designed and used to estimate the pose of the UAV. The navigation states are estimated without a magnetic compass, which is beneficial because magnetic sensors are sensitive to electromagnetic disturbances on-board the UAV. A significant part of this research concerns methods for handling the most dominant error sources in georeferencing. This includes calibration of thermal cameras and how to detect mounting misalignment errors. The final contribution is a thorough experimental validation consisting of more than 4000 images in two independent flight experiments, and a comprehensive analysis of the results.

*Outline of Article*

Real-time georeferencing is described in Section 2 and the navigation system used to estimate the UAV pose is described in Section 3. Section 4 describes thermal camera calibration as well as data synchronization and calibration of mounting misalignment errors. Section 5 presents the experimental validation of the system before the paper is concluded in Section 6. The final part are two appendices that describe the navigation system in detail and how errors in the UAV pose affect georeferencing, respectively.

## 2. Real-time Georeferencing

Georeferencing is used as a term for finding the north-east-down (NED) coordinates of a single pixel in the rest of this work, and covered in e.g. (Leira et al., 2015b; Hemerly, 2014). The method presented here is conceptually similar, but reviewed here for understanding challenges in real-time georeferencing that are addressed later.

Figure 1 illustrates the most important coordinate frames. A pixel in the image plane is denoted $(u, v)$ where $u$ is the horizontal coordinate and $v$ is the vertical coordinate. The camera-fixed coordinate frame moves with the UAV and is denoted {c}. The x axis of {c} points in the direction of $u$, the y axis in the direction of $v$ and the z axis straight out the camera lens. The body-fixed frame of the UAV, denoted {b}, moves with the UAV and is defined in (Beard and McLain, 2012). The x axis of {b} points out the nose of the airframe, the y axis points sideways out the right wing and the z axis points out the belly. The NED frame is denoted {n} and is defined locally as a tangent plane at the location of the experiment. The x axis points north, the y axis east and the z axis down.

Pixel coordinates have two degrees of freedom and NED coordinates have three. Hence, finding NED coordinates of a single pixel is an ill-posed problem. A common way to avoid this issue (in direct georeferencing) is to assume that all pixels within an image are located in the same plane. This is known as the flat-earth assumption and necessary for obtaining NED coordinates of a single pixel unless an elevation map exists. It is reasonable for an image showing the sea surface, but not in an area with mountains.

The pixel coordinates $(u, v)$ are related to {c} through the pinhole camera model (Hutchinson et al., 1996):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{f}{z^c} \begin{bmatrix} x^c \\ y^c \\ \frac{z^c}{f} \end{bmatrix}, \quad z^c \neq 0 \tag{1}$$

where $(x^c, y^c, z^c)$ are the camera-fixed coordinates and $f$ is the focal length of the camera lens. $z^c$ is the depth and unknown for a monocular camera. Equation (1) can also be formulated as

$$z^c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x^c \\ y^c \\ z^c \end{bmatrix}}_{\mathbf{p}^c} = \mathbf{A}\mathbf{p}^c \tag{2}$$

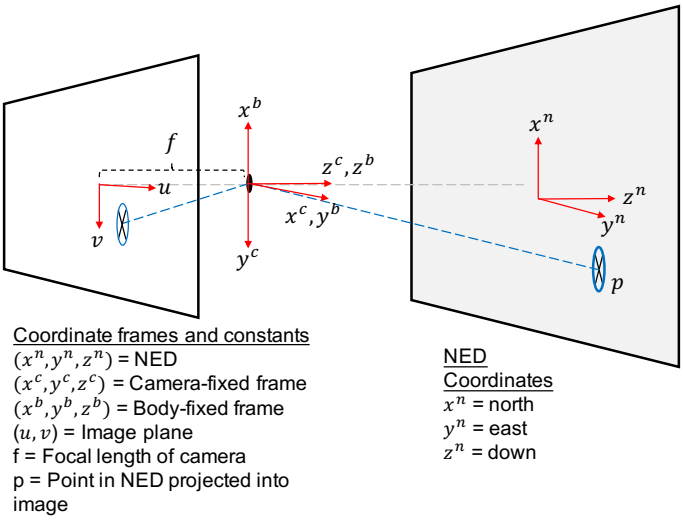


Figure 1: Illustration of coordinate frames.

where $\mathbf{A}$ is the camera intrinsic matrix. $f_x$ and $f_y$ are the focal lengths (expressed in pixels) in the horizontal and vertical direction, respectively. $c_x$ and $c_y$ are the principle points (center of lens) and should theoretically be in the image center. Note that the origin of the pixel coordinates $(u, v)$ is assumed to be in the top-left corner of the image in (2). Decomposing $\mathbf{p}^c$ in {n} is achieved by utilizing a transformation $\mathbf{G}_n^c$ between {c} and {n} (Leira et al., 2015b)

$$z^c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A}\mathbf{G}_n^c \underline{\mathbf{p}}^n \tag{3}$$

where $\underline{\mathbf{p}}^n$ is the homogeneous coordinate vector of the pixel decomposed in {n}. $\mathbf{G}_n^c$ is the homogeneous transformation

$$\mathbf{G}_n^c := \begin{bmatrix} \mathbf{R}_n^c, & -\mathbf{R}_n^c \mathbf{r}_{nc}^n \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1, & \mathbf{r}_2, & \mathbf{r}_3, & -\mathbf{R}_n^c \mathbf{r}_{nc}^n \end{bmatrix}$$

where $\mathbf{R}_n^c$ is the rotation matrix that transforms a vector in {n} to {c}. The same notation is used for all rotation matrices in this paper ($\mathbf{R}_b^a$ transforms a vector decomposed in a frame {b} to a frame {a}). The column vectors in $\mathbf{R}_n^c$ are denoted $\mathbf{r_1}$, $\mathbf{r_2}$ and $\mathbf{r_3}$, and $\mathbf{r}_{nc}^n$ is the position of the origin of {c} relative to {n} decomposed in {n}.

Without loss of generality, $\mathbf{R}_n^c$ is split into three consecutive rotations and expressed as

$$\mathbf{R}_n^c = (\mathbf{R}_b^n \mathbf{R}_c^b)^{-1} = (\mathbf{R}_b^n (\mathbf{R}_m^c \mathbf{R}_b^m)^{-1})^{-1} \tag{4}$$

where $\mathbf{R}_b^n$ is the rotation matrix between {n} and {b}, defined according to the zyx convention and specified in terms of the Euler angles (roll ($\phi$), pitch ($\theta$), yaw ($\psi$)) (Fossen, 2011):

$$\mathbf{R}_b^n = \begin{bmatrix} c\psi c\theta & -s\psi c\phi + c\psi s\theta s\phi & s\psi s\phi + c\psi c\phi s\theta \\ s\psi c\theta & c\psi c\phi + s\phi s\theta s\psi & -c\psi s\phi + s\theta s\psi c\phi \\ -s\theta & c\theta s\phi & c\theta c\phi \end{bmatrix} \tag{5}$$

where $c\psi = \cos(\psi)$ and $s\psi = \sin(\psi)$. The reference frame {m} is referred to as the gimbal frame, and is used if the camera is mounted in a gimbal and not strapped on the airframe. $\mathbf{R}_b^m$ relates the body frame to the gimbal frame and is parametrized by the gimbal angles. If the camera is strapped directly to the airframe, $\mathbf{R}_b^m$ is simply chosen as the identity matrix. The second rotation, $\mathbf{R}_m^c$, relates the camera-fixed frame to the gimbal frame and used to align the image plane with the gimbal frame in a specific way.

$\mathbf{R}_b^m$ depends on the gimbal. In this research, it is assumed that the gimbal has two degrees of freedom, namely pan and tilt angles. It is further assumed that the gimbal is mounted so that {b} is aligned with {m} when the gimbal has zero pan ($\psi_{gb}$) and tilt ($\theta_{gb}$). In {b}, pan and tilt movements correspond to a rotation along the body z and y axes, respectively. Hence, the rotation is defined as

$$\begin{aligned} \mathbf{R}_b^m &= (\mathbf{R}_z(\psi_{gb})\mathbf{R}_y(\theta_{gb}))^\top = \mathbf{R}_y^\top(\theta_{gb})\mathbf{R}_z^\top(\psi_{gb}) \\ &= \begin{bmatrix} \cos\psi_{gb}\cos\theta_{gb} & \sin\psi_{gb}\cos\theta_{gb} & -\sin\theta_{gb} \\ -\sin\psi_{gb} & \cos\psi_{gb} & 0 \\ \cos\psi_{gb}\sin\theta_{gb} & \sin\psi_{gb}\sin\theta_{gb} & \cos\theta_{gb} \end{bmatrix} \end{aligned} \tag{6}$$

where $\mathbf{R}_z(\cdot)$ and $\mathbf{R}_y(\cdot)$ are principle rotations about the z and y axes, respectively (Fossen, 2011). The pinhole camera model (1) is defined so that the x axis of {c} is aligned with the horizontal direction in the image plane ($u$) and not the body-fixed x axis. Therefore, $\mathbf{R}_m^c$ is a rotation of -90 degrees about the camera z axis:

$$\mathbf{R}_m^c = \mathbf{R}_z(-90^o) = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

In general, $\mathbf{R}_m^c$ depends on how the camera and gimbal are mounted with respect to the body of the UAV, and how the pinhole camera model is defined.

Only two coordinates in NED can be recovered by a single pixel $(u, v)$ as explained in the beginning of this section. However, since objects at the sea surface are of interest, the down position of pixels is close to zero as long as the origin of {n} is placed at the mean sea level. Consequently, one can identify the north-east (NE) coordinates using a single pixel and set the down position to zero. The NE coordinates ($\mathbf{p}_t^n$) of the pixel $(u, v)$ are given by (3) as

$$\begin{bmatrix} p_{t,N}^n \\ p_{t,E}^n \\ 1 \end{bmatrix} = z^c \mathbf{G}_{NE}^{-1} \mathbf{A}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{8}$$

where $\mathbf{G}_{NE}$ is defined as

$$\mathbf{G}_{NE} := \begin{bmatrix} \mathbf{r}_1, & \mathbf{r}_2, & -\mathbf{R}_n^c \mathbf{r}_{nc}^n \end{bmatrix} \tag{9}$$

The depth $z^c$ is left out of the calculation by normalizing the coordinates on the left side so that the down-component gets a value of 1. This concludes the direct georeferencing algorithm, which is summarized by the following steps that are applied whenever a new image is received:

1. Detect object(s) and find the pixel position $(u, v)$, e.g. through image processing.
2. Get UAV pose and gimbal angles from the navigation system.
3. Calculate $\mathbf{R}_n^c$ with equation (4)-(7) and find $\mathbf{G}_{NE}$ with (9).
4. Use (8) to calculate NE coordinates.

The georeferencing equations can be executed in real-time on a small embedded computer, but the overall computation time is obviously limited by the complexity of the object detection algorithm in step 1.

## 3. Navigation System

Real-time georeferencing rely on accurate information about the UAV pose as outlined in Section 2. Therefore, it is necessary to design a real-time navigation system that accurately estimates the navigation states. The attitude dynamics are in general nonlinear and all degrees of freedom need to be considered. Navigation filters can be probabilistic, such as the extended Kalman filter (EKF) or the Multiplicative EKF (MEKF) (Sola, 2017), or deterministic such as nonlinear observers (Hosen et al., 2016; Fusini et al., 2018; Mahony et al., 2008; Euston et al., 2008; Grip et al., 2013). The main advantages with nonlinear observers are reduced computational complexity, easy implementation, and the possibility of proving global stability for the error dynamics through Lyapunov analysis since linearization is avoided in contrast to EKF-based techniques. The main advantage with EKF-based designs is that the state uncertainties are available directly through the covariance matrix.

In this research, a state of the art nonlinear observer is used because of the aforementioned properties of nonlinear observers. The observer design is described in Appendix A. The structure of the observer is originally presented in (Grip et al., 2013). The translational motion observer is based on measurements from an accelerometer and RTK-GNSS receivers. The attitude observer utilizes angular rate measurements and reference vectors that are extracted from an accelerometer, and pitch and yaw measurements from two RTK-GNSS receivers with separate antennas (Sollie et al., 2019).

The nonlinear observer represents position in the Earth-Centered-Earth-Fixed (ECEF) coordinate frame since GPS measurements are given in ECEF. The georeferencing equations are represented in {n} so a transformation between ECEF and {n} is needed. ECEF coordinates are transformed into longitude ($l$), latitude($\mu$) and height ($h$) (Vermeille, 2004) by using the WGS-84 reference ellipsoid. Local NED coordinates are acquired from $l$, $\mu$ and $h$ (Fossen, 2011; Farrell, 2008). The origin of {n} is placed at a user-defined $(l_0, \mu_0, h_0)$. Note that a local flat earth assumption is present when using NED coordinates and is only valid in proximity of $(l_0, \mu_0, h_0)$ because a tangent plane is used when defining {n}.

## 4. Sensor Calibration and Synchronization

The core of the georeferencing system is the thermal camera used to detect objects. A thermal camera is preferred because of the distinct thermal signature of marine vessels and other floating objects compared to the sea surface. More specifically, the sea surface emits less infrared radiation than a typical floating object. Consequently, it is (in most cases) easier to detect floating objects in thermal images than in visual spectrum images. Algorithms for detection of floating objects is not within the scope of this work, but described in several articles (Helgesen et al., 2017b; Deng et al., 2018; Rodin and Johansen, 2018).

### *4.1. Calibration of a Thermal Camera*

Georeferencing depends on the camera intrinsic matrix, which can be extracted from the camera and lens specification. However, in practice, the true camera model often deviates from the theoretical model and calibration is necessary. Calibration of visual spectrum cameras is addressed extensively in the literature and most methods are based on (Zhang, 2000).

Calibration is of similar importance for thermal cameras. However, because the image sensor is sensitive to a different spectrum than the visual, the standard chessboard calibration surface is not clearly visible in thermal images. Therefore, another calibration surface is needed. Thermal camera calibration is addressed before (Leira et al., 2015b; Yahyanejad et al., 2011; Hartmann et al., 2012), but a common calibration surface has not been accepted for thermal cameras. This research uses a new surface to conduct the calibration.

Most calibration patterns for thermal cameras use some sort of heat source to generate a visible structure. The pattern proposed here is based on heating a circle grid, which is detected during calibration. The circle grid is a 3D-printed square plastic plate. The plate is mounted on a wooden surface, which is covered with a conductive (resistive) coating that is heated when electric current flows through. By connecting a power source to the coating, the wooden surface is warmed up quickly. When the circle grid is attached to the wood, the heat escapes through the circles and they seem warmer than the rest of the plastic plate. The main advantage with this surface is that the circle pattern is visible rapidly. The heating time is just about a minute and the power source keeps the wood at a high temperature throughout the duration of the calibration. In addition, the temperature difference between the circles and the rest of the plastic plate is large because the plate conducts (or transfers) heat poorly, giving a high signal to noise ratio. Therefore, the calibration surface proposed here is advantageous compared to surfaces that are heated by lamps or cooled down passively because the temperature difference remains large for a longer period. An example of an image captured during calibration is displayed in Figure 2.

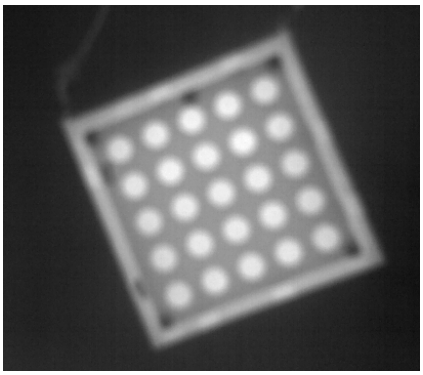

Figure 2: A thermal image used for camera calibration.

### *4.2. Synchronization of data*

The UAV pose must be known accurately when an image is captured. Therefore, robust time stamping of the capture time and a navigation system with sufficiently high update rate are needed. Furthermore, accurate synchronization between the time frames of the camera and the navigation system is required. Errors in the magnitude of fractions of a second are troublesome, and especially vital during fast maneuvers where the attitude is changing quickly. The solution developed in this work uses SyncBoard (Albrektsen and Johansen, 2017) (renamed to SenTiBoard more recently). The precision of the time stamping using SyncBoard is tested to be 1 µs for navigation sensors, such as IMUs and GNSS receivers.

SyncBoard has two main tasks in this work. It is used to synchronize the sensors in the navigation system internally, namely an IMU and two RTK-GPS receivers. In addition, it is used to synchronize the camera frame captures with the navigation system. An alternative is to use the on-board computer time for when an image is stored, but this solution is not accurate since there may be a significant bias between the on-board computer time and GPS time. Moreover, the on-board computer clock may drift during the duration of the experiment, and can be busy with other tasks when images are received.

The thermal camera used in this research is a FLIR Tau2. The internal frame rate of the camera is 30 Hz, but the output rate of images is reduced to 7.5 Hz because of export regulations. External triggering of image capture is not possible. Therefore, to obtain the desired accuracy, SyncBoard synchronizes the internal camera clock with GPS every second and thereby the camera and the navigation system. A potential drift in the camera clock is removed with this approach. Note that this solution neglects the delay between exposure of the pixels and when the image is ready on the camera output bus, but this must be accepted when external triggering is infeasible. Nevertheless, the delay is only minor because the internal time between consecutive images is 33.3 ms so any potential camera latency is significantly smaller than that.

### *4.3. Calibration of misalignment errors*

Camera calibration and time synchronization are key in georeferencing. Another error source is mounting misalignment errors, which can be even more critical. Remember from Section 2 that it is assumed that the gimbal frame {m} coincides with {b} when the gimbal has zero pan and tilt angles. In practice, it is hard to align these frames perfectly, which results in significant georeferencing errors as shown in Appendix B. This section derives a method for estimating the mounting misalignment. More specifically, the goal is to identify the rotation matrix between the body-fixed frame and the camera when the pan and tilt angles are zero. The procedure utilizes the georeferencing algorithm from Section 2 and is conducted after the gimbal and camera are mounted in the airframe.

The main requirement is to use a motion-capture system that provides information about the attitude and position of an object, relative to a fixed frame with high precision. For example OptiTrack (OptiTrack Motion Capture Systems, 2018) is a motion-capture system based on several cameras, and is able to give the attitude and position for a set of markers that form a rigid-body. Four markers are needed to define the body-fixed frame of the UAV (one on each wing, one in the front and one in the back). Moreover, a set of four markers is also mounted on a second object (the thermal camera calibration surface).

The intention behind the procedure is to use the thermal camera to capture images of the second object. When the position and attitude of the camera are known (given by motion-capture system), it is possible to calculate where the object theoretically should be located by the georeferencing equations. This is compared with where the object is located according to the motion-capture system. Moreover, by using several images, an optimization problem is solved so that the rotation matrix between {b} and {c} is calculated based on where the object is and should be located. Note that the body frame the gimbal is aligned with through this method is given directly by the markers. The precision of the calibration is, therefore, limited by how accurate the markers are aligned with the true geometric body of the UAV.

The gimbal has two degrees of freedom (pan and tilt). In addition, it is beneficial to add a roll rotation because there can be mounting misalignment errors in all degrees of freedom. The theoretical NED positions for the second object (based on the pixel coordinates that are selected manually from the images) are calculated as

$$\frac{1}{z^c}\begin{bmatrix} p^n_{t,N} \\ p^n_{t,E} \\ 1 \end{bmatrix} = \mathbf{G}^{-1}_{NE}\mathbf{A}^{-1}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{10}$$

The theoretical values $[p^n_{t,N}, p^n_{t,E}]^\top$ are compared with the values measured by the motion-capture system, and is formulated as an optimization problem with several known points in multiple images where the position and attitude

of the camera and/or the objects vary. The optimization problem is stated as

$$J = \underset{\psi_{gb},\theta_{gb},\phi_{gb}}{\text{minimize}} \sum_{i=1}^{n_{\text{images}}} \sum_{k=1}^{n_{\text{points}}} f(i,k) \tag{11}$$

where $\psi_{gb}, \theta_{gb}$ and $\phi_{gb}$ are misalignments in pan, tilt and roll, respectively. The objective function is defined as

$$f(i,k) = ||\mathbf{p}_t^n(i,k) - \mathbf{p}_m^n(i,k)|| \tag{12}$$

and the subscript $m$ and $t$ are the measured (based on the motion-capture system) and theoretical (based on the position in the image) coordinates, respectively. Moreover, the subscripts $i$ and $k$ correspond to point $k$ in image $i$. Three points in every image are needed because the optimization problem is formulated with respect to three unknown parameters. Optimization is conducted for several images to reduce the uncertainty a single image provides and assumed to mitigate the influence of noise. The problem is solved with the nonlinear least squares algorithm. Since misalignment errors are generally small, initialization is straightforward.

## 5. Experimental Validation and Results

This section presents the experimental results of this research and is divided in five parts. The first part describes the sensor suite used to collect data. The results of the optical camera calibration are presented in the second part, and the third part describes calibration of misalignment errors. The fourth and fifth parts present georeferencing results from two independent flight experiments.

### *5.1. Sensor Suite*

Figure 3 shows the fixed-wing UAV used in the experiments. A payload tailored for small fixed-wing UAVs driven by an electrical engine was developed. All of the sensors in the navigation system, except the GPS antennas, were placed in one small stack together with SyncBoard and the on-board computer. The camera was mounted in a retractable pan/tilt gimbal. The following sensors and systems were a part of the payload:

- FLIR Tau2 thermal camera with resolution of 640 × 512 pixels. The focal length is 19 mm, giving a field of view of $32^o \times 26^o$.
- A ThermalGrabber (ThermalCapture Grabber USB 640 - TEAX Technology, 2018) for extracting the digital image.
- Analog Devices ADIS 16490 IMU measuring linear acceleration and angular rate at a frequency of 250 Hz.
- 2x RTK-GPS based on uBlox NEO-M8T receivers measuring position velocity, pitch and heading at a rate of 5 Hz.

Figure 3: The NTNU Cruiser-Mini fixed-wing UAV with the thermal camera visible near the front.

- SyncBoard (Albrektsen and Johansen, 2017).
- Odroid-XU4 on-board computer.

The measurements of the pitch and heading were acquired by measuring the baseline between the antennas of the two receivers on the UAV and the base station, and comparing this vector to the baseline in the body frame. The on-board computer stored images from the camera, and measurements from the GPS receivers and IMU with time stamping from SyncBoard. Figure 4 shows the structure of the system, and a thermal image captured during the flight experiments is shown in Figure 5.

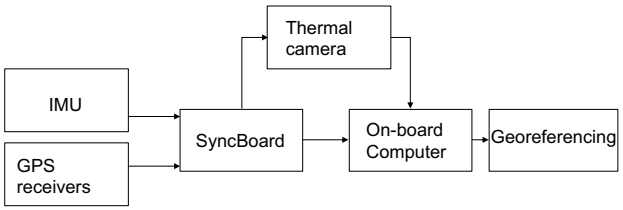


Figure 4: Structure of Georeferencing system.

### *5.2. Camera Calibration Results*

Camera calibration was conducted with a set of 30 images and the following intrinsic matrix was obtained:

$$\mathbf{A}_{\text{Tau2}} = \begin{bmatrix} 1159.2 & 0 & 313 \\ 0 & 1167.8 & 265 \\ 0 & 0 & 1 \end{bmatrix}$$

The values are close to the theoretical matrix for the camera, which is (based on the camera and lens specification)

$$\mathbf{A}_{\text{Theoretical}} = \begin{bmatrix} 1117.7 & 0 & 320 \\ 0 & 1117.7 & 256 \\ 0 & 0 & 1 \end{bmatrix}$$

The principal point of the camera was estimated to be 7 and 9 pixels from the theoretical center in the horizontal

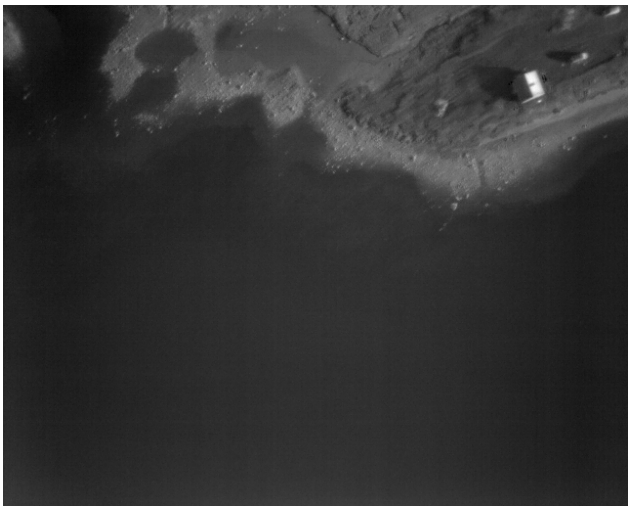

Figure 5: A thermal image captured during the experiments. The dark area is the sea surface.

and vertical direction, respectively. Moreover, the focal lengths are close to the theoretical values. Distortion coefficients were also estimated, but are not included here because they affected the results only marginally.

### *5.3. Calibration of Misalignment Errors*

The mounting misalignment of the gimbal was calculated to be (based on the method in Section 4.3)

$$\phi_{gb} = -1.7^o, \quad \theta_{gb} = 3.9^o, \quad \psi_{gb} = 1.9^o$$

Challenges related to sensor synchronization were avoided in the calibration process since the camera was stationary during image capture.

The calibration process was verified by projecting the calibration points from the image plane back into NED and comparing that with the true position. The mean reprojection error was 0.0229 m and the calibration was conducted at a range of approximately 1.2 m. The remaining error is a combination of errors in the marked pixel position, the camera intrinsic matrix and the camera pose (the motion-capture system is not perfect).

If it is assumed that the remaining reprojection error is caused by errors in the orientation of the camera, the error grows to 6.6 m when the range is increased to 350 m, which is the operating altitude in the first experiment. The expected error is 3.8 m when the altitude is 200 m as in the second experiment. This is only a rough indication for the error one can expect in the experimental data as the error is caused by other sources as well, which are not proportional to the altitude. Moreover, the range is greater than the altitude in turns.

Although the gimbal misalignment has been identified, a misalignment between the body frame as defined by the IMU, and the body frame defined by the geometry of the UAV (which is the body frame used in the gimbal calibration) can be present. IMU misalignment was detected by comparing the Euler angles computed by the navigation system with zero when the UAV was leveled by a leveler. IMU calibration was also conducted, giving an offset in roll, pitch and yaw of $1.1^o$, $-5.7^o$ and $0.25^o$, respectively. These angles are obviously equally important for the accuracy of georeferencing. Note that gyro bias is estimated in the navigation system, but not a potential accelerometer bias. However, as the IMU calibration was conducted when the accelerometer was warm, accelerometer bias variations is likely not affecting the results significantly. Moreover, the run bias stability of the accelerometer is small according to the manufacturer.

### *5.4. Flight Experiment 1 - Georeferencing of base station GPS Antenna*

The first flight experiment was carried out near Agdenes, in the outer Trondheim fjord in Norway. The goal was to collect data for georeferencing of a stationary base station GPS-antenna located two meters above sea level. The GPS receiver connected to the antenna had RTK capability, which means that the true position of the antenna was known with centimeter accuracy. The antenna was chosen as target to have a static object with known ground truth position. Nearly 8000 images were captured, and the flight experiment lasted for approximately 25 min. The GPS antenna was in the field of view and marked manually in 552 images. An uncertainty of a few pixels was potentially added during this process, but the mean error should be close to zero. The data have been processed after the flight, but georeferencing has also been conducted in real-time on the on-board computer more recently.

The marked pixel positions were used together with the UAV pose from the nonlinear observer to acquire NED positions of the antenna. The UAV was equipped with a Pixhawk autopilot running Arduplane software and using its own set of sensors (Pixhawk 4 - By Dronecode, 2018), and navigation data from the autopilot were also used in georeferencing for comparison. Note that the autopilot was calibrated for misalignment errors in the same way as the nonlinear observer so the difference experienced later is mainly caused by the difference in sampling rate (250 Hz for the observer and 10 Hz for the autopilot) and the quality of the sensors. Moreover, the autopilot was synchronized with the camera through SyncBoard. The timing error for the images without using SyncBoard (offset between GPS-time and the on-board computer clock) was 0.44 s during this experiment.

The path of the UAV is displayed in Figure 6, and the local NED frame is placed so that the target position is in the origin. The antenna was only in the field of view in a fixed part of the loiter motion near the origin so it was observed from approximately the same attitude and altitude for each set of measurements. The UAV operated at an altitude of 350 m above sea level. Operating at greater altitudes increases the ground coverage of a single pixel and reduces the precision of georeferencing. A single pixel covers a square with sides equal to approximately

$(8.95 \times 10^{-4} \times$ altitude) m for the lens used in the experiments. Moreover, because the georeferencing error, as a consequence of errors in the camera orientation, is proportional to the altitude of the UAV, larger errors must be expected when increasing the altitude in general.

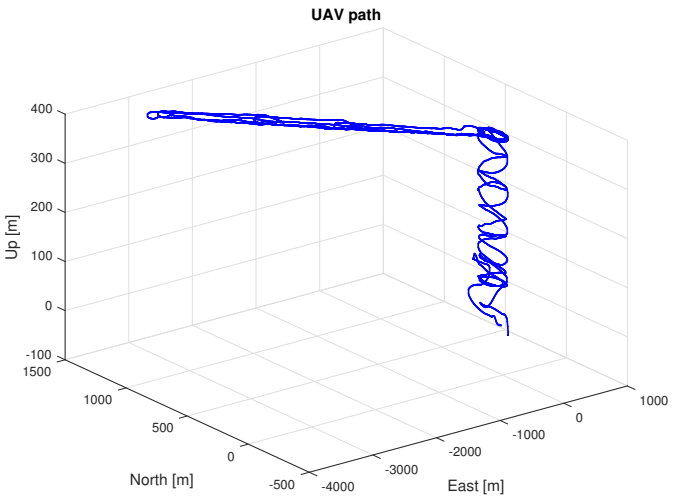


Figure 6: The UAV path estimated by the observer.

Figure 7 shows the UAV attitude estimated by the nonlinear observer compared with the estimates from the autopilot. Only the time span where detections occurred is shown. There are only minor differences between the autopilot and the observer, which indicate that both provide accurate estimates. Figure 8 shows the NED velocities, which are also similar. A true reference is not available for the velocity and attitude. Nevertheless, since the observer and autopilot used different GPS-receivers and IMUs, the estimates from both the autopilot and the observer are more credible when they are comparable.

Figure 9 shows the georeferenced positions of the antenna using the observer, while Figure 10 shows the georeferenced positions using the autopilot. The measurements are centered around the true position, and distributed in a way which is reminiscent of a Gaussian distribution without a bias. This is useful in target tracking where several measurements are used to estimate the target states.

Table 1 summarizes the main results using the observer and autopilot navigation data. The performance is comparable, but the standard deviation of the measurements is 4.10 m and 6.47 m for the observer, and 4.83 m and 7.84 m for the autopilot. The mean georeferenced position for all images using the observer is $-0.18$ m and $-0.12$ m in north and east, respectively. Since the true position is in the origin, the exact location (with centimeter accuracy) of the GPS antenna is known when all measurements are used. The mean north and east positions using the autopilot are 0.29 m and 0.09 m, respectively. The small difference in mean position indicates that both the observer and autopilot are competitive. Moreover, it shows that the quality of the sensors is less important than proper misalignment

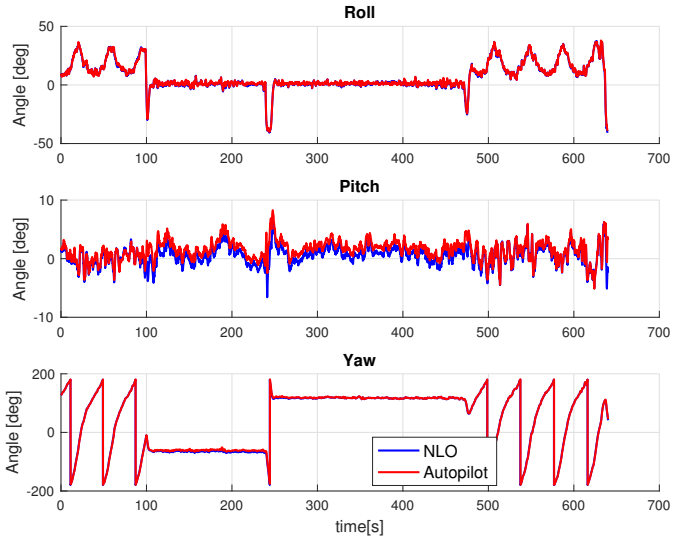


Figure 7: The UAV attitude estimated by the nonlinear observer (blue) and autopilot (red).

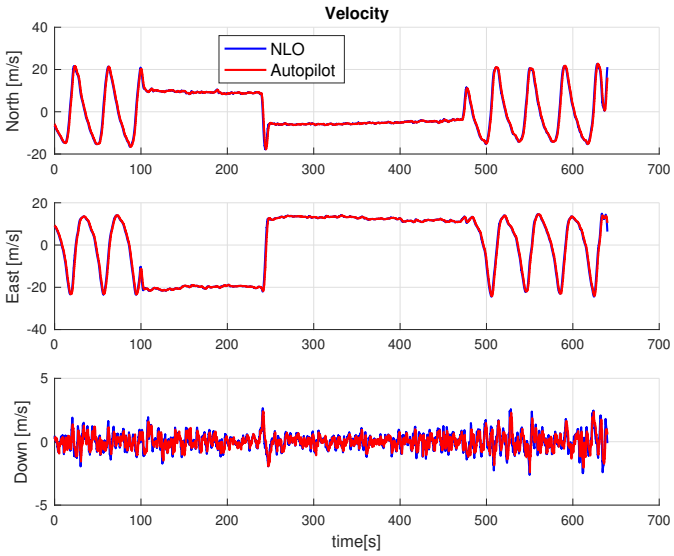


Figure 8: The NED velocities estimated by the nonlinear observer (blue) and the autopilot (red).

calibration and time synchronization.

The mean error for the georeferenced position using a single image is 6.40 m with the observer (7.70 m for the autopilot) at an operating altitude of 350 m. This is a significant result as the accuracy is comparable to previous work achieved at lower altitudes (Leira et al., 2015b). In fact, the error is also slightly lower than what was expected from the remaining reprojection error after the calibration process (6.6 m). Nevertheless, the assumption behind this value was that the error was solely caused by inaccuracies in the camera orientation, which is a simplification. A part of the error is related to factors such as errors in the UAV pose and in the detection process. Moreover,

Table 1: Main results of experiment 1. SD = standard deviation

| | **Observer** | **Pixhawk Autopilot** | **Reference** |
|---|---|---|---|
| Mean north position | −0.18 m | 0.29 m | 0 m |
| Mean east position | −0.12 m | 0.09 m | 0 m |
| SD north position | 4.10 m | 4.83 m | - |
| SD east position | 6.47 m | 7.84 m | - |
| Mean absolute error (single image) | 6.40 m | 7.70 m | - |

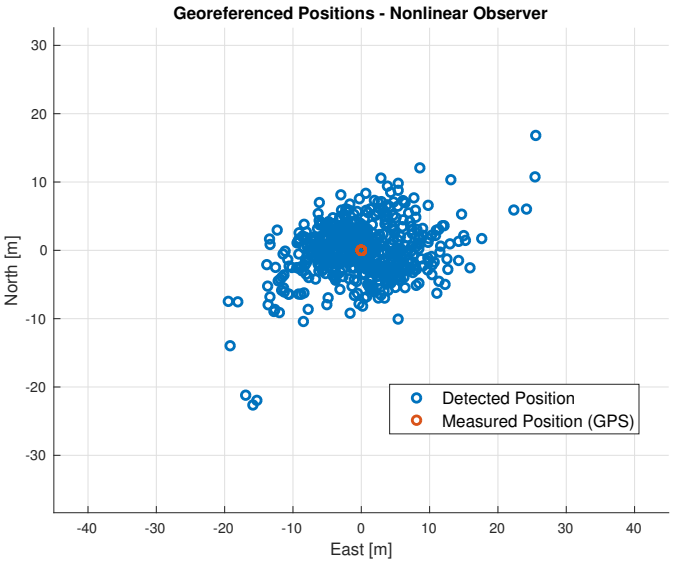


Figure 9: The georeferenced antenna position together with the true position with UAV pose from the nonlinear observer.

the calibration process identified the misalignment error for the gimbal, but the attitude of the UAV was assumed known through the motion-capture system. However, in a field experiment, the UAV pose is also uncertain and that can both increase and decrease the error.

Figure 11 and 12 show the corresponding georeferencing results without time synchronization of the camera (the image timestamps are assigned directly by the on-board computer) and without IMU misalignment calibration, respectively. Moreover, these results are summarized in Table 2. The mean error in a single image is 11.10 m without time synchronization compared to 6.40 m with time synchronization. The results without IMU misalignment calibration are dreadful in comparison. The mean error in a single image is 40.89 m which is explained by the mean position being −38.16 m and 9.28 m from the true position in north and east, respectively. Moreover, the distribution does not resemble a normally distributed variable. The ba-

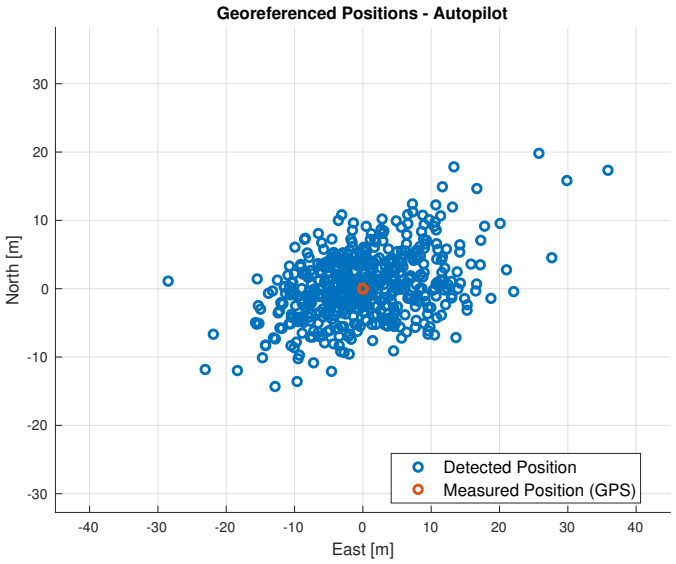


Figure 10: The georeferenced antenna position together with the true position with UAV pose from the Pixhawk/ArduPlane autopilot.

nana shape of the georeferenced points indicates that an error in the attitude is present since this is the expected shape during a turn if there is a bias in roll or pitch. This is clearly a significant reduction in performance even though the misalignment of the IMU was only 1.1° in roll, −5.7° in pitch and 0.25° in yaw. These results support the need for a proper and reliable misalignment calibration and time synchronization.

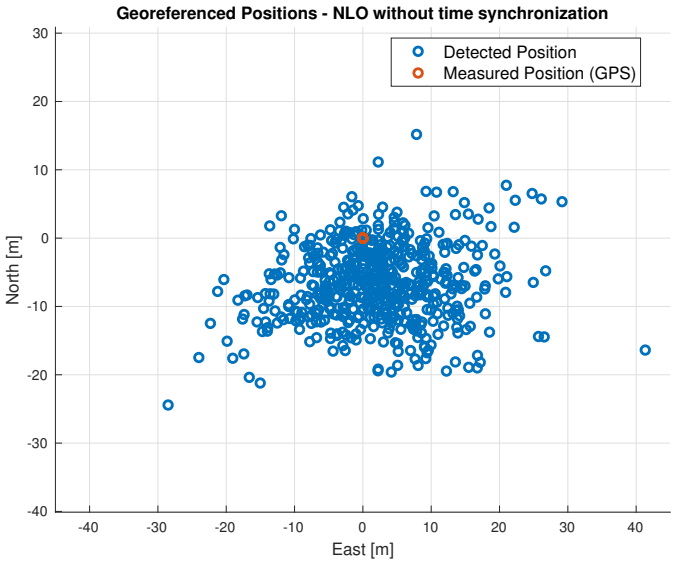


Figure 11: The georeferenced antenna position using the nonlinear observer without time synchronization.

Figure 13 shows the measurement and estimation error when the georeferenced measurements (using the observer) are filtered in a Kalman filter with a motion model cor-

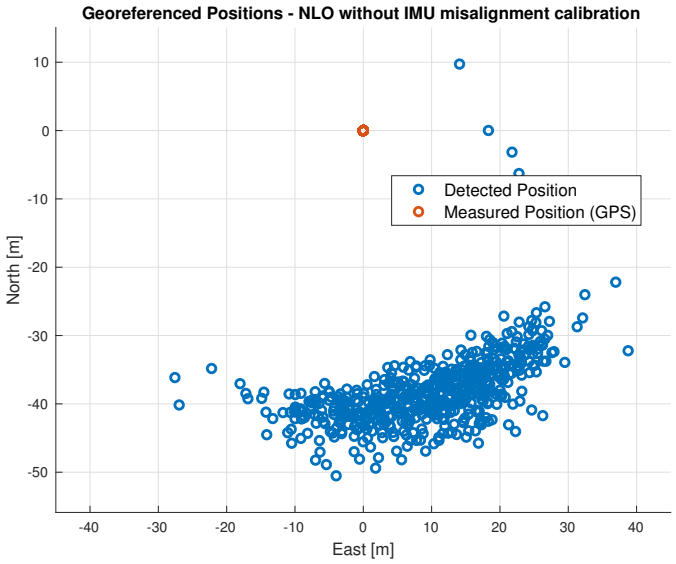


Figure 12: The georeferenced antenna position using the nonlinear observer without IMU misalignment calibration.

Table 2: Main results of experiment 1 with and without time synchronization and IMU misalignment calibration. SD = standard deviation

| | **Observer** | **Observer without time sync.** | **Observer without misalignment calibration** |
|---|---|---|---|
| Mean north position | −0.18 m | −6.67 m | −38.16 m |
| Mean east position | −0.12 m | 2.04 m | 9.28 m |
| SD north position | 4.10 m | 5.64 m | 5.40 m |
| SD east position | 6.47 m | 8.96 m | 10.63 m |
| Mean absolute error (single image) | 6.40 m | 11.10 m | 40.89 m |

responding to an object at rest (zero velocity). The estimation error is less than a meter after the first batch of measurements. Moreover, the estimation error in the end is less than 0.5 m. In practice, this confirms that it is possible to track slowly moving objects accurately with georeferencing. The largest measurement error occurs in the end where a single measurement has an error just surpassing 30 m. The overall best measurement has an absolute error below 0.3 m.

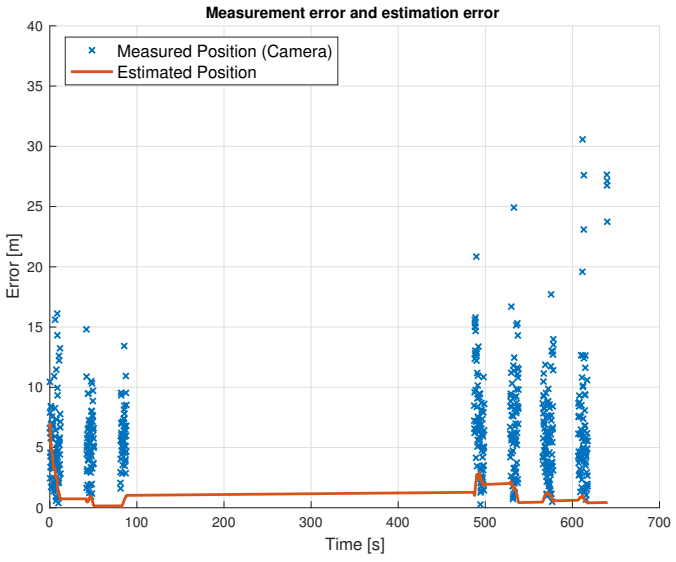


Figure 13: Measurement and estimation error when a Kalman filter is used to filter the georeferenced measurements of the antenna.

Figure 14 shows the estimated autocorrelation for the innovation process in the north and east positions. The innovation process is defined as the difference between the received measurement and the predicted measurement in the Kalman filter. The autocorrelation shows that the innovation process is correlated in both north and east, which means that the innovation in the previous measurement is correlated with the innovation in the current measurement. From a practical point of view, this is expected because consecutive images are captured from almost the same pose. Therefore, a systematic error in the camera orientation in one image is most likely present at the time when the next image is captured, and thus the georeferencing error is correlated for consecutive images.

In field experiments, it is not realistic to expect that the innovations are white. Moreover, since navigation errors are the most significant challenge, the errors in georeferencing depend on the output of a navigation filter. The estimation error in the navigation filter is often correlated for two consecutive estimates and propagates colored noise into the georeferencing algorithm. Nevertheless, the results show that it is beneficial to observe the target for a longer period so that errors in the navigation states are averaged as the target is observed from new poses and over a longer time period. Furthermore, the innovations illustrate that a Markov (or other colored noise model) augmentation could be beneficial when georeferenced measurements captured from a moving platform are used in a tracking system. Additionally, the spread of the points in Figure 9 indicates that the measurements can be approximated as a normally distributed variable for a large batch of measurements, even though consecutive innovations are correlated. It should also be emphasized that from an application point of view, the resulting georeferencing error in terms of root-mean-square error is often more important than fulfilling the theoretical requirements for optimality. This is especially relevant when tracking few targets as consistency is more important in multiple target tracking

or with a sensor where clutter is expected.

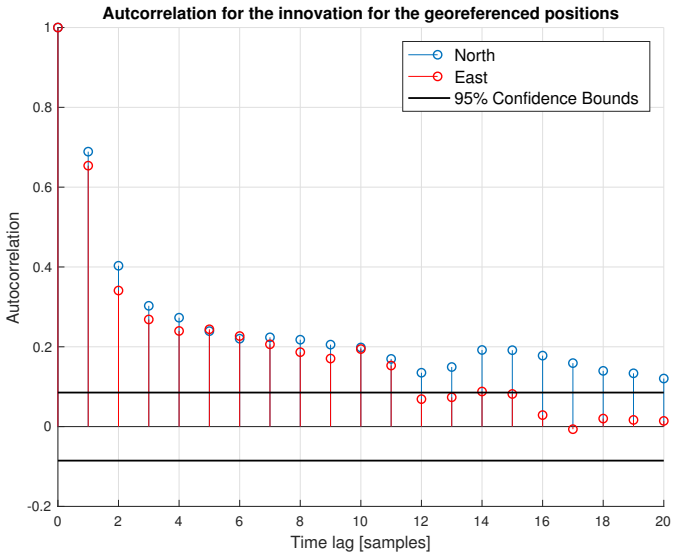


Figure 14: Autocorrelation for the innovations in the north and east georeferenced positions.

There is a difference in the autocorrelation for the north and east innovation processes. The acceleration is larger in east than north during antenna detections. Moreover, the antenna is observed from the same part of the loiter motion for every set of measurements, which gives the same variations in velocity and attitude. The attitude is not varying significantly during a coordinated-turn so the acceleration in east is the parameter that varies the most. This most likely explains why the correlation is smaller in the east innovation process since larger variations usually lead to better observability conditions in the corresponding degree of freedom. If the antenna was observed during other maneuvers or from another location on the loiter motion, a different behavior is expected.

### *5.5. Flight Experiment 2 - Georeferencing of small marine vessel*

A second independent experiment was carried out to study the accuracy of the georeferencing system in a relevant remote sensing application. The motivation behind this experiment was to measure the position of the marine vessel displayed in Figure 15. The motion of the vessel was mostly caused by drift in the sea water. However, a few short maneuvers were conducted during the experiment. The position of the vessel was measured with a single-frequency GPS receiver with a rate of 0.5 Hz for reference (ground truth). Consequently, the reference is more uncertain than for the base station because it does not have RTK capability or a high update rate. Nevertheless, the GPS position is used as a measure for the true position, but the error in the reference is potentially a few meters.

The position of the vessel in the image was detected automatically with the method presented in (Leira et al., 2015a; Helgesen et al., 2017b), which returns the center of the detected object. Several detections have been investigated manually to ensure that issues like false and inaccurate detections are limited. Nevertheless, the exact pixel position is not detected perfectly in every image, but the detection error is normally minor and most often at a level corresponding to a single pixel. The experiment lasted for almost 50 min and more than 22000 images were captured. The vessel was detected in 3635 images and georeferencing was conducted for each detection. The exact same system parameters (with respect to IMU and gimbal misalignments) as in the first experiment were used in this experiment. Thus, the second experiment was independent, but used the same calibration and can, therefore, verify the results from the previous experiment. In addition, the amount of data is larger and increases the credibility of the system. The timing error for the images without using SyncBoard was 0.29 s in this experiment compared to 0.44 s in the previous experiment.

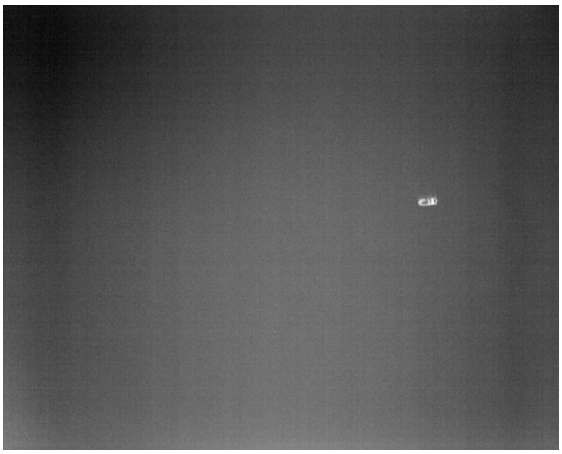

Figure 15: The small marine vessel used in the second experiment.

Figure 16 shows the path of the UAV. The UAV operated at an altitude of approximately 200 m during the periods where the vessel was recorded. The vessel was in the field of view of the camera in many different periods and the variation in UAV poses during detection is much greater in this experiment.

Figure 17 shows the distribution of georeferenced positions for a subset of 1250 images using the nonlinear observer. Figure 18 shows the distribution using the autopilot. A subset of 1250 images is shown in Figures 17 and 18 to enlarge the level of details. Figure 19 shows the distribution of georeferenced points in the entire flight using the nonlinear observer and verifies that the system works well on all images. The segment of 1250 images can be recognized in the upper left corner of Figure 19. The distributions in Figure 17 and 18 fit well with the GPS reference. The distribution from the nonlinear observer is perhaps following the trajectory of the reference better than the distribution from the autopilot for the subset of 1250 images, but the difference is minor.

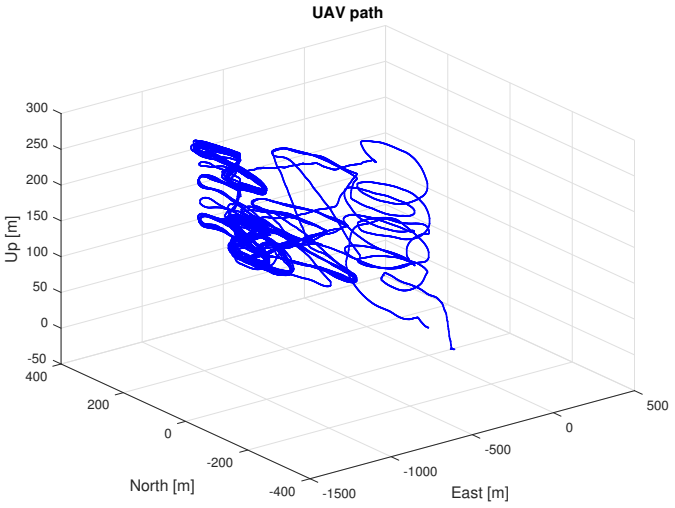


Figure 16: The UAV path estimated by the nonlinear observer in the second experiment.

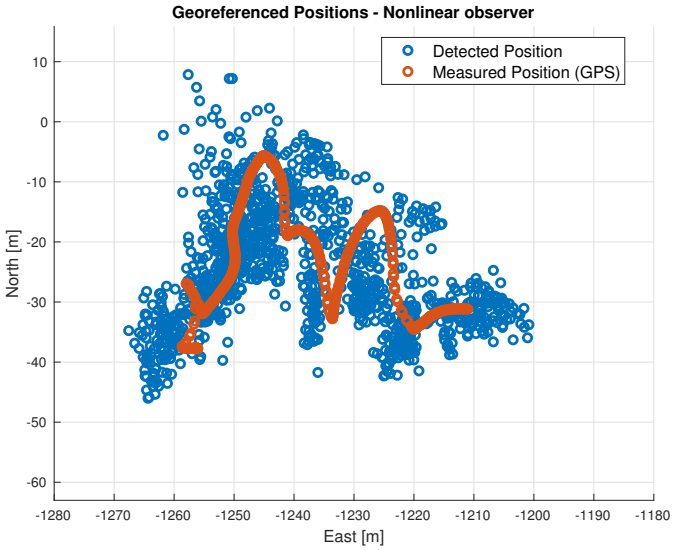


Figure 17: The distribution of georeferenced positions using the nonlinear observer for a subset of 1250 images.

Table 3: Main results of experiment 2. SD = standard deviation

| | **Observer** | **Pixhawk Autopilot** |
|---|---|---|
| Mean north error | 1.98 m | 0.37 m |
| Mean east error | −0.76 m | 3.38 m |
| SD north error | 7.77 m | 7.83 m |
| SD east error | 7.96 m | 7.50 m |
| Mean absolute error (single image) | 10.25 m | 9.37 m |

Table 3 summarizes the accuracy of the georeferenced positions for both the nonlinear observer and the autopilot for the entire flight. Table 4 shows the same metrics using the subset of 1250 images. The mean error for the entire flight using the observer is 1.98 m and −0.76 m in north and east, respectively. The standard deviation of the error is comparable for the observer and autopilot. The mean error of a single image is 10.25 m and 9.31 m for the observer and autopilot, respectively. Thus, the performance is comparable and in line with what was observed in the first experiment. The observer has the best performance for the reduced set of images, but the difference is still minor.

Table 4: Main results of experiment 2 for the subset of 1250 images used in Figures 17 and 18. SD = standard deviation

| | **Observer** | **Pixhawk Autopilot** |
|---|---|---|
| Mean north error | −0.68 m | −3.53 m |
| Mean east error | −0.37 m | 2.89 m |
| SD north error | 4.88 m | 6.37 m |
| SD east error | 6.09 m | 4.99 m |
| Mean absolute error (single image) | 7.29 m | 7.40 m |

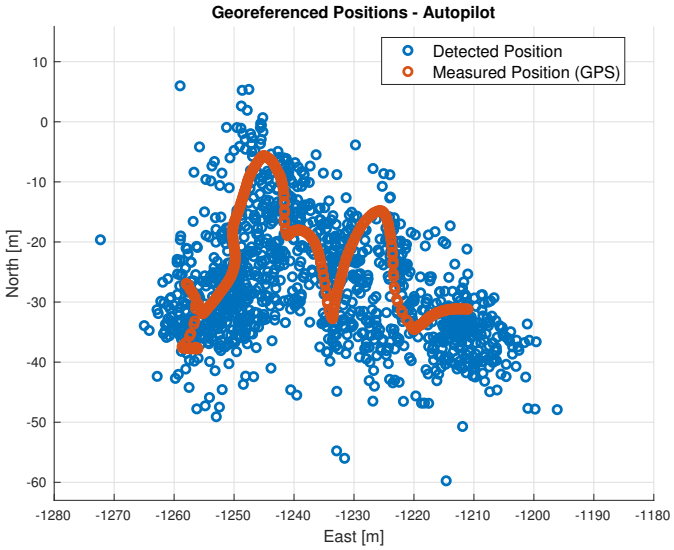


Figure 18: The distribution of georeferenced positions using the autopilot for the subset of 1250 images.

The mean absolute error in a single image is larger than in the first experiment (10.25 m compared to 6.40 m), even though the altitude of the UAV was lower. This is explained by the reference, which is more uncertain. In addition, the vessel obviously moves more than the stationary GPS base station antenna and the GPS receiver used as reference was not located exactly in the center of the vessel as assumed in the object detection algorithm. Since the images are captured from a larger set of different UAV poses and maneuvers than in the first experiment, the accuracy in this experiment is reasonable. The mean position has an overall accuracy just surpassing 2 m, which means that the target position is located in a trustworthy manner

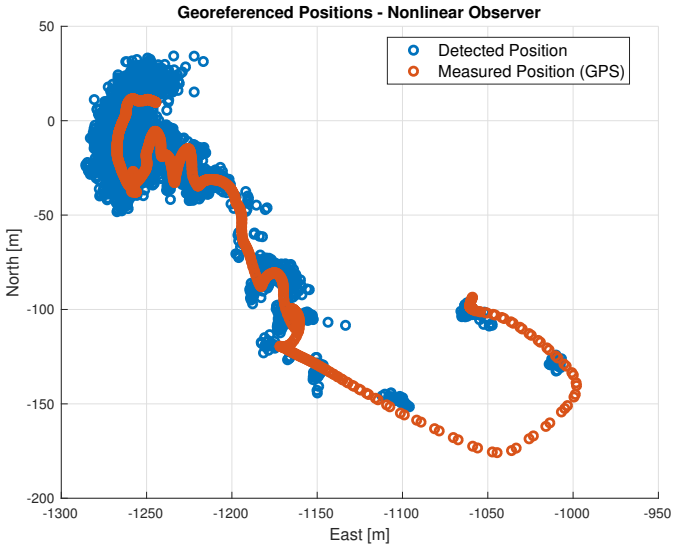


Figure 19: The distribution of georeferenced positions using the nonlinear observer for all images.

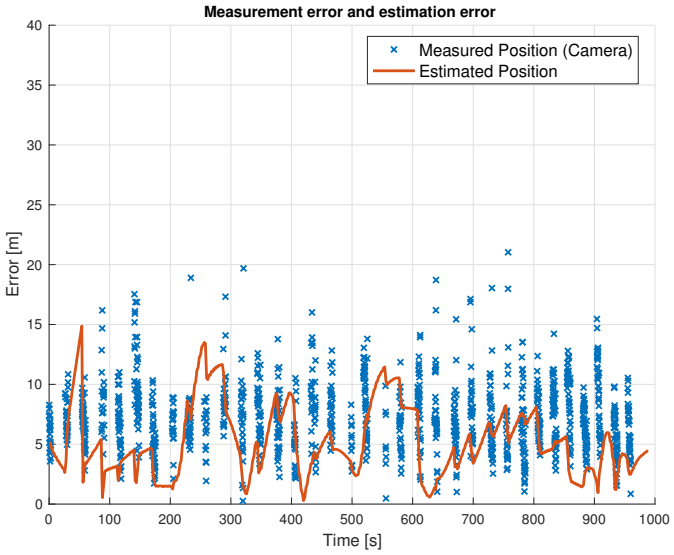


Figure 20: The georeferencing error and the estimation error when a Kalman filter is used to track the target. This shows the tracking results for the subset of 1250 images.

and that the true position can be estimated from a small set of images.

Figure 20 shows the measurement and estimation error when a Kalman filter with a constant velocity motion model is used to track the vessel for the subset of 1250 images. Figure 21 shows the tracking results for the entire flight. The initial error is larger for the entire flight since the vessel had a larger velocity in the beginning and thus the Kalman filter is struggling initially (the filter is initialized with zero speed). It is possible to tune the Kalman filter in a better way to reduce the initial error, but is not necessary since the intention of these figures is to show the achievable estimation error and the magnitude of the error in the measurements. The largest measurement error for the entire flight is almost 40 m, but the majority of the measurements have an error below 20 m. The measurement error grows somewhat at the end as displayed in Figure 21, but the estimation error is still small and within a few meters. The better part of the measurements are between five and ten meters from the reference.

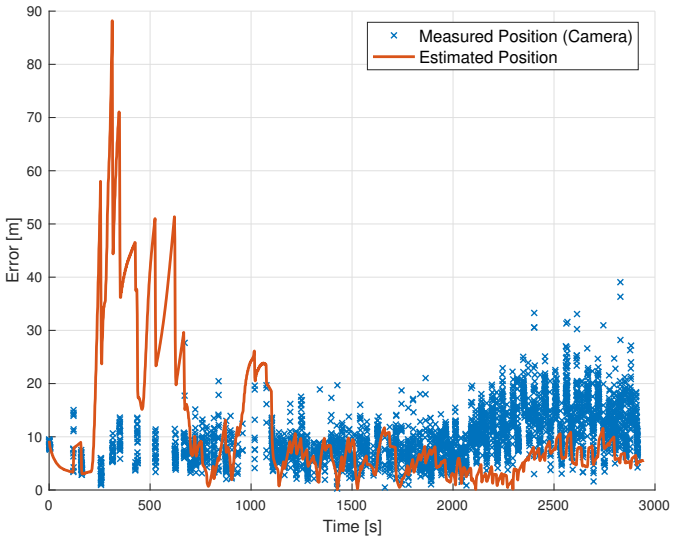


Figure 21: The georeferencing error and the estimation error when a Kalman filter is used to track the target. This shows the tracking results for the entire flight.

## 6. CONCLUSIONS

A system for real-time georeferencing of detected objects using a thermal camera in fixed-wing UAVs has been proposed and validated experimentally in this work. A camera calibration surface for thermal cameras has been described and calibration of misalignment errors has also been covered. To ensure operational flexibility, direct georeferencing was used without the need for ground reference points. The need for a time-synchronized and accurate navigation system is demanding for fixed-wing UAVs. This work has confirmed that it is necessary to calibrate for misalignment errors and handle synchronization among the sensors in an accurate manner. The results also indicate that these considerations are more important than the quality of the sensors in the navigation system. Moreover, thermal cameras have limited spatial resolution compared to visual spectrum imagery. Therefore, georeferencing using thermal images is more sensitive to noise during object detection because a single pixel covers a larger geographical area.

The effect of IMU misalignment calibration and time synchronization depends on the path. On straight-line segments, the attitude is constant and time synchroniza-

tion is less vital. This is because synchronization errors in level flight only lead to an error in the UAV position. The attitude on the other hand is correct and Appendix B shows that attitude errors are more severe. However, during turns or when the attitude is changing quickly, proper synchronization is key as synchronization errors give inaccuracies in the attitude. Misalignment errors are devastating during all maneuvers, but is especially troublesome at high altitudes. Therefore, proper mounting of the sensors is perhaps the most important consideration.

## Appendix A. Observer Design

This appendix describes the nonlinear observer used as navigation system (Section 3) in detail. The observer design is based on a kinematic model that is assumed to represent the true relationship between the states.

### *Appendix A.1. Strapdown Equations*

The states of interest in the navigation system are the position ($\mathbf{p}$), linear velocity ($\mathbf{v}$) and attitude of the body frame, denoted {b}, relative the Earth-fixed-Earth-Centered (ECEF) coordinate frame, denoted {e}. The two former are decomposed in {e}. A review of methods for attitude estimation is presented in (Crassidis et al., 2007). In this research, attitude is parametrized as a unit quaternion ($\boldsymbol{q}_b^e$) and a rotation matrix between ECEF and the body-fixed frame of the UAV, denoted $\mathbf{R}_b^e$.

The kinematics (strapdown equations) are used to design the observer equations, and described mathematically as

$$\dot{\boldsymbol{p}}_{eb}^e = \boldsymbol{v}_{eb}^e \tag{A.1}$$

$$\dot{\boldsymbol{v}}_{eb}^e = -2\boldsymbol{S}(\boldsymbol{\omega}_{ie}^e)\boldsymbol{v}_{eb}^e + \boldsymbol{R}_b^e\boldsymbol{f}_{ib}^b + \boldsymbol{g}_b^e \tag{A.2}$$

$$\dot{\boldsymbol{q}}_b^e = \frac{1}{2}\boldsymbol{q}_b^e \otimes \begin{pmatrix} 0 \\ \boldsymbol{\omega}_{ib}^b \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 0 \\ \boldsymbol{\omega}_{ie}^e \end{pmatrix} \otimes q_b^e \tag{A.3}$$

with

$$\boldsymbol{\omega}_{ie}^e = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \omega_{ie}, \tag{A.4}$$

where $\omega_{ie}$ is the Earth's rotation rate. The nonlinear observer is structurally the same as in (Grip et al., 2013).

### *Appendix A.2. Attitude Observer*

The nonlinear observer for estimating the attitude between $\{b\}$ and $\{e\}$ is given similar to (Grip et al., 2013),

$$\Sigma_1 : \begin{cases} \dot{\hat{\boldsymbol{q}}}_b^e = \frac{1}{2}\hat{\boldsymbol{q}}_b^e \otimes \begin{pmatrix} 0 \\ \hat{\boldsymbol{\omega}}_{ib}^b \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 0 \\ \boldsymbol{\omega}_{ie}^e \end{pmatrix} \otimes \hat{\boldsymbol{q}}_b^e, & \text{(A.5a)} \\ \hat{\boldsymbol{\omega}}_{ib}^b = \boldsymbol{\omega}_{\text{IMU}}^b - \hat{\boldsymbol{b}}_{\text{ars}}^b + \hat{\boldsymbol{\sigma}}_{ib}^b, & \text{(A.5b)} \\ \dot{\hat{\boldsymbol{b}}}_{\text{ars}}^b = \text{Proj}\left(\hat{\boldsymbol{b}}_{\text{ars}}^b, -k_I\hat{\boldsymbol{\sigma}}_{ib}^b\right), & \text{(A.5c)} \end{cases}$$

where $\hat{\boldsymbol{b}}_{\text{ars}}^b$ is the estimate of the angular rate sensor (ARS) bias. $\text{Proj}(\star,\star)$ denotes the ARS bias projection algorithm ensuring that $\|\hat{\boldsymbol{b}}_{\text{ars}}^b\|_2 \leq M_{\hat{b}_{\text{ars}}}$ for $M_{\hat{b}_{\text{ars}}} > M_{b_{\text{ars}}}$ (Grip et al., 2012), and $k_I$ is the gain associated with the ARS bias estimation. Moreover, the nonlinear injection term, $\hat{\boldsymbol{\sigma}}_{ib}^b$, is given as

$$\hat{\boldsymbol{\sigma}}_{ib}^b = k_1\underline{\boldsymbol{v}}_1^b \times \boldsymbol{R}^\intercal(\hat{\boldsymbol{q}}_b^e)\underline{\boldsymbol{v}}_1^e + k_2\underline{\boldsymbol{v}}_2^b \times \boldsymbol{R}^\intercal(\hat{\boldsymbol{q}}_b^e)\underline{\boldsymbol{v}}_2^e, \tag{A.6}$$

where the measurement vectors $\underline{\boldsymbol{v}}_{1,2}^b$ and reference vectors $\underline{\boldsymbol{v}}_{1,2}^e$ are calculated using

$$\underline{\boldsymbol{v}}_1^b = \underline{\boldsymbol{f}}^b, \qquad \underline{\boldsymbol{v}}_1^e = \underline{\boldsymbol{f}}^e, \tag{A.7}$$

$$\underline{\boldsymbol{v}}_2^b = \underline{\boldsymbol{f}}^b \times \underline{\boldsymbol{p}}^b, \quad \underline{\boldsymbol{v}}_2^e = \underline{\boldsymbol{f}}^e \times \underline{\boldsymbol{p}}^e. \tag{A.8}$$

Furthermore, the measurement and corresponding reference vector pairs in (A.7)–(A.8) are constructed as

$$\underline{\boldsymbol{f}}^b = \frac{\boldsymbol{f}_{\text{IMU}}^b}{\|\boldsymbol{f}_{\text{IMU}}^b\|_2}, \qquad \underline{\boldsymbol{f}}^e = \frac{\text{sat}_{M_f}(\hat{\boldsymbol{f}}_{ib}^e)}{\|\text{sat}_{M_f}(\hat{\boldsymbol{f}}_{ib}^e)\|_2}, \tag{A.9}$$

$$\underline{\boldsymbol{p}}^b = \frac{\boldsymbol{p}_{ba_1,1}^b - \boldsymbol{p}_{ba_1,0}^b}{\|\boldsymbol{p}_{ba_1,1}^b - \boldsymbol{p}_{ba_1,0}^b\|_2} \quad \underline{\boldsymbol{p}}^e = \frac{\boldsymbol{p}_{\text{RTK},1}^e - \boldsymbol{p}_{\text{RTK},0}^b}{\|\boldsymbol{p}_{\text{RTK},1}^e - \boldsymbol{p}_{\text{RTK},0}^b\|_2}, \tag{A.10}$$

where $\boldsymbol{p}_{ba_k,k}^b$ is the relative vector between the two antenna positions given in body. $\hat{\boldsymbol{f}}_{ib}^e$ is the estimated specific force, provided by the TMO, presented next. $\mathbf{p}_{RTK,0}^b$ and $\mathbf{p}_{RTK,1}^b$ are the position of the GPS antennas given in {b}.

### *Appendix A.3. Translational Motion Observer*

The translational motion observer (TMO) is similar to that of (Grip et al., 2013), and given as follows:

$$\Sigma_2 : \begin{cases} \dot{\hat{\boldsymbol{p}}}_{eb}^e = \hat{\boldsymbol{v}}_{eb}^e + \vartheta\boldsymbol{K}_{pp}^0\tilde{\boldsymbol{y}}_{eb}^e & \text{(A.11a)} \\ \dot{\hat{\boldsymbol{v}}}_{eb}^e = -2\boldsymbol{S}(\boldsymbol{\omega}_{ie}^e)\boldsymbol{v}_{eb}^e + \hat{\boldsymbol{f}}_{ib}^e + \\ \qquad \boldsymbol{g}_b^e(\hat{\boldsymbol{p}}_{eb}^e) + \vartheta^2\boldsymbol{K}_{vp}^0\tilde{\boldsymbol{y}}_{eb}^e & \text{(A.11b)} \\ \dot{\hat{\boldsymbol{\xi}}}_{ib}^e = -\boldsymbol{R}(\hat{\boldsymbol{q}}_b^e)\boldsymbol{S}(\hat{\boldsymbol{\sigma}}_{ib}^b)\boldsymbol{f}_{\text{IMU}}^b + \vartheta^3\boldsymbol{K}_{\xi p}^0\tilde{\boldsymbol{y}}_{eb}^e & \text{(A.11c)} \\ \hat{\boldsymbol{f}}_{ib}^e = \boldsymbol{R}(\hat{\boldsymbol{q}}_b^e)\boldsymbol{f}_{\text{IMU}}^b + \boldsymbol{\xi}_{ib}^e, & \text{(A.11d)} \end{cases}$$

where

$$\tilde{\boldsymbol{y}}_{eb}^e = \boldsymbol{p}_{RTK,k}^e - \hat{\boldsymbol{p}}_{eb}^e \tag{A.12}$$

, where $k \in [0,1]$, and $\boldsymbol{K}_\star$ are gains associated with the RTK position measurement. $\boldsymbol{\xi}_{ib}^e$ is an auxiliary state used to estimate $\boldsymbol{f}_{ib}^e$. $\vartheta$ is a high-gain like parameter used to guarantee stability. Furthermore, by noting the linear time-varying (LTV) structure of (A.11) and defining

$$\boldsymbol{x} := \left(\boldsymbol{p}_{eb}^e;\quad \boldsymbol{v}_{eb}^e;\quad \boldsymbol{\xi}_{ib}^e\right), \tag{A.13}$$

the TMO can be written on LTV form as

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}(t)\boldsymbol{u} + \boldsymbol{D}(t,\hat{\boldsymbol{x}}) + \boldsymbol{K}(t)(\boldsymbol{y} - \boldsymbol{C}\hat{\boldsymbol{x}}), \tag{A.14}$$

with the system matrices,

$$\boldsymbol{A} = \begin{pmatrix} \mathbf{0}_{3\times3} & \boldsymbol{I}_3 & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \boldsymbol{I}_3 \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{pmatrix}, \boldsymbol{B}(t) = \begin{pmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \boldsymbol{R}(\hat{\boldsymbol{q}}_b^e) & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \boldsymbol{R}(\hat{\boldsymbol{q}}_b^e) \end{pmatrix}, \tag{A.15}$$

the measurement matrix,

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{I}_3 & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \end{pmatrix}, \tag{A.16}$$

the vector,

$$\boldsymbol{D}(t,\hat{\boldsymbol{x}}) = \begin{pmatrix} \boldsymbol{0}_{3\times1}; & -2\boldsymbol{S}(\boldsymbol{\omega}^e_{ie})\hat{\boldsymbol{v}}^e_{eb} + \boldsymbol{g}^e_b(\hat{\boldsymbol{p}}^e_{eb}); & \boldsymbol{0}_{3\times1} \end{pmatrix}, \tag{A.17}$$

and the gain matrix,

$$\boldsymbol{K}(t) = \begin{pmatrix} \boldsymbol{K}_{pp} \\ \boldsymbol{K}_{vp} \\ \boldsymbol{K}_{\xi p} \end{pmatrix} = \begin{pmatrix} \vartheta \boldsymbol{K}^0_{pp} \\ \vartheta^2 \boldsymbol{K}^0_{vp} \\ \vartheta^3 \boldsymbol{K}^0_{\xi p} \end{pmatrix}, \tag{A.18}$$

where

$$\boldsymbol{K}^0(t) = \begin{pmatrix} (\boldsymbol{K}^0_{pp})^\intercal & (\boldsymbol{K}^0_{vp})^\intercal & (\boldsymbol{K}^0_{\xi p})^\intercal \end{pmatrix}^\intercal \tag{A.19}$$

is given as $\boldsymbol{K}^0(t) = \boldsymbol{P}(t)\boldsymbol{C}^\intercal\boldsymbol{R}^{-1}(t)$, with $\boldsymbol{P}(t) = \boldsymbol{P}^\intercal(t) > 0$ being the solution of the time-scaled Riccati equation

$$\begin{aligned} \frac{1}{\vartheta}\dot{\boldsymbol{P}}(t) = {} & \boldsymbol{A}\boldsymbol{P}(t) + \boldsymbol{P}(t)\boldsymbol{A}^\intercal - \boldsymbol{P}\boldsymbol{C}^\intercal\boldsymbol{R}^{-1}(t)\boldsymbol{C}^\intercal\boldsymbol{P}(t) \\ & + \boldsymbol{B}(\hat{\boldsymbol{q}}^e_b)\boldsymbol{Q}(t)\boldsymbol{B}^\intercal(\hat{\boldsymbol{q}}^e_b). \end{aligned} \tag{A.20}$$

Finally, the input is given as

$$\boldsymbol{u} = \left( \boldsymbol{f}^b_{\text{IMU}}; -\boldsymbol{S}(\hat{\boldsymbol{\sigma}}^b_{ib})\boldsymbol{f}^b_{\text{IMU}} \right). \tag{A.21}$$

This results in the origin of the error dynamics being semi-globally exponentially stable. See (Grip et al., 2013) for details on the stability analysis. In addition, a given lever arm from the IMU to any of the GPS antennas may be incorporated into the TMO by replacing (A.12) with

$$\tilde{\boldsymbol{y}}^e_{eb,k} = \boldsymbol{p}^e_{RTK,k} - \hat{\boldsymbol{p}}^e_{eb} - \mathbf{R}(\boldsymbol{q}^e_b)\mathbf{r}^b_b, \tag{A.22}$$

where $\mathbf{r}^b_b$ represents the lever arm and $k \in [0,1]$ is the index of the RTK position solution.

## Appendix B. Analysis of how the UAV pose influences the georeferencing

This appendix seeks to analyze how an error in the UAV navigation system (and thus the camera pose) influences georeferencing. This is achieved with three simplified cases that are meant to illustrate level flight (zero roll, pitch and yaw), turn (roll angle of 25 degrees) and ascent (pitch angle of 20 degrees). All cases are simulated at an altitude of 100 and 400 meters. Moreover, the UAV has north and east positions of zero and the camera parameters are equal to the ones used in the flight experiments. The object is assumed to be in the middle of the image. Note that the magnitude of the georeferencing error also depends on the position in the image so larger errors must be expected closer to the boundaries. Therefore, the values shown here can be interpreted as the best-case scenario for different navigation errors since the georeferencing error increases further away from the image center. Table B.5 shows the error in the georeferenced position if a small bias is added to the UAV navigation states in level flight. Table B.6 and Table B.7 show the error during turn and ascent, respectively.

Table B.5: Error in georeferencing when there exist a bias in the UAV attitude and position for level flight.

| **Offset** | **Error with altitude 100m** | **Error with altitude 400m** |
|---|---|---|
| Roll offset 3° | 5.3 m | 21 m |
| Pitch offset 3° | 5.3 m | 21.2 m |
| Yaw offset 3° | 0.5 m | 1.7 m |
| North offset 10 m | 10 m | 10 m |
| East offset 10 m | 10 m | 10 m |
| Down offset 10 m | 0.8 m | 0.8 m m |
| Roll and pitch offset 3° | 7.5 m | 30 m |

Table B.6: Error in georeferencing when there exist a bias in the UAV attitude and position during a turn (roll angle of 25 degrees).

| **Offset** | **Error with altitude 100m** | **Error with altitude 400m** |
|---|---|---|
| Roll offset 3° | 6.8 m | 27 m |
| Pitch offset 3° | 5.3 m | 21.2 m |
| Yaw offset 3° | 2.7 m | 10.7 m |
| North offset 10 m | 10 m | 10 m |
| East offset 10 m | 10 m | 10 m |
| Down offset 10 m | 5.1 m | 5.1 m |
| Roll and pitch offset 3° | 9 m | 36 m |

A factor worth noticing is that an error in the down position (altitude) leads to the same error at both 100 and 400 meters, but that the error is larger during a turn and ascent compared to level flight. This is expected as the ray from the camera center points straight towards the ground in level flight, but a longer distance during a turn or ascent. In addition to these factors, it is important to notice that the error is proportional to the altitude of the UAV when offsets in the attitude are considered. Notice also the magnitude of the error at an altitude of 400 meters with a small bias in both roll and pitch (30 m to 36 m). This illustrates the necessity of synchronizing the sensors and conduct a proper calibration since even small biases give large errors in georeferencing.


### Acknowledgments

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

Table B.7: Error in georeferencing when there exist a bias in the UAV attitude and position during ascent (pitch angle of 20 degrees).

| **Offset** | **Error with altitude 100m** | **Error with altitude 400m** |
|---|---|---|
| Roll offset 3° | 5.8 m | 23 m |
| Pitch offset 3° | 6.5 m | 25.9 m |
| Yaw offset 3° | 2.4 m | 9.5 m |
| North offset 10 m | 10 m | 10 m |
| East offset 10 m | 10 m | 10 m |
| Down offset 10 m | 4.5 m | 4.5 m |
| Roll and pitch offset 3° | 8.9 m | 35.4 m |

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. The authors would also like to thank Krzysztof Cisek and Dr. Robert Rogne for their contributions in the development of the payload.

## References

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

Beard, R.W., McLain, T.W., 2012. Small unmanned aircraft: Theory and practice. Princeton university press.

Choi, K., Lee, I., 2013. A sequential aerial triangulation algorithm for real-time georeferencing of image sequences acquired by an airborne multi-sensor system. Remote Sensing 5, 57–82. doi:10.3390/rs5010057.

Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing 92, 79–97. doi:10.1016/j.isprsjprs.2014.02.013.

Crassidis, J.L., Markley, F.L., Cheng, Y., 2007. Survey of nonlinear attitude estimation methods. Journal of Guidance, Control, and Dynamics 30, 12–28. doi:10.2514/1.22452.

Daakir, M., Pierrot-Deseilligny, M., Bosser, P., Pichard, F., Thom, C., Rabot, Y., Martin, O., 2017. Lightweight uav with on-board photogrammetry and single-frequency gps positioning for metrology applications. ISPRS Journal of Photogrammetry and Remote Sensing 127, 115 – 126. doi:10.1016/j.isprsjprs.2016.12.007.

Deng, C., Wang, S., Huang, Z., Tan, Z., Liu, J., 2014. Unmanned aerial vehicles for power line inspection: A cooperative way in platforms and communications. Journal of Communications 9, 687–692. doi:10.12720/jcm.9.9.687-692.

Deng, Z., Sun, H., Zhou, S., Zhao, J., Lei, L., Zou, H., 2018. Multi-scale object detection in remote sensing imagery with convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing 145, 3 – 22. doi:10.1016/j.isprsjprs.2018.04.003.

Euston, M., Coote, P., Mahony, R., Kim, J., Hamel, T., 2008. A complementary filter for attitude estimation of a fixed-wing uav. 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems , 340–345doi:10.1109/IROS.2008.4650766.

Farrell, J.A., 2008. Aided Navigation: GPS with High Rate Sensors. McGraw-Hill.

Fossen, T., 2011. Handbook of Marine Craft Hydrodynamics and Motion Control. John Wiley & Sons.

Fusini, L., Fossen, T.I., Johansen, T.A., 2018. Nonlinear observers for gnss- and camera-aided inertial navigation of a fixed-wing uav. IEEE Transactions on Control Systems Technology 26, 1884–1891. doi:10.1109/TCST.2017.2735363.

Grip, H.F., Fossen, T.I., Johansen, T.A., Saberi, A., 2012. Attitude estimation using biased gyro and vector measurements with time-varying reference vectors. IEEE Transactions on Automatic Control 57, 1332—1338. doi:10.1109/TAC.2011.2173415.

Grip, H.F., Fossen, T.I., Johansen, T.A., Saberi, A., 2013. Nonlinear observer for gnss-aided inertial navigation with quaternion-based attitude estimation. American Control Conference , 272–279.doi:10.1109/ACC.2013.6579849.

Hartmann, W., Tilch, S., Eisenbeiss, H., Schindler, K., 2012. Determination of the uav position by automatic processing of thermal images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39, B6. doi:10.5194/isprsarchives-XXXIX-B6-111-2012.

Harwin, S., Lucieer, A., 2012. Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (uav) imagery. Remote Sensing 4, 1573–1599. doi:10.3390/rs4061573.

Helgesen, H.H., Leira, F.S., Fossen, T.I., Johansen, T.A., 2017a. Tracking of ocean surface objects from unmanned aerial vehicles with a pan/tilt unit using a thermal camera. Journal of Intelligent & Robotic Systems doi:10.1007/s10846-017-0722-3.

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

Hemerly, E.M., 2014. Automatic georeferencing of images acquired by uav's. International Journal of Automation and Computing 11, 347–352. doi:10.1007/s11633-014-0799-0.

Hosen, J., Helgesen, H.H., Fusini, L., Fossen, T.I., Johansen, T.A., 2016. Vision-aided nonlinear observer for fixed-wing unmanned aerial vehicle navigation. Journal of Guidance, Control, and Dynamics 39, 1777–1789. doi:10.2514/1.G000281.

Hutchinson, S., Hager, G., Corke, P., 1996. A tutorial on visual servo control. IEEE Transactions on Robotics and Automation 12, 651–670. doi:10.1109/70.538972.

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

Leira, F.S., Johansen, T.A., Fossen, T.I., 2017. A uav ice tracking framework for autonomous sea ice management. The International Conference on Unmanned Aircraft Systems , 581–590doi:10.1109/ICUAS.2017.7991435.

Leira, F.S., Trnka, K., Fossen, T.I., Johansen, T.A., 2015b. A ligth-weight thermal camera payload with georeferencing capabilities for small fixed-wing uavs. The International Conference on Unmanned Aircraft Systems , 485–494doi:10.1109/ICUAS.2015.7152327.

Mahony, R., Hamel, T., Pflimlin, J.M., 2008. Nonlinear complementary filters on the special orthogonal group. IEEE Transactions on Automatic Control 53, 1203–1218. doi:10.1109/TAC.2008.923738.

Nikolic, J., Burri, M., Rehder, J., Leutenegger, S., Huerzeler, C., Siegwart, R., 2013. A uav system for inspection of industrial facilities. Proceedings of the 2013 IEEE Aerospace Conference , 1–8doi:10.3929/ethz-a-010061810.

OptiTrack Motion Capture Systems, 2018. URL: http://optitrack.com. last accessed 2018-08-14.

Pajares, G., 2015. Overview and current status of remote sensing applications based on unmanned aerial vehicles (uavs). Photogrammetric Engineering & Remote Sensing 81, 281–329. doi:10.14358/PERS.81.4.281.

Pixhawk 4 - By Dronecode, 2018. URL: https://docs.px4.io/en/flight_controller/pixhawk4.html. last accessed 2018-08-03.

Rehak, M., Mabillard, R., Skaloud, J., 2013. A micro-uav with the capability of direct georeferencing. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci , 317–323doi:10.5194/isprsarchives-XL-1-W2-317-2013.

Rodin, C.D., Johansen, T.A., 2018. Detectability of objects at the sea surface in visible light and thermal camera images. MTS/IEEE OCEANS Conference .

Sakamaki, J.Y., Beard, R.W., Rice, M., 2017. Tracking multiple ground objects using a team of unmanned air vehicles, in: Fossen, T.I., Pettersen, K.Y., Nijmeijer, H. (Eds.), Sensing and Control for Autonomous Vehicles: Applications to Land, Water and Air Vehicles. Springer International Publishing, pp. 249–268. doi:10.1007/978-3-319-55372-6_12.

Sola, J., 2017. Quaternion kinematics for the error-state kf. URL: http://www.iri.upc.edu/people/jsola/JoanSola/objectes/notes/kinematics.pdf. last accessed 2018-02-02.

Sollie, M.L., Bryne, T.H., Johansen, T.A., 2019. Pose estimation of uavs based on ins aided by two independent low-cost gnss receivers. The International Conference on Unmanned Aircraft Systems .

ThermalCapture Grabber USB 640 - TEAX Technology, 2018. URL: http://thermalcapture.com/thermalcapture-grabber-usb-640/. last accessed 2018-08-14.

Toth, C., Jóźków, G., 2016. Remote sensing platforms and sensors: A survey. ISPRS Journal of Photogrammetry and Remote Sensing 115, 22 – 36. doi:10.1016/j.isprsjprs.2015.10.004.

Vermeille, H., 2004. Computing geodetic coordinates from geocentric coordinates. Journal of Geodesy 78, 94—95. doi:10.1007/s00190-004-0375-4.

Xiang, H., Tian, L., 2011. Method for automatic georeferencing aerial remote sensing (rs) images from an unmanned aerial vehicle (uav) platform. Biosystems Engineering 108, 104 – 113. doi:10.1016/j.biosystemseng.2010.11.003.

Yahyanejad, S., Misiorny, J., Rinner, B., 2011. Lens distortion correction for thermal cameras to improve aerial imaging with small-scale uavs. IEEE International Symposium on Robotic and Sensors Environments (ROSE) , 231–236doi:10.1109/ROSE.2011.6058528.

Yahyanejad, S., Rinner, B., 2015. A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale uavs. ISPRS Journal of Photogrammetry and Remote Sensing 104, 189 – 202. doi:10.1016/j.isprsjprs.2014.07.015.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 1330–1334. URL: https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/.

Zhou, G., Li, C., Cheng, P., 2005. Unmanned aerial vehicle (uav) real-time video registration for forest fire monitoring. The IEEE International Geoscience and Remote Sensing Symposium. 3, 1803–1806. doi:10.1109/IGARSS.2005.1526355.