2016 International Conference on
Unmanned Aircraft Systems (ICUAS)
June 7-10, 2016. Arlington, VA USA

Tracking of Marine Surface Objects from Unmanned Aerial Vehicles with a Pan/Tilt Unit using a Thermal Camera and Optical Flow

Håkon Hagen Helgesen[1], Frederik Stendahl Leira[1], Tor Arne Johansen[1] and Thor I. Fossen[1]

Abstract— **This paper presents a vision-based tracking system for marine surface objects utilizing images captured in a fixed-wing unmanned aerial vehicle (UAV) with a retractable pan/tilt gimbal. A formula for calculating the North-East velocities of objects detected in the images is derived. The formula utilizes optical flow (OF) and compensates for the UAV and gimbal motions. It can be used together with georeferencing to obtain position and velocity measurements of objects detected in the images. The tracking system is suitable with both infrared and visual spectrum images. A flight experiment has been conducted near the Azores, Portugal in 2015 to gather thermal images with marine surface objects. The tracking system has been evaluated off-line and the results show that a marine vessel can be tracked in the North-East-Down (NED) plane.**

I. INTRODUCTION

Unmanned aerial vehicles are in their golden age and the field of applications grows rapidly. Visual sensors, such as an infrared or visual spectrum camera, are often a part of the UAV payload today. Cameras can be useful for search and rescue applications [1] [2], navigation [3] [4], vehicle tracking [5] and obviously also in many other applications.

An application where UAVs equipped with a visual sensor can be useful is autonomous ship control. Autonomous ships need to obey the International Regulations for Preventing Collisions at Sea (COLREGS) [6]. The main challenge related to COLREGS is collision avoidance. Autonomous ships require a system for detecting and keeping track of obstacles near the planned path of the ship. On-board sensors such as a radar, LIDAR and camera can be used to detect objects in the environment of the ship [7] [8] [9]. However, in order to have a robust system it might not be sufficient to only place sensors on-board the ship. Limited range and resolution as well as small objects floating in the waterline can make it challenging to detect objects. UAVs can be used to overcome some of the challenges and make a robust system in combination with other sensors on-board the ship. UAVs equipped with a visual sensor can be used to scan the planned path of the ship and detect objects that are difficult to detect with ship sensors [10]. It is necessary to keep track of the objects for a certain time period for making decisions that obey COLREGS. This is where object detection and tracking become important.

[1]Håkon Hagen Helgesen, Frederik Stendahl Leira, Tor Arne Johansen and Thor I. Fossen are with the NTNU Centre for Autonomous Marine Operations and Systems, Department of Engineering Cybernetics, Norwegian University of Science and Technology, NTNU, 7491 Trondheim, Norway. (hakon.helgesen@itk.ntnu.no, frederik.s.leira@itk.ntnu.no, tor.arne.johansen@itk.ntnu.no, thor.fossen@ntnu.no)

Object detection is the process of detecting objects of importance with respect to some criterion. Tracking is the process of generating a time-dependent position (often also velocity) trajectory for objects detected in a sequence of images. Object detection and tracking have been studied thoroughly and the research on the topics is mature [11]. However, the focus has traditionally been directed towards applications where the sensor is at rest or moving slowly. This is especially the case for segmentation techniques used to find moving objects. Fixed-wing UAVs are operating at relatively high velocity, which causes the images captured on-board to be more contaminated by blur than images captured at rest. Moreover, the scene changes rapidly, which makes many conventional detection methods inappropriate for UAVs. Therefore, suitable algorithms for analyzing images captured from a fixed-wing UAV operating at high speed is an interesting research area.

Visual spectrum images might not be the best solution for object detection at sea. A thermal camera is a more attractive option since the temperature or emissivity difference between the sea surface and surface objects (such as a marine vessel) is often significant. A thermal camera has successfully been utilized to detect and track objects at sea in [12]. Furthermore, by using a georeferencing method, e.g [13], the position in an earth-fixed coordinate frame, such as the NED reference frame, can be obtained.

Optical flow can be defined as a velocity field that transforms one image into the next in a sequence of images [14] [15]. One distinguishes between sparse (OF calculated at a subset of the pixels or features) and dense (OF calculated at every pixel) OF methods. [14], [16] and [17] are examples of dense methods, while [18] and [19] are examples of sparse methods. The OF at a single pixel is represented as a vector describing the movement of the pixel (feature) between two images. Sparse methods often use point detectors, such as SIFT [18], SURF [19] or the KLT detector [20], to find features which can be used to calculate OF vectors. Different OF methods are described in [21] [22].

OF can be used to detect moving objects from a camera at rest. However, it is more difficult to detect moving objects when the camera is mounted in the payload of a moving UAV. The attitude, angular and linear velocities as well as gimbal motion are all important factors that make it hard to detect moving objects with OF. This is because the UAV motion often affects the OF more than the object motion.

OF has recently been used to recover the linear and angular velocities of a fixed-wing UAV [15], and the recovered linear velocity has been used as a measurement for estimating the

978-1-4673-9333-1/16/$31.00 ©2016 IEEE

attitude of a UAV using a nonlinear observer [3] [4].

A. Contributions of this Paper

This paper develops [15] and [3] further to derive the relationship between OF and the linear and angular velocities of a fixed-wing UAV equipped with a pan/tilt gimbal, which is a novel contribution. The gimbal motion leads to a more complicated relationship between OF and velocity, but increases the flexibility and operational range of the system. One advantage with the gimbal is the fact that a single area (or target) can be monitored over a longer period of time since the gimbal can to some extent compensate for the UAV motion. Therefore, the camera can be pointed at a fixed location during different UAV maneuvers or adjusted to keep a moving target within the field of view of the camera [23].

The relationship between OF and velocity can be used to estimate the linear and angular velocities of a UAV equipped with a pan/tilt gimbal by assuming that all features related to the OF vectors are at rest (belongs to the background). However, another interesting possibility, which is considered in this paper, is that a 2-dimensional velocity vector of the features can be obtained when the velocity of the UAV is known. This can be very useful at sea since surface objects only move in the 2-dimensional NE plane. Thus, it is possible to compute the NE velocities of the features and identify features that belong to moving objects. Additionally, by using a georeferencing technique the NED positions of the features can be obtained. By combining the velocity and position of features belonging to the same object, it is possible to track the object and estimate the velocity. This is convenient since information about the velocity can be used to predict the future trajectory of the object.

The proposed method is validated by off-line processing of thermal images captured at a flight experiment near the Azores island outside of Portugal in the summer of 2015. The method is suitable with both visual spectrum and thermal cameras, but thermal images were preferred to easier separate marine vessels from the sea. An experiment with a visual spectrum camera is described in [24]. The images contain several marine vessels operating at sea. SIFT [18] has been used to extract features and calculate OF vectors in the images from the experiment. Feature extraction in thermal images is not described extensively in the literature, but works well in the images gathered at the flight. Navigation data from an extended Kalman filter [25] are used together with the OF vectors to find the NE velocities of detected features. The NED positions of the features are obtained by georeferencing, and all features belonging to the same object have been grouped together. A tracking system, which estimates the position and velocity of the object is implemented, and it utilizes the calculated position and velocity as measurements. The estimates are compared with the position and velocity of the object, which is measured by GPS and used as a reference.

B. Organization of this Paper

The remainder of this paper is divided into five sections. Section II defines the notation necessary for understanding the derivations in the rest of the paper. Section III derives the relationship between OF and velocity of a moving target, and presents the tracking system. Section IV describes the experiment. The results are presented in Section V, before the paper is concluded in Section VI.

II. Notation and Preliminaries

Vectors and matrices are represented by lowercase and uppercase bold letters, respectively. $\mathbf{X}^{-1}$ denotes the inverse of a matrix and $\mathbf{X}^\top$ the transpose of a matrix or vector. A vector $\mathbf{x} = [x_1, x_2, x_3]^\top$ is represented in homogeneous coordinates as $\underline{\mathbf{x}} = [x_1, x_2, x_3, 1]^\top$. The operator $S(\mathbf{x})$ transforms the vector $\mathbf{x}$ into the skew-symmetric matrix

$$S(\mathbf{x}) = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}$$

and $\mathbf{0}_{m\times n}$ is a matrix of zeros with dimension $m \times n$.

Several reference frames are considered in this paper, but the three most important are: the body-fixed frame $\{\mathrm{B}\}$, the North-East-Down (NED) frame $\{\mathrm{N}\}$ (Earth-fixed, considered inertial) and the camera-fixed frame $\{\mathrm{C}\}$. The rotation from $\{\mathrm{N}\}$ to $\{\mathrm{B}\}$ is represented by the matrix $\mathbf{R}_b^n \equiv \mathbf{R} \in SO(3)$, with $SO(3)$ representing the Special Orthogonal group. Similar transformations exist between the other reference frames.

A vector decomposed in $\{\mathrm{B}\}$, $\{\mathrm{N}\}$ and $\{\mathrm{C}\}$ has superscript ^{b}, ^{n} and ^{c}, respectively. A point in the environment decomposed in $\{\mathrm{N}\}$ is $\mathbf{t}^n = [x^n, y^n, z^n]^\top$: note that a point located at sea level corresponds to $z^n = 0$. The same point decomposed in $\{\mathrm{B}\}$ is $\mathbf{t}^b = [x^b, y^b, z^b]^\top$.

The Greek letters ϕ, θ, and ψ represent the roll, pitch, and yaw angles, respectively, defined according to the zyx convention for principal rotations [26]. ψ_{gb} and θ_{gb} are the gimbal pan and tilt angles, which correspond to a rotation about the body z- and y-axis, respectively. A 2-dimensional camera image has coordinates (r, s) in the image plane. The derivative $[\dot{r}, \dot{s}]^\top$ of the image plane coordinates is the OF. r' and s' will also be used as the image plane coordinates. $s\theta$ and $c\theta$ denote the sine and cosine functions with θ as input. The subscript $_f$ is used to indicate that the corresponding parameter is related to a feature detected in the image. It should not be mixed with the letter f, which will be used for the focal length of a camera.

III. Computer Vision

This section presents the computer vision system necessary for detecting and tracking objects at the sea surface. The first part focuses on OF and briefly mentions some important methods. The second part explains how the NED positions of a pixel in the image can be recovered by georeferencing. The third part derives the relationship between OF and the NE velocities of detected objects (features). The latter part describes the tracking system. Classification [27] and data association [28] are not within the scope of this paper.

A. Optical Flow

A single OF vector can be understood as the 2-dimensional translation (in the image plane) of a feature detected in two consecutive images. Sparse methods, which calculates the OF vectors at a subset of the pixels based on feature extraction, will be the focus in this paper.

SIFT [18], which is the method utilized in the experiment, locates scale and rotation invariant features. In practice it means that features, which change in size and/or orientation with respect to the camera (between two images), can be detected in both images. This is a significant advantage in images captured from a UAV since the scale and rotation of objects in the images often change rapidly. Another advantage with SIFT (and other point detectors) is the fact that only the current image is used to find features. Thus, a change in background, which is bound to happen, does not necessarily affect the detection rate. This is not the case for methods relying on some sort of background subtraction/modeling.

Each detected feature has a descriptor, which is a vector consisting of properties related to the feature. The descriptors of features detected in consecutive images are matched together with a FLANN nearest neighbor search [29] to find common features within the images. OF vectors are created as the difference between the image plane position of common features in the images.

B. Recovering the NED Positions of a Pixel

[13] established how the NED positions of pixels can be recovered by georeferencing for a camera mounted in a pan/tilt gimbal. In this paper the georeferencing algorithm will be changed slightly, but the end result will be the same. Therefore, the reader who want the complete derivation should study [13] for a more comprehensive description.

The pinhole camera model [30] relates a point in the image plane with coordinates given by a camera-fixed coordinate frame $\{\mathrm{C}\}$. The relationship between the frames is displayed in Figure 1 and can be described as

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

Equation (1) describes the connection between the image plane coordinates (r, s) and the camera-fixed coordinates (x^c, y^c, z^c). z^c is the distance between the lens aperture and the plane the captured point is located in, and f is the focal length of the camera. Equation (1) can be expressed in matrix form as

$$z^c \begin{bmatrix} r \\ s \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^c \\ y^c \\ z^c \end{bmatrix} = \mathbf{A}\mathbf{p}^c \tag{2}$$

where $\mathbf{p}^c$ is the coordinates of a feature at pixel (r, s) decomposed in $\{\mathrm{C}\}$ (keep in mind that (r, s) should be represented in meters). It is more convenient to express (2) as a function of the NED coordinates instead of the camera-fixed coordinates since the origin of $\{\mathrm{C}\}$ moves with the UAV. It can be achieved by utilizing a transformation $\mathbf{G}_n^c$ between $\{\mathrm{C}\}$ and $\{\mathrm{N}\}$ [13]

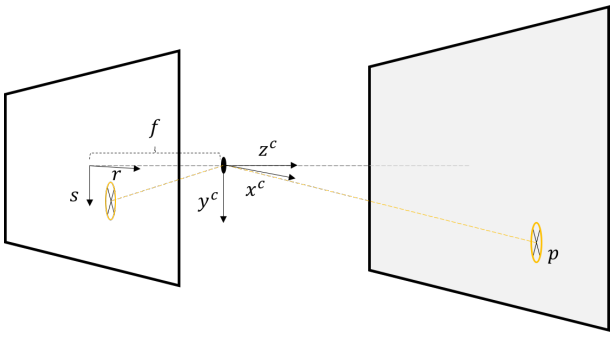

Fig. 1. Illustration of the pinhole camera model. The letter p marks the feature position in $\{\mathrm{N}\}$.

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

where $\underline{\mathbf{p}}^n$ is the homogeneous coordinates of the feature expressed in $\{\mathrm{N}\}$. $\mathbf{G}_n^c$ is defined as

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

where $\mathbf{R}_n^c$ is the rotation matrix between $\{\mathrm{C}\}$ and $\{\mathrm{N}\}$ defined in [13], with column vectors $\mathbf{r_1}$, $\mathbf{r_2}$ and $\mathbf{r_3}$, and $\mathbf{r}_{nc}^n$ is the the position of the origin of $\{\mathrm{C}\}$ relative to $\{\mathrm{N}\}$ decomposed in $\{\mathrm{N}\}$. By assuming that the origin of $\{\mathrm{C}\}$ coincides with the origin of $\{\mathrm{B}\}$, $\mathbf{r}_{nc}^n$ can be simplified as the NED positions of the UAV. In practice, for the experiment described in Section IV, the origin of $\{\mathrm{C}\}$ is located within centimeters of the origin of $\{\mathrm{B}\}$. Therefore, the assumption is reasonable, especially since the operating altitude of the UAV (in the experiment) is about 100 meters which makes an error of a few centimeters negligible.

Only two coordinates of the NED positions can be recovered by the pixel coordinates (r, s). Hence, the down position cannot be recovered directly. However, since we are observing objects at the sea surface, the down position of pixels in the image is close to zero as long as the origin of the NED frame is placed at sea level. It is necessary to assume that all objects in the image are located at sea level (unless a digital elevation map is available) and have a limited height compared to the altitude of the UAV. In practice, an object height of 10 meters did not degrade the results significantly when the UAV operated at an altitude of 100 meters, but the accuracy obviously decreases with the height of the object.

The NE coordinates of the pixel (r, s) are given by (3) as

$$\begin{bmatrix} N_{obj} \\ E_{obj} \\ 1 \end{bmatrix} = z^c \mathbf{G}_{NE}^{-1} \mathbf{A}^{-1} \begin{bmatrix} r \\ s \\ 1 \end{bmatrix} \tag{4}$$

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

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

The rotation matrix $\mathbf{R}_n^c$ depends on the roll (ϕ), pitch (θ) and yaw (ψ) angles as well as the gimbal pan (ψ_{gb}) and tilt (θ_{gb}) angles.

In order to find the NE coordinates, z^c needs to be calculated with an altitude measurement. When deriving the relationship between OF and velocity, it will also be necessary to calculate x^c and y^c, which are the x- and y coordinates of the feature at pixel (r, s) decomposed in $\{\mathrm{C}\}$. They can be expressed as a function of the UAV navigation states and gimbal orientation. Assume that the homogeneous coordinates of the feature can be written as $\underline{\mathbf{t}}^n = [x^n, y^n, z^n, 1]^\top$ and $\underline{\mathbf{t}}^c = [x^c, y^c, z^c, 1]^\top$ decomposed in $\{\mathrm{N}\}$ and $\{\mathrm{C}\}$, respectively. The relationship between the coordinates is

$$\underline{\mathbf{t}}^c = \mathbf{T}_n^c \underline{\mathbf{t}}^n \tag{5}$$

where $\mathbf{T}_n^c$ is the homogeneous transformation between $\{\mathrm{C}\}$ and $\{\mathrm{N}\}$ defined as

$$\mathbf{T}_n^c := \begin{bmatrix} \mathbf{R}_n^c & -\mathbf{R}_n^c \mathbf{r}_{nc}^n \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}$$

By inserting (5) into the pinhole camera model (1) the equation can be solved with respect to x^n and y^n (by assuming z^n known). The solution decomposed in $\{\mathrm{C}\}$ is calculated with (5) and given as

$$\mathbf{t}_*^c = \begin{bmatrix} x^c \\ y^c \\ z^c \end{bmatrix} = \frac{1}{\overline{r's\psi_{gb}s\theta + fc\theta_{gb}c\phi c\theta + s'c\psi_{gb}c\theta_{gb}s\theta -}} \\ \frac{}{\overline{fc\psi_{gb}s\theta_{gb}s\theta + r'c\psi_{gb}c\theta s\phi + s'c\phi c\theta s\theta_{gb} -}} \\ \frac{}{\overline{s'c\theta_{gb}c\theta s\psi_{gb}s\phi + fc\theta s\psi_{gb}s\theta_{gb}s\phi}} \begin{bmatrix} -r'(z_{uav}^n - z_f^n) \\ -s'(z_{uav}^n - z_f^n) \\ -f(z_{uav}^n - z_f^n) \end{bmatrix}$$

where r' and s' are the pixel coordinates and s and c are the sine and cosine functions. z_{uav}^n is the down position of the UAV and z_f^n is the down position of the feature at pixel (r, s) (assumed to be zero). $\mathbf{t}_*^c$ only depends on known parameters, and thus the camera-fixed coordinates of features are known as long as all features are located at sea level (z_f^n is zero), which is a reasonable assumption.

The georeferencing algorithm depends on measurements (or estimates) of the UAV NED positions, the Euler angles (roll, pitch and yaw), the gimbal orientation (pan and tilt angles), the focal length of the camera and the pixel position in the image plane. The accuracy of the obtained NED positions depends on the sensors used to measure or estimate these parameters. The NED positions of the UAV can be measured by GPS, but the down position is not very accurate with single frequency GPS receivers without differential correction. The Euler angles can be estimated with an inertial measurement unit (IMU) and some heading reference. The gimbal orientation cannot necessarily be measured, but a set-point should always be available. The focal length of the camera is given in the camera (lens) specification, but a more accurate estimate of the focal length is obtained with camera calibration. The pixel position is known from the feature extraction.

C. Transformation Between Optical Flow and Velocity

This section derives the relationship between OF and velocity. Assume that a feature at pixel position (r, s) is of interest. Differentiation of the pinhole camera model (1) yields

$$\begin{bmatrix} \dot{r} \\ \dot{s} \end{bmatrix} = \frac{1}{z_f^c} \begin{bmatrix} f & 0 & -f\frac{x_f^c}{z_f^c} \\ 0 & f & -f\frac{y_f^c}{z_f^c} \end{bmatrix} \begin{bmatrix} \dot{x}_f^c \\ \dot{y}_f^c \\ \dot{z}_f^c \end{bmatrix} \tag{6}$$

where $[\dot{r}, \dot{s}]^\top$ is the OF vector of the feature. The vector $[\dot{x}_f^c, \dot{y}_f^c, \dot{z}_f^c]^\top$ on the right-hand side is recognized as [15]

$$\dot{\mathbf{p}}_f^c = \begin{bmatrix} \dot{x}_f^c \\ \dot{y}_f^c \\ \dot{z}_f^c \end{bmatrix} = \mathbf{v}_{f/c}^c + \boldsymbol{\omega}_{f/c}^c \times (\mathbf{p}_f^c - \mathbf{o}_f^c) \tag{7}$$

where $\mathbf{v}_{f/c}^c$ and $\boldsymbol{\omega}_{f/c}^c$ are the linear and angular velocities of the feature with respect to $\{\mathrm{C}\}$ decomposed in $\{\mathrm{C}\}$, respectively. $\mathbf{p}_f^c = [x_f^c, y_f^c, z_f^c]^\top$ is the position of the feature decomposed in $\{\mathrm{C}\}$. $\mathbf{o}_f^c$ is the feature point of rotation decomposed in $\{\mathrm{C}\}$ such that $(\mathbf{p}_f^c - \mathbf{o}_f^c)$ is the arm of rotation. All rotations seen in the image are rotations about the camera center, hence the rotation point $\mathbf{o}_f^c$ coincides with the origin of $\{\mathrm{C}\}$. Since it is also assumed that the origin of $\{\mathrm{C}\}$ coincides with $\{\mathrm{B}\}$, the rotation of features caused by the UAV motion will be about the camera center. Thus, $\mathbf{o}_f^c$ is simply the zero vector.

The assumption of $\{\mathrm{C}\}$ coinciding with $\{\mathrm{B}\}$ has been tested experimentally. It was not possible to find an increase in accuracy when the distance between the origins was accounted for (when the true distance is limited to a few centimeters). Therefore, since the following derivation is simplified with the assumption, it is not accounted for in this paper. In situations where the origin of $\{\mathrm{C}\}$ is far from the origin of $\{\mathrm{B}\}$ one should be aware of the simplification.

Equation (6) might be rewritten by inserting (7):

$$\begin{bmatrix} \dot{r} \\ \dot{s} \end{bmatrix} = \frac{1}{z_f^c} \begin{bmatrix} \mathbf{B} & | & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{f/c}^c \\ \boldsymbol{\omega}_{f/c}^c \times \mathbf{p}_f^c \end{bmatrix} \tag{8}$$
$$\mathbf{B} = \begin{bmatrix} f & 0 & -f\frac{x_f^c}{z_f^c} \\ 0 & f & -f\frac{y_f^c}{z_f^c} \end{bmatrix}$$

By the properties of the crossproduct

$$\begin{aligned} \boldsymbol{\omega}_{f/c}^c \times \mathbf{p}_f^c &= -\mathbf{p}_f^c \times \boldsymbol{\omega}_{f/c}^c \\ &= -\mathbf{S}(\mathbf{p}_f^c)\boldsymbol{\omega}_{f/c}^c \end{aligned}$$

it is possible to rewrite (8) as

$$\begin{aligned} \begin{bmatrix} \dot{r} \\ \dot{s} \end{bmatrix} &= \frac{1}{z_f^c} \begin{bmatrix} \mathbf{B} & | & -\mathbf{B} \cdot \mathbf{S}(\mathbf{p}_f^c) \end{bmatrix} \begin{bmatrix} \mathbf{v}_{f/c}^c \\ \boldsymbol{\omega}_{f/c}^c \end{bmatrix} \\ &= \mathbf{M}(f, \mathbf{p}_f^c) \begin{bmatrix} \mathbf{v}_{f/c}^c \\ \boldsymbol{\omega}_{f/c}^c \end{bmatrix} \end{aligned} \tag{9}$$

A relationship between the linear and angular velocities and OF is now established through (9) where

$$\mathbf{M}(f,\mathbf{p}_f^c) = \frac{1}{z_f^c}\cdot \begin{bmatrix} f & 0 & -f\frac{x_f^c}{z_f^c} & -f\frac{x_f^c}{z_f^c}y_f^c & fz_f^c + f\frac{x_f^c}{z_f^c}x_f^c & -fy_f^c \\ 0 & f & -f\frac{y_f^c}{z_f^c} & -fz_f^c - f\frac{y_f^c}{z_f^c}y_f^c & f\frac{y_f^c}{z_f^c}x_f^c & fx_f^c \end{bmatrix}$$

Since the velocities of the UAV in the body-fixed frame is known from GPS velocity measurements and attitude estimates, it is possible to find the OF caused by the UAV motion. It will from now on be referred to as the theoretical flow. The theoretical flow $[\dot{r}_t, \dot{s}_t]^\top$ is defined as

$$\begin{bmatrix}\dot{r}_t \\ \dot{s}_t\end{bmatrix} := \mathbf{M}(f,\mathbf{p}_f^c)\begin{bmatrix}\mathbf{v}_{T/c}^c \\ \boldsymbol{\omega}_{T/c}^c\end{bmatrix} \tag{10}$$

where $\mathbf{v}_{T/c}^c$ and $\boldsymbol{\omega}_{T/c}^c$ are the linear and angular velocities of the sea surface (NED) with respect to $\{\mathrm{C}\}$ decomposed in $\{\mathrm{C}\}$. These velocities can be calculated with the navigation states of the UAV. Since the origin of $\{\mathrm{C}\}$ coincides with the origin of $\{\mathrm{B}\}$, both $\{\mathrm{B}\}$ and $\{\mathrm{C}\}$ have the same linear velocity with respect to $\{\mathrm{N}\}$. Therefore, it can be expressed as

$$\mathbf{v}_{T/c}^c = \mathbf{R}_b^c\mathbf{v}_{T/c}^b = \mathbf{R}_b^c\mathbf{v}_{T/b}^b = -\mathbf{R}_b^c\mathbf{v}_{b/T}^b$$

where $\mathbf{v}_{b/T}^b$ is the body-fixed linear velocity of the UAV with respect to $\{\mathrm{N}\}$ decomposed in $\{\mathrm{B}\}$, which is known from the navigation system.

$\boldsymbol{\omega}_{T/c}^c = \boldsymbol{\omega}_{n/c}^c$ can be expressed as

$$\begin{aligned}\boldsymbol{\omega}_{T/c}^c &= \boldsymbol{\omega}_{T/b}^c + \boldsymbol{\omega}_{b/c}^c \\ &= \mathbf{R}_b^c(\boldsymbol{\omega}_{T/b}^b + \boldsymbol{\omega}_{b/c}^b) \\ &= -\mathbf{R}_b^c(\boldsymbol{\omega}_{b/T}^b + \boldsymbol{\omega}_{c/b}^b)\end{aligned}$$

where $\boldsymbol{\omega}_{b/T}^b$ is the angular velocity of $\{\mathrm{B}\}$ with respect to $\{\mathrm{N}\}$ decomposed in $\{\mathrm{B}\}$. That is the angular velocity of the UAV which can be measured by gyros or be estimated in the navigation system. $\boldsymbol{\omega}_{c/b}^b$ is the angular velocity of $\{\mathrm{C}\}$ with respect to $\{\mathrm{B}\}$ decomposed in $\{\mathrm{B}\}$. It is given by the gimbal motion and should be accounted for. A pan/tilt gimbal can only rotate about the body z- and y-axis. Thus, the angular velocity caused by the gimbal can be approximated as [31]

$$\begin{aligned}\boldsymbol{\omega}_{c/b}^b &= \boldsymbol{\omega}_z(\dot{\psi}_{gb}) + \mathbf{R}_z(\psi_{gb})\boldsymbol{\omega}_y(\dot{\theta}_{gb}) \\ &= \begin{bmatrix}0\\0\\\dot{\psi}_{gb}\end{bmatrix} + \mathbf{R}_z(\psi_{gb})\begin{bmatrix}0\\\dot{\theta}_{gb}\\0\end{bmatrix}\end{aligned}$$

where $\mathbf{R}_z(\psi_{gb})$ is a principle rotation about the z-axis with a rotation given by the pan angle ψ_{gb} [26]. $\dot{\psi}_{gb}$ and $\dot{\theta}_{gb}$ are the derivatives of the pan and tilt angles, respectively. They need to be measured or approximated by e.g. a first-order Taylor-series approximation (or a higher-order method). A first-order Taylor-series approximation is utilized in the experiment described in Section IV.

The theoretical flow can now be calculated with (10). It is still, however, some work needed before the velocity of the feature itself is identified. The OF is a sum of the camera motion with respect to NED and the motion of the features in the images with respect to NED. Moreover,

$$\begin{bmatrix}\mathbf{v}_{f/c}^c \\ \boldsymbol{\omega}_{f/c}^c\end{bmatrix} = \begin{bmatrix}\mathbf{v}_{f/T}^c + \mathbf{v}_{T/c}^c \\ \boldsymbol{\omega}_{f/T}^c + \boldsymbol{\omega}_{T/c}^c\end{bmatrix} \tag{11}$$

Equation (11) can be inserted into (9) where $\dot{r}_m$ and $\dot{s}_m$ now are defined as the measured OF obtained with e.g. SIFT. Consequently,

$$\begin{bmatrix}\dot{r}_m \\ \dot{s}_m\end{bmatrix} = \mathbf{M}(f,\mathbf{p}_f^c)\begin{bmatrix}\mathbf{v}_{f/T}^c \\ \boldsymbol{\omega}_{f/T}^c\end{bmatrix} + \mathbf{M}(f,\mathbf{p}_f^c)\begin{bmatrix}\mathbf{v}_{T/c}^c \\ \boldsymbol{\omega}_{T/c}^c\end{bmatrix} \tag{12}$$

where the second term is recognized as the theoretical flow. Thus, it is possible to rewrite (12) as

$$\begin{aligned}\begin{bmatrix}\dot{r}_m \\ \dot{s}_m\end{bmatrix} &= \mathbf{M}(f,\mathbf{p}_f^c)\begin{bmatrix}\mathbf{v}_{f/T}^c \\ \boldsymbol{\omega}_{f/T}^c\end{bmatrix} + \begin{bmatrix}\dot{r}_t \\ \dot{s}_t\end{bmatrix} \\ \begin{bmatrix}\dot{r}_m - \dot{r}_t \\ \dot{s}_m - \dot{s}_t\end{bmatrix} &= \mathbf{M}(f,\mathbf{p}_f^c)\begin{bmatrix}\mathbf{R}_n^c\mathbf{v}_{f/T}^n \\ \mathbf{R}_n^c\boldsymbol{\omega}_{f/T}^n\end{bmatrix}\end{aligned} \tag{13}$$

Since the main motivation is to locate surface objects in the NE plane, the angular velocity of the features located on the object is assumed to be zero (constant object heading between successive images). Therefore, the final three columns of $\mathbf{M}(f,\mathbf{p}_f^c)$ disappears and (13) can be further simplified. Consequently,

$$\begin{bmatrix}\dot{r}_m - \dot{r}_t \\ \dot{s}_m - \dot{s}_t\end{bmatrix} = \frac{1}{z_f^c}\begin{bmatrix} f & 0 & -f\frac{x_f^c}{z_f^c} \\ 0 & f & -f\frac{y_f^c}{z_f^c}\end{bmatrix}\mathbf{R}_n^c\mathbf{v}_{f/T}^n \tag{14}$$

In addition, since the down velocity is zero, the last column of $\mathbf{R}_n^c$ can be discarded in (14) and the NE velocities (v_x^n and v_y^n) of the feature can be calculated as

$$\begin{aligned}\begin{bmatrix}\dot{r}_m - \dot{r}_t \\ \dot{s}_m - \dot{s}_t\end{bmatrix} &= \frac{1}{z_f^c}\begin{bmatrix} f & 0 & -f\frac{x_f^c}{z_f^c} \\ 0 & f & -f\frac{y_f^c}{z_f^c}\end{bmatrix}\begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2\end{bmatrix}\begin{bmatrix}v_x^n \\ v_y^n\end{bmatrix} \\ \begin{bmatrix}v_x^n \\ v_y^n\end{bmatrix} &= \left(\frac{1}{z_f^c}\begin{bmatrix} f & 0 & -f\frac{x_f^c}{z_f^c} \\ 0 & f & -f\frac{y_f^c}{z_f^c}\end{bmatrix}\begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2\end{bmatrix}\right)^{-1}\begin{bmatrix}\dot{r}_m - \dot{r}_t \\ \dot{s}_m - \dot{s}_t\end{bmatrix}\end{aligned} \tag{15}$$

The NE velocities of the feature are now identified. By combining all features related to an object, the mean velocity can be used as a measurement in a target tracking system. The velocity calculation depends on measurements of the parameters described at the end of Section III-B (for georeferencing). Additionally, the body-fixed linear and angular velocities of the UAV and an approximation of the pan and tilt angle derivatives need to be obtained. The linear velocity can be measured by GPS and the angular velocity can be measured by gyros. The pan and tilt angles can be approximated by a first-order Taylor-series approximation as described earlier.

D. Target Tracking System

A tracking system utilizing a Kalman filter [25] is used to track surface objects. For simplicity it is assumed that a single target is of interest and that there exist a method for identifying every OF vector related to the target, e.g. by using the detection method described in [12]. Multiple target tracking can be accomplished by the association method described in [12].

A motion model for the target is required in order to use a Kalman filter. How to choose a motion model is described in [32]. In this paper, a constant velocity model (white noise acceleration) is chosen. This is because the dynamics of typical surface objects are assumed to be slow. The position and velocity in the NE plane are of interest. The constant velocity motion model can be described in continuous time as

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{E}\mathbf{w} \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \mathbf{v} \end{aligned} \tag{16}$$

where $\mathbf{x} = [p_x^n, p_y^n, v_x^n, v_y^n]^\top$, $\mathbf{w} = [w_x, w_y]^\top$, $\mathbf{y} = [p_x^n, p_y^n, v_x^n, v_y^n]^\top$ and $\mathbf{v} = [v_{px}, v_{py}, v_{vx}, v_{vy}]^\top$. The state vector $\mathbf{x}$ consists of the position and velocity in the NE plane, the vector $\mathbf{w}$ is white Gaussian process noise, $\mathbf{y}$ is the measurement vector (obtained by georeferencing and OF) and $\mathbf{v}$ is white Gaussian measurement noise. The matrices $\mathbf{A}$, $\mathbf{E}$ and $\mathbf{C}$ are

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{E} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The motion model (16) can be discretized in order to get the discrete state-space equations, which are necessary for implementation in a computer.

The tracking system receives a measurement of the position and velocity of the target in the NE-plane. This is achieved by finding the mean position and velocity of every feature related to the object. Obviously, it is also possible to treat each feature independently and feed the Kalman filter with more than one measurement for each element of the state vector, but this is not utilized in the experiment. Since the mean position of the features is used as a position measurement, it might be somewhat inaccurate for large targets because features not necessarily are uniformly distributed on the target. Furthermore, this is especially an issue if only a part of the target is visible in the image. Increased accuracy on the position measurements is achieved if the center of the target is obtained, with e.g. [12].

IV. Experimental Setup

A flight experiment consisting of two different flights has been conducted in the Azores outside of Portugal, the summer of 2015. The X8 Skywalker fixed-wing UAV interfaced with a retractable pan/tilt gimbal was used to gather data. A light-weight payload [13] with a FLIR Tau2 640 thermal camera with a focal length of 19mm and resolution of 640×480 pixels was used to capture images from the flights. The experiment was conducted at sea with marine vessels operating in the area. An image captured at the experiment is displayed in Figure 2. The thermal camera has a frame rate of 7.5 frames per second, and has been calibrated with the method proposed in [13] to increase the accuracy of the camera intrinsic matrix. This is important because the accuracy of the georeferencing algorithm and the velocity calculation from OF rely heavily on the conversion between pixels and meters.

Fig. 2. A thermal image captured at the flight experiment in the Azores, the summer of 2015.

The images and the navigation data gathered at the experiment have been processed off-line. The Open Source Computer Vision Library (OpenCV) [33] has been used to implement SIFT, which is used for feature extraction and OF calculation. Matched features between successive images are assigned a value indicating the uncertainty of the match. Matches with more than twice the uncertainty of the best match have been removed to increase the reliability of the OF vectors. Figure 3 displays a single image captured in the experiment with OF vectors acquired by SIFT.

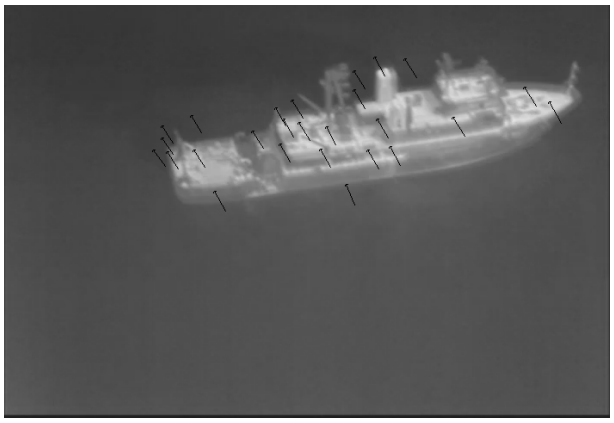

Fig. 3. Optical flow vectors acquired by SIFT on an image captured at the flight experiment.

The OF vectors computed with SIFT have been combined with navigation data from the UAV (estimated by an extended Kalman filter [25]) to calculate the velocity

of detected features. The navigation data are stored with a frequency of 10 Hz. The gimbal pan and tilt angles were both controlled manually and automatically with a gimbal controller [23] during the flight. The pan and tilt angles cannot be measured directly and only the commanded set-point is available. Therefore, a possible weakness in the results is the accuracy of the pan and tilt angles. Another limitation is the delay between a change in set-point and when the gimbal actually reaches the new orientation. However, the delay has been compensated for by estimating the delay at two points in time. This is achieved by identifying occasions where the pan angle changes significantly and locating the corresponding image where the change can be observed. Nevertheless, the delay might differ at other points in time and the accuracy of the estimated delay is limited by the frame rate of the camera.

The relationship between OF and velocity requires time-synchronized data. The images, navigation data and gimbal orientation are stored by the on-board computer. The data are not synchronized in hardware, and thus the time stamp is given by the on-board computer. In practice this means that the time stamps can be somewhat uncertain when the on-board computer has a lot of tasks. This is because a delay will be added to when a sensor obtained the measurement. Moreover, the GPS receiver has a typical delay of 100-200 ms [34]. However, the images have been synchronized off-line by adjusting the time stamp for images where the time between subsequent images differs substantially from the frame rate. Furthermore, the mean time between consecutive images (given by the time stamps without any adjustment) was in accordance with the frame rate of the camera.

The tracking system is discretized and implemented in Matlab. Prediction is performed for every received image. Measurements are used to correct the prediction whenever the target is detected in the images and OF vectors are available. The overall goal of the target tracking system is to be able to predict the trajectory of the target when measurements are unavailable. Therefore, the experiment contains periods where the target is outside of the field of view of the camera.

In order to use a Kalman filter it is necessary to choose covariance matrices for the process and measurement noise ($\mathbf{Q}$ and $\mathbf{R}$). The covariance matrix of the process noise was for simplicity chosen to be diagonal and constant:

$$\mathbf{Q} = E[\mathbf{w}\mathbf{w}^\top] = \begin{bmatrix} a_{max} & 0 \\ 0 & a_{max} \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

$\mathbf{Q}$ was designed under the assumption that marine vessels have slow dynamics. A maximum acceleration of $a_{max} = 3m/s^2$ in both North and East was assessed to be reasonable, although smaller accelerations are expected in practice. This is because the noise also needs to account for the uncertainty of the assumed model since the true model is unknown.

$\mathbf{R} = E[\mathbf{v}\mathbf{v}^\top]$ was designed with respect to the expected accuracy of the georeferencing algorithm and the velocity calculation. The standard deviation in position was approximately $10m$ in [13] for both the north and east position, but that was obtained with a fixed gimbal position. Therefore, the performance of the georeferencing is likely to be less accurate with gimbal motion. Since gimbal motion affects the georeferencing and velocity calculation significantly, two different noise levels were created. R_l is used as the measurement covariance at every time step where the change in pan and tilt angles (set-point) between consecutive images are below $0,7$ degrees. R_h is used when a change in pan or tilt is above the same threshold. The maximum and mean changes in tilt are $1,92$ and $0,3$ degrees (between consecutive images) during the time the target is visible in the images. The pan angle is constant in the tracking period. The matrices are designed as

$$\mathbf{R}_l = \begin{bmatrix} 150 & 0 & 0 & 0 \\ 0 & 150 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}, \mathbf{R}_h = \begin{bmatrix} 250 & 0 & 0 & 0 \\ 0 & 250 & 0 & 0 \\ 0 & 0 & 50 & 0 \\ 0 & 0 & 0 & 50 \end{bmatrix}$$

The tracking system was initialized with the first available measurement for position and velocity. The initial value of the covariance matrix was chosen to be

$$\mathbf{P}_{init} = \begin{bmatrix} 300 & 0 & 0 & 0 \\ 0 & 300 & 0 & 0 \\ 0 & 0 & 150 & 0 \\ 0 & 0 & 0 & 150 \end{bmatrix}$$

to ensure that the tracking system is not overconfident in the beginning of the tracking period.

The tracking system is able to run in real-time on a MacBook Pro (2015 version) with an Intel dual core i7 processor when images arrive at a frequency of 7.5 Hz. A non-optimized implementation of SIFT in OpenCV is used, and it can process more than 13 images each second. The processing frequency can be increased by decreasing the image resolution.

V. Results

This section presents the results from the off-line processing of the data gathered at the flight experiment. The results are divided into two sections because two different simulations have been conducted (one from each flight). The first part looks into tracking of a large marine vessel at rest (displayed in Figure 2). The second part looks into tracking of a small maneuvering vessel (displayed in Figure 12). The goal is to estimate the position and velocity of the vessels. The tests are based on two different sets of images and the corresponding navigation data.

A. Tracking of Marine Vessel at Rest

The first test is based on images with a large marine vessel, which is located in the camera field of view for a short period on two separate occasions. The main challenge is to predict the trajectories of the ship when measurements are unavailable. The vessel is approximately 70 meters long and has a width of 13 meters. An image captured of the vessel is displayed in Figure 2. Figure 4 shows the UAV path estimated by the navigation system and the path of the vessel (measured by GPS) for a period of 80 seconds, which

is the tracking period. Figure 5 shows the gimbal pan and tilt angles in the same time span. The vessel is at rest during the tracking period, but the system has no knowledge of the behavior of the vessel beforehand.

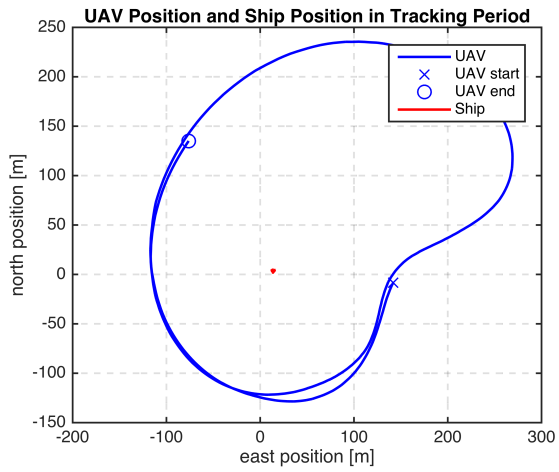

Fig. 4. Position in the NE plane for the UAV and the ship during the tracking period in the first test.

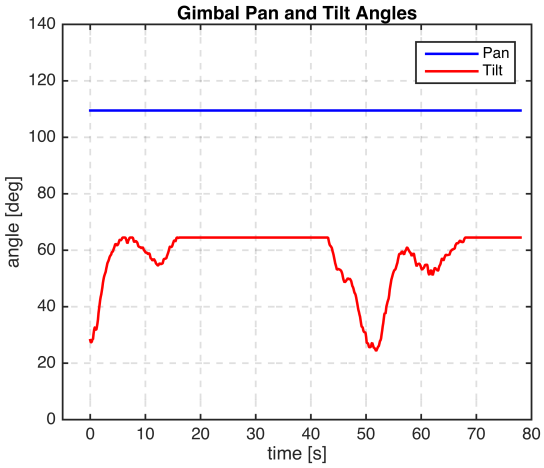

Fig. 5. Gimbal orientation during the tracking period in the first test.

The vessel is not in the camera field of view in the the time interval between 20 and 50 seconds and after 72 seconds. Furthermore, SIFT is not able to find features on the ship in some images. 600 images were captured in the tracking period and features were detected on the vessel in 250 images. A part of the vessel is visible in approximately 400 images. However, 100 of these images only contain a very small part of the vessel. The whole vessel is visible in 200 images.

Figure 6 and 7 show the theoretical flow and the OF measured by SIFT in horizontal (r) and vertical (s) direction in the image plane. Since the vessel is at rest, the theoretical flow should be equal to the measured OF. The noise level is fairly large in s and the accuracy is better in r. Still, taking the uncertainty related to synchronization of data and the accuracy of the sensors into account, the results look reasonable.

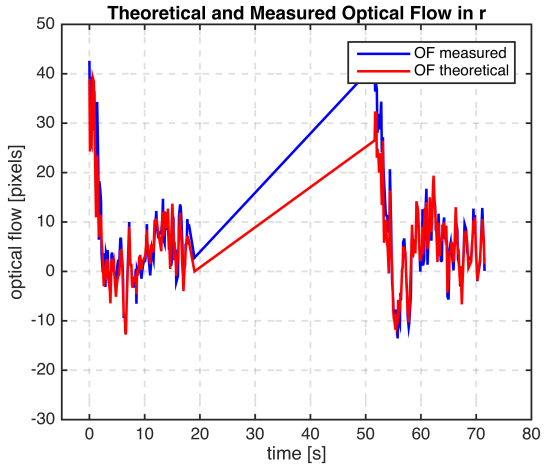

Fig. 6. Comparison of theoretical and measured optical flow in the horizontal direction (r). They are, in theory, equal for objects at rest.

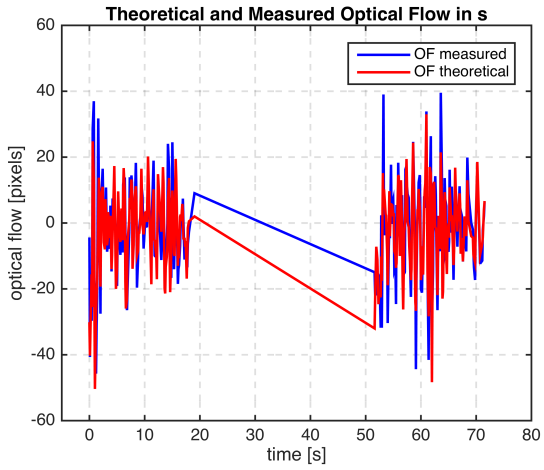

Fig. 7. Comparison of theoretical and measured optical flow in vertical direction (s). They are, in theory, equal for objects at rest.

Figure 8 shows the measured position of the vessel (obtained by georeferencing the images) together with the UAV position. The measured position does not vary significantly, which is the expected behavior for a target at rest. Moreover, it is important to notice that the measured position is little affected by the UAV position (also attitude and gimbal orientation), which shows that the georeferencing is reliable. Figure 9 shows the velocity of the ship obtained by OF. The noise level is quite large, but the mean error is within $1m/s$ in both the North and East velocities.

Figure 10 and 11 display the estimates of the ship position and velocity from the tracking system. The position estimate is compared with the GPS measured position as a reference. The GPS is located close to the center of the ship. The estimated north and east position are close to the GPS measured position after an initial transient period. More

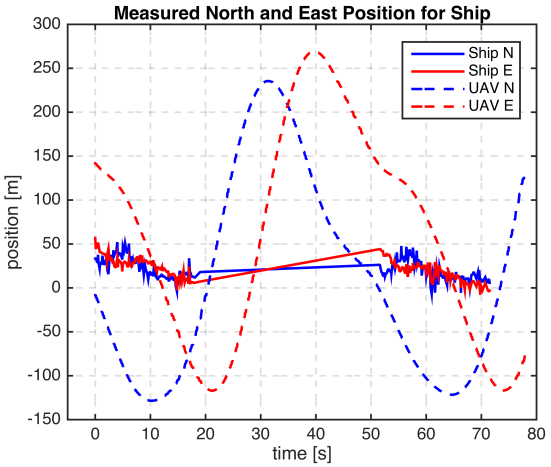

Fig. 8. Measured position of ship obtained with georeferencing together with UAV position. The ship was not in the field of view of the camera in the time interval $[20, 50]$ and $[72, 80]$.

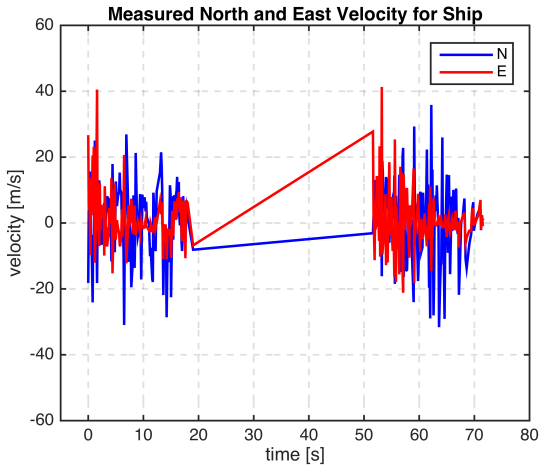

Fig. 9. Measured velocity of ship obtained with optical flow. The ship was not in the field of view of the camera in the time interval $[20, 50]$ and $[72, 80]$.

importantly, the predicted position does not drift significantly in the period between 20 and 50 seconds when measurements are unavailable. Furthermore, the estimated position is within 15 meters of the GPS position in the end. The estimated velocity is also quite accurate. It is important to notice that the tracking system almost manages to converge to the reference (zero) just before the target is outside the field of view. This is obviously why the predicted position does not drift far away from the reference in the period where the target is outside the field of view. With this in mind, the tracking system works as intended. Nevertheless, it is important to also be aware of the fact that the performance strongly depends on the navigation states of the UAV, and thus it is advantageous with accurately calibrated sensors and time-synchronized data.

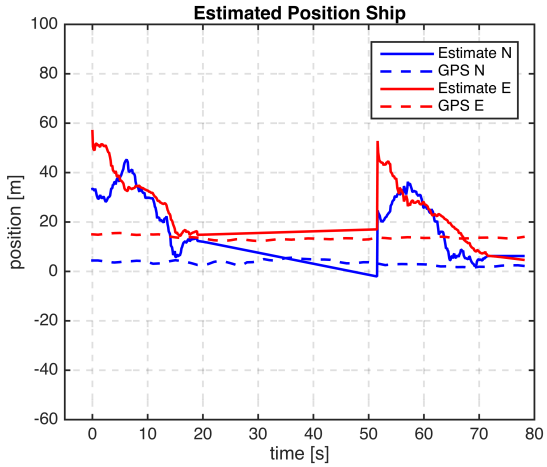

Fig. 10. Estimated position of ship compared with position measured by GPS in the first test.

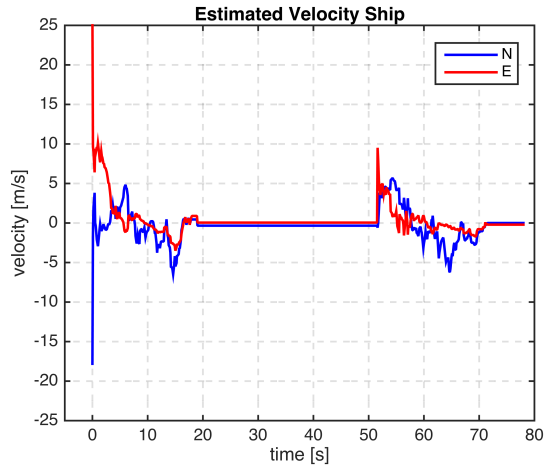

Fig. 11. Estimated velocity of ship in the first test. A reference is not available, but the ship was at rest during the tracking period and the velocity expected to be close to zero.

B. Tracking of Maneuvering Marine Vessel

The second test is based on images with a smaller marine vessel. The vessel is displayed in Figure 12. Figure 13 shows the UAV path estimated by the navigation system together with the path of the vessel (measured by GPS) for approximately 55 seconds, which is the tracking period in the second test. The vessel is only in the field of view of the camera in the time interval between 0 and 5 seconds and from 37 to 48 seconds. Thus, the estimates are in a very large part of the tracking period solely based on prediction.

420 images were captured in the tracking period and the vessel was visible in 120 images. Features were detected on the vessel in 99 images. Images where the absolute value of the calculated velocity was above $15m/s$ in North or East were discarded because these velocities are assumed to exceed the physical constraints of the vessel. 13 images were removed because the measured velocity surpassed the threshold. Therefore, 86 measurements were available in

the tracking period. 30 of these measurements are located in the time between 0 and 5 seconds. The remaining 56 measurements are located in the time interval between 37 and 48 seconds.

Fig. 12. A thermal image captured of the small vessel used in the second test.

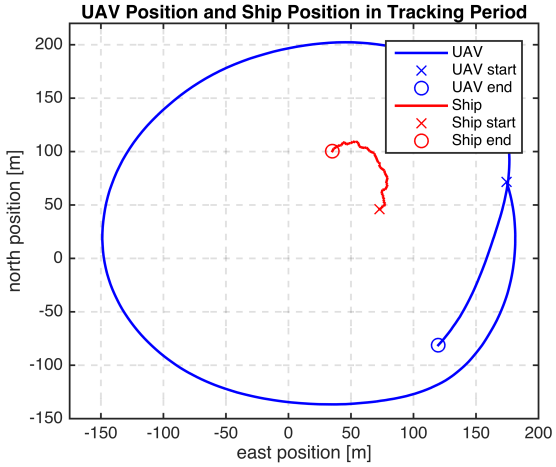

Fig. 13. Position in the NE plane for the UAV and the vessel during the tracking period in the second test.

Figure 14 and 15 display the estimated position and speed of the vessel. The estimates are compared with GPS measurements of position and speed. The estimated position is close to the reference during the whole tracking period. Obviously the estimates are more accurate in the time intervals when measurements are available. Nevertheless, the predicted position is quite reasonable in both North and East when measurements are unavailable, especially since the target operates outside the field of view of the camera for 30 seconds without new observations. The estimated speed is also quite accurate. It is slightly above the GPS measured speed. A different tuning of the tracking system could have increased the accuracy of the results. The same tuning is chosen for both tests because it shows that the same parameters can be used to track both a moving target and a target at rest. This is obviously an advantage since the system is more robust when different objects can be tracked with the same parameters. However, the chosen tuning might not be ideal for either of the tests.

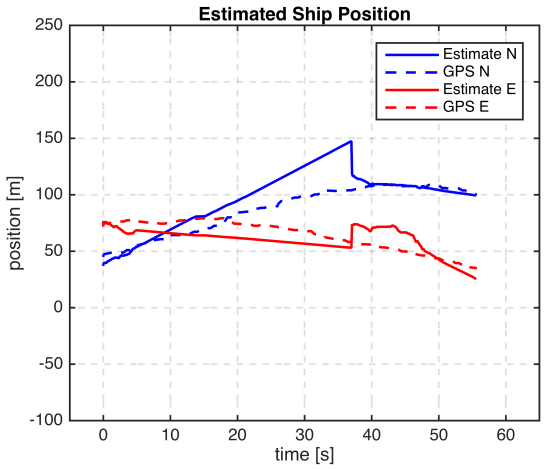

Fig. 14. Estimated position of the vessel compared with the position measured by GPS in the second test.

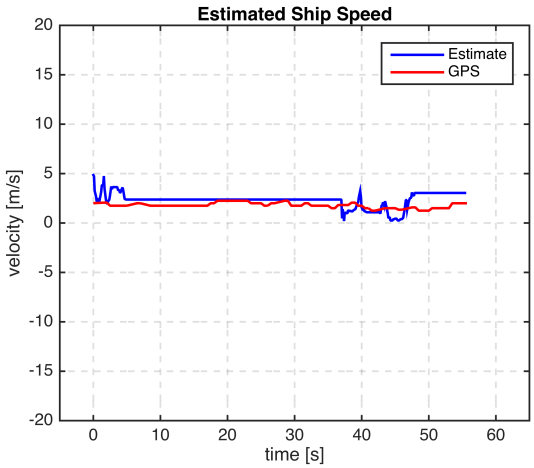

Fig. 15. Estimated velocity of the vessel compared with the GPS measured speed in the second test.

VI. CONCLUSIONS

This paper presented a vision-based tracking system for marine surface objects utilizing images captured in a fixed-wing unmanned aerial vehicle with a retractable pan/tilt gimbal and a thermal camera. A method for calculating the NED positions and velocities of objects, based on the pixel position and optical flow, has been derived. This is challenging because it is necessary to compensate for the fast UAV and gimbal motions, and the thermal images have low resolution and frame rate. Experimental results show that it is possible to estimate the position and velocity of a ship with thermal images captured from a fixed-wing UAV operating at high speed. More importantly, the system is able to predict the position trajectory of the ship quite well when the ship is outside the field of view of the camera.

ACKNOWLEDGMENT

This work has been carried out at the NTNU Centre for Autonomous Marine Operations and Systems. This work was supported by the Research Council of Norway through the Centres of Excellence funding scheme, Project number 223254. The authors are grateful for the support from the University of Porto, the University of the Azores, UAV operators Lars Semb and Krzysztof Cisek and REP15 coordinators João Tasso Sousa and Kanna Rajan.

REFERENCES

[1] P. Doherty and P. Rudol, "A uav search and rescue scenario with human body detection and geolocalization," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4830, pp. 1–13, 2007.

[2] P. Rudol and P. Doherty, "Human body detection and geolocalization for uav search and rescue missions using color and thermal imagery," in *IEEE Aerospace Conference*, 2008.

[3] 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 *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015.

[4] J. Hosen, H. H. Helgesen, L. Fusini, T. I. Fossen, and T. A. Johansen, "A vision-aided nonlinear observer for fixed-wing uav navigation," in *AIAA Guidance, Navigation, and Control Conference*, 2016.

[5] Y.-C. Chung and Z. He, "Low-complexity and reliable moving objects detection and tracking for aerial video surveillance with small uavs," in *IEEE International Symposium on Circuits and Systems*, 2007.

[6] "COLREGs - convention on the international regulations for preventing collisions at sea, international maritime organization (IMO)," 1972.

[7] L. Elkins, D. Sellers, and W. R. Monach, "The autonomous maritime navigation (amn) project: Field tests, autonomous and cooperative behaviors, data fusion, sensors, and vehicles," *Journal of Field Robotics*, vol. 27, no. 6, pp. 790–818, 2010.

[8] M. T. Wolf, C. Assad, Y. Kuwata, A. Howard, H. Aghazarian, D. Zhu, T. Lu, A. Trebi-Ollennu, and T. Huntsberger, "360-degree visual detection and target tracking on an autonomous surface vehicle," *Journal of Field Robotics*, vol. 27, no. 6, pp. 819–833, 2010.

[9] T. Huntsberger, H. Aghazarian, A. Howard, and D. C. Trotz, "Stereo vision–based navigation for autonomous surface vessels," *Journal of Field Robotics*, vol. 28, no. 1, pp. 3–18, 2011.

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

[11] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Surveys*, vol. 38, no. 4, 2006.

[12] 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 *IEEE Aerospace Conference, Big Sky, US*, 2015.

[13] F. S. Leira, K. Trnka, T. I. Fossen, and T. A. Johansen, "A ligth-weight thermal camera payload with georeferencing capabilities for small fixed-wing uavs," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015.

[14] B.K.P.Horn and B.G.Schunk, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–204, 1981.

[15] M. Mammarella, G. Campa, M. L. Fravolini, and M. R. Napolitano, "Comparing optical flow algorithms using 6-dof motion of real-world rigid objects," *IEEE Transactions on systems, Man, and Cybernetics*, vol. 42, no. 6, pp. 1752–1762, 2012.

[16] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2749, pp. 363–370, 2003.

[17] B.Lucas and T.Kanade, "An iterative image restoration technique with an application to stereo vision," *Proc. DARPA Image Underst, Workshop*, pp. 121–130, 1981.

[18] D. Lowe, "Object recognition from local scale-invariant features," *Proc. Int. Conf. Computer Vision*, pp. 1150–1157, 1999.

[19] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[20] J. Shi and C. Tomasi, "Good features to track," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 593–600, 1994.

[21] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[22] H. Chao, Y. Gu, and M. Napolitano, "A survey of optical flow techniques for robotics navigation applications," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 73, no. 1-4, pp. 361–372, 2014.

[23] E. Skjong, S. Nundal, F. Leira, and T. 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 *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015.

[24] H. H. Helgesen, "Object detection and tracking based on optical flow in unmanned aerial vehicles," master thesis, NTNU, Trondheim, 2015.

[25] R. Brown and P. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, 4th ed. John Wiley and Sons, Inc., 2012.

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

[27] A. Jain, M. Murty, and P. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[28] Y. Bar-Shalom, T. Kirubarajan, and X. Lin, "Probabilistic data association techniques for target tracking with applications to sonar, radar and eo sensors," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8 II, pp. 37–54, 2005.

[29] M. Muja and D. G. Lowe, "Flann, fast library for approximate nearest neighbors," in *International Conference on Computer Vision Theory and Applications (VISAPP'09)*, 2009.

[30] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, 1996.

[31] O. Egeland and J. T. Gravdahl, *Modeling and simulation for automatic control*. Marine Cybernetics Trondheim, Norway, 2002.

[32] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking: dynamic models," in *Proceedings of SPIE*, vol. 4048, 2000.

[33] G. Bradski, *Dr. Dobb's Journal of Software Tools*, 2000.

[34] J. Hansen, T. Fossen, and T. Johansen, "Nonlinear observer for ins aided by time-delayed gnss measurements: Implementation and uav experiments," in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015.