Air-to-Ground Target Tracking in a GPS-Denied Environment using Optical Flow Estimation

Yoko Watanabe* and Patrick Fabiani†
Department of Systems Control and Flight Dynamics
ONERA, Toulouse 31055, France

Guy Le Besnerais‡
Department of Modeling and Information Processing
ONERA, Châtillon 92322, France

This paper proposes a visual navigation system for an unmanned aerial vehicle to track a moving ground object in a GPS-denied environment. Image processing combines a target tracker which provides pixel coordinates of a ground target, and the estimation of optical flow around the detected target position. An extended Kalman filter is applied to estimate global position and velocity of the target as well as those of the own-ship aerial vehicle by fusing the image processing outputs with onboard inertial sensor measurements. The image processor and the estimation filter are tested on onboard camera images and vehicle state data that are synchronically recorded in actual flights.

I. INTRODUCTION

Unmanned aerial vehicles, or UAVs, are expected to play an important role in both military and commercial operations: for example, reconnaissance and surveillance mission, disaster observation, and search and rescue operation. Towards a need of UAVs performing such a complex mission in an unknown remote site, an autonomous flight system of UAVs has been progressively developed in recent years. In particular, vision-based navigation, guidance and control problem is one of the most focused on research topics in automation of UAVs. This paper addresses a visual air-to-ground target tracking problem in an urban site which may possibly deny an access to GPS signals.

As seen in nature among birds and flying insects, a 2D passive vision sensor can be used as an exclusive information source for a navigation purpose in a 3D environment. In fact, there is a large body of research on bio-mimic approaches of vision-based UAV navigation. The most well-known study in this research community is by Srinivasan et al.,[1] who have analyzed how honeybees utilize vision information to navigate themselves. Through experiments, they discovered that honeybees use cues derived from *optical flow* for flight control and odometory. Inspired by this study, optical flow-based obstacle avoidance algorithms for UAVs have been suggested.[2,3] Their results demonstrated the effectiveness of optical flow information in UAV navigation. However, those bio-mimic approaches are applicable only to local navigation but not to global navigation.

Our work on visual target tracking is a part of a research project which aims to achieve a multi-agent cooperative operation including UAVs, unmanned ground vehicles (UGVs) and software agents. When operating in a group of agents, it is important to localize objects and agents themselves with respect to a commonly defined world reference in order to share information. Therefore, a goal of this paper is to globally localize a ground target as well as an own-ship aerial vehicle by using vision when GPS signals are occasionally disrupted. Vision-aided self-localization in a GPS-denied environment has been investigated

*Post-Doctoral Researcher, Email: Yoko.Watanabe@onera.fr
†Director, Email: Patrick.Fabiani@onera.fr
‡Research Engineer, Email: Guy.Le_Besnerais@onera.fr

by others. Many of the works are based on an idea of simultaneous localization and mapping (SLAM) problem that was firstly addressed by Smith and Chesseman[4] and is well-described by Durrant-Whyte and Bailey.[5] Visual SLAM problem with UAV applications have been addressed in some studies.[6,7] In their work, it is suggested to integrate vision-based measurements (optical flow and range) with inertial navigation system (INS) in order to simultaneously localize an UAV and feature points (or landmarks) that are selected through image processing. One of the drawbacks of this approach is that it assumes a fixed environment and cannot handle an environment containing moving objects. In addition, the computational load of estimation increases as the number of feature points increases. Similar to their work, this paper adopts an integrated INS/vision navigation system but for simultaneous self-localization and localization of a moving ground target. First, a visual target tracker detects the 2D position of a target on an image. Then, feature points are sampled and a optical flow vector is calculated for each feature point. This optical flow computation is performed only on a small part of the image, selected in a proximity of the detected target position so that all the sampled feature points can be assumed to lie on a ground surface. Finally, an extended Kalman filter (EKF) is designed to estimate global position and velocity of the target and the UAV from those image processing results (target position and optical flow) and onboard inertial sensor measurements. The suggested EKF design utilizes an affine approximation of an optical flow field instead of an actual measurement set of optical flow vectors. In this way, we can eliminate redundant information and avoid the large number of measurements.

This paper is organized in the following manner: Section II defines a visual air-to-ground target tracking problem. Section III describes image processing algorithms including a visual target tracker and optical flow estimation. Section IV presents an EKF-based navigation filter design. Section V shows simulation results, and Section VI includes concluding remarks.

II. PROBLEM FORMULATION

This paper considers a visual air-to-ground target tracking problem in an urban environment, where an access to GPS signals can be denied. An UAV is required to pursue a moving ground target while globally localizing it by using a single onboard camera and inertial sensors. Figure 1 depicts an entire flight system of an UAV for this problem. This paper focuses on image processing and navigation filter design, and assumes that a guidance and control system for tracking is available.

A. Reference Frames

Before deriving mathematical formulations of the problem, the following reference frames are defined to express state variables. Figure 2 graphically summarizes these frames.

- **Local Reference Frame (F_I) :** A frame locally fixed to a world. Absolute states of the UAV and the target are expressed in this frame. A NED (North-East-Down) frame fixed on the Earth surface is used in this paper. Neglecting rotation and curvature of the Earth, the NED frame can be considered as an inertial frame.
- **UAV Body Frame (F_B) :** A frame fixed to a UAV body. Its origin is at the center of gravity of the body. The X-axis and the Z-axis are pointing forward and downward, respectively.
- **Camera Frame (F_C) :** A frame fixed to the camera. Its origin is at the camera's focus, and its Z-axis alignes with the camera's optical axis. The X-axis and the Y-axis are parallel to the 2D image frame axes defined below.
- **Image Frame (F_i) :** A 2D frame defined on the image plane. Its x-axis and y-axis are taken in the direction of image width and length, respectively.

B. Dynamics of UAV, Camera and Target

Let $\boldsymbol{X}_v$, $\boldsymbol{V}_v$ be the position and velocity vectors of the UAV and $\boldsymbol{a}_v$ be its acceleration input, expressed in the inertial frame F_I. Then, the UAV dynamics are given by

$$\dot{\boldsymbol{X}}_v = \boldsymbol{V}_v, \quad \dot{\boldsymbol{V}}_v = \boldsymbol{a}_v. \tag{1}$$

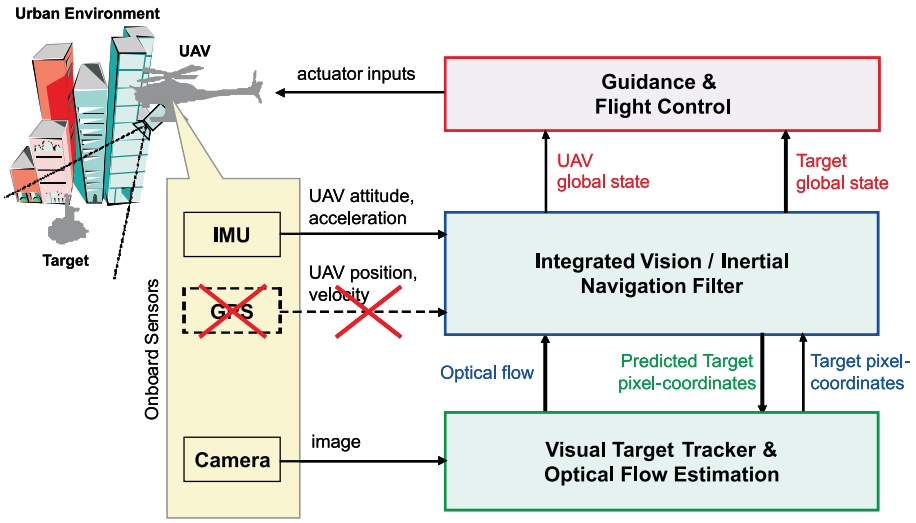


Figure 1. Visual Target Tracking System

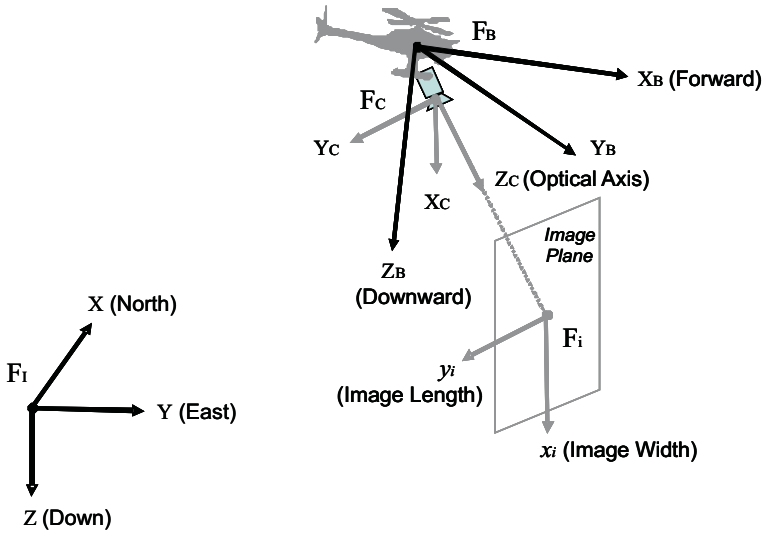


Figure 2. Reference Frames

The UAV attitude is determined by roll, pitch and yaw angles ($\phi,\ \theta,\ \psi$). A direction cosine matrix (DCM) from F_I to F_B, denoted by L_{BI}, is calculated from those attitude angles as follows.

$$L_{BI} = L_1(\phi)L_2(\theta)L_3(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where $L_i(\cdot)$ represents a rotation around the i-th axis. The camera is fixed to the UAV body with a displacement $\Delta\boldsymbol{X}_c$ and mount angles ($\phi_c,\ \theta_c,\ \psi_c$) with respect to F_B. Then, an absolute camera position $\boldsymbol{X}_c$ and a DCM from F_I to F_C are given as follows.

$$\boldsymbol{X}_c = \boldsymbol{X}_v + L_{BI}^T \Delta\boldsymbol{X}_c \tag{3}$$

$$L_{CI} = L_{CB}L_{BI} \tag{4}$$

where L_{CB} is a DCM from F_B to F_C that is determined by the mount angles in a similar manner as in (2).

$$L_{CB} = L_1(\phi_c)L_2(\theta_c)L_3(\psi_c) = const.$$

Let $\boldsymbol{X}_t$, $\boldsymbol{V}_t$ and $\boldsymbol{a}_t$ be the position, velocity and acceleration vectors of the target in F_I. Then, the target dynamics become

$$\dot{\boldsymbol{X}}_t = \boldsymbol{V}_t, \quad \dot{\boldsymbol{V}}_t = \boldsymbol{a}_t. \tag{5}$$

Since this paper supposes a ground target, its motion is restricted on a ground surface. Assuming a flat ground $Z = Z_0$, the target state must satisfy the following constraints.

$$\begin{cases} Z_t = Z_0 = const. \\ V_{t_Z} = a_{t_Z} = 0 \end{cases} \tag{6}$$

A relative position of the target with respect to the camera expressed in F_C is defined by

$$\begin{aligned} \boldsymbol{X}_t^c &= L_{CI}\left(\boldsymbol{X}_t - \boldsymbol{X}_c\right) \\ &= L_{CI}\left(\boldsymbol{X}_t - \boldsymbol{X}_v\right) - L_{CB}\Delta\boldsymbol{X}_c. \end{aligned} \tag{7}$$

Assuming a pinhole camera model shown in Figure 3, pixel image coordinates of the target (in F_i) can be expressed by the following nonlinear function of $\boldsymbol{X}_t^c$.

$$\boldsymbol{x}_t = \boldsymbol{h}(\boldsymbol{X}_t^c) = \frac{f}{Z_t^c}\begin{bmatrix} X_t^c \\ Y_t^c \end{bmatrix} \tag{8}$$

where f (pixels) is the focal length of the camera.

C. Visual Target Tracking

An objective of this paper is to develop a visual navigation system which estimates the global target state $\{\boldsymbol{X}_t,\ \boldsymbol{V}_t\}$ by using information extracted from a sequence of images and other onboard sensors. Since the camera observes only a relative state of the target, the camera state (i.e. the UAV state) needs to be globally localized in order to obtain a global state of the target. Therefore, a navigation system is designed to estimate an augmented state $\{\boldsymbol{X}_v,\ \boldsymbol{V}_v,\ \boldsymbol{X}_t,\ \boldsymbol{V}_t\}$. An integrated GPS/INS navigation can provide a very accurate estimate of the UAV state.[8] However, its accuracy highly relies on GPS information. It is well-known that an INS-only navigation introduces a large drift due to an accumulation of measurement biases. Thus, this paper proposes to make use of vision information not only for target tracking but also for compensation of such an estimation drift when GPS signals are not accessible.

III. IMAGE PROCESSING ALGORITHMS

Two tasks are devoted to image processing: target tracking and optical flow estimation. The image processing algorithms have been developed based on basic image processing routines that can be found for instance on the Kovesi's website.[9] The algorithms are validated on real image sequences recorded during flights of the ONERA ReSSAC UAV, see Section V. An example of frame and image processing results is illustrated in Figure 4.

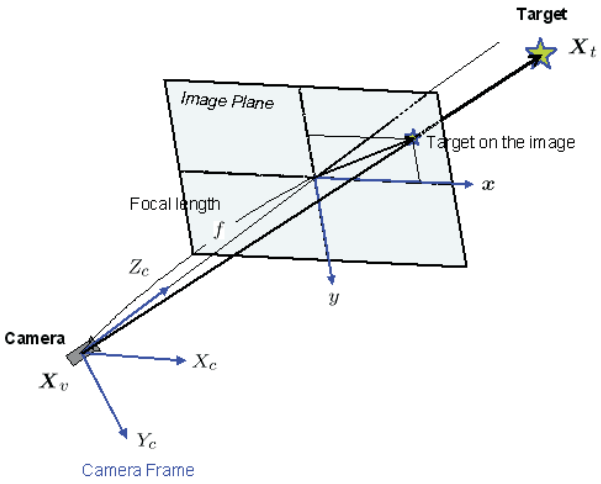


Figure 3. Pinhole Camera Model

A. Target Tracking

The goal of the target tracker is to find the pixel image coordinates $\boldsymbol{x}_{t_k}$ of the target in the current image I_k, knowing a prediction of this position given by propagating the last image position $\boldsymbol{x}_{t_{k-1}}$ with its dynamic model derived from (1, 5, 8) in Section II. Note that the initial target position needs to be given by the user or by any automatic detection algorithm, for example based on color attributes. There are several approaches to video target tracking, see the recent survey by Yilmaz et al.[10] However, their study is out of the scope of this paper, which focuses more on the UAV navigation application of such a technique.

In this paper, the video tracking problem is made simpler by the following facts: (i) the approximate size of the object is known; (ii) the target's graylevel is in general significantly higher than the background. Hence our target tracker simply consists in convolving the current image by a Gaussian kernel whose size is fitted to the estimated object size, and in selecting a position attaining the maximum in this convolved image. When the algorithm selects more than one positions, the one closest to the predicted target position is chosen. If the detected position is far from the predicted one, the observation is discarded. This simple target tracker performs very fast and accurate, as shown in Figure 4-a), in the real sequences at hand.

B. Optical Flow Estimation

This paper proposes utilizing the background visual motion, i.e. *optical flow*, for the UAV own-ship navigation purpose. Since this paper supposes a ground target, we can assume that the image area around the detected target position corresponds to the ground surface. Under this assumption, the optical flow estimation is focused on the surroundings of the target by masking the image around the position given by the target tracker, while leaving out a central zone corresponding to the target itself (Figure 4-a)). Consider a point feature that appears at a position $\boldsymbol{x}_{p_k}$ on the current image I_k and has appeared at $\boldsymbol{x}_{p_{k-1}}$ on the previous image I_{k-1}. Then, its optical flow vector is given by

$$\Delta \boldsymbol{x}_p = \boldsymbol{x}_{p_k} - \boldsymbol{x}_{p_{k-1}}. \tag{9}$$

On the masked image, estimates of the background optical flow vectors $\{\hat{\Delta}\boldsymbol{x}_{p_n}\}_{n=1...,N}$ (typically $N = 50$) are obtained by point features tracking. First, as shown in Figure 4-b), the feature points are detected both on I_k and I_{k-1} by using a Harris-Stephen operator.[9] Then, feature matching between the two images is simply done by a back and forth correlation computed on 15×15 windows. Since the feature points are localized in the center of highly textured windows, their motion can be measured with a good precision. An example of computed optical flow vectors is given in Figure 4-c). Finally, an *affine* motion model given by

$$\boldsymbol{x}_{p_k} = A_k \boldsymbol{x}_{p_{k-1}} + \boldsymbol{b}_k \tag{10}$$

is robustly fitted to the estimated flow vectors. This affine approximation is reasonable under the hypothesis that the ground is locally flat around the target. In this paper, the RANSAC program[9] is applied to eliminate

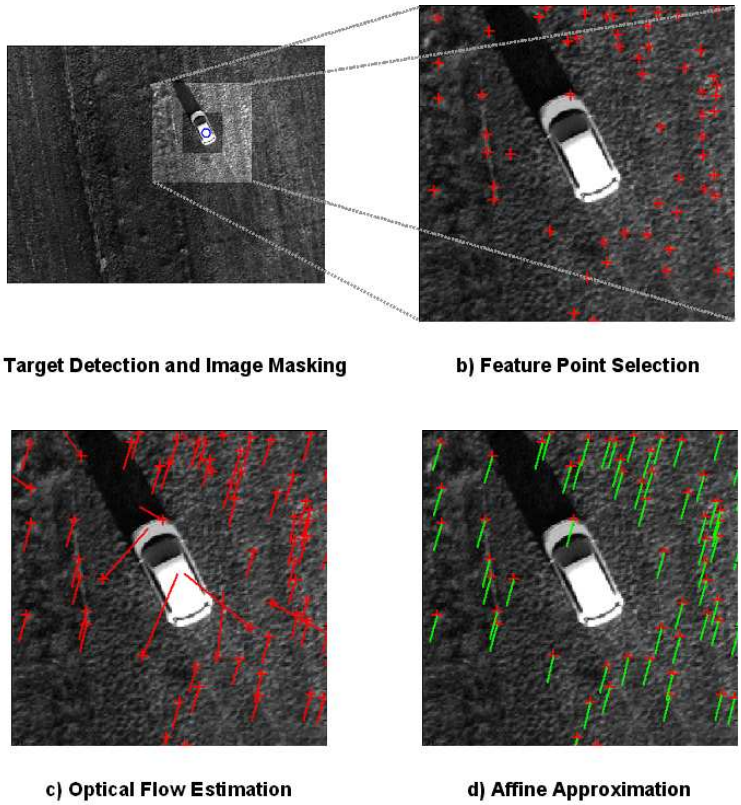


Figure 4. Image Processing Results

outliers and calculate a linear least-squares estimate of the affine parameters $(A_k,\ \boldsymbol{b}_k)$ on the inliers vectors. Figure 4-d) shows the affine approximation of the optical flow field given in Figure 4-c). In the navigation filter design discussed in the next section, the affine motion model (10) is reformulated in terms of the optical flow vector as follows.

$$\begin{aligned} \Delta \boldsymbol{x}_p &= (I - A_k^{-1})\boldsymbol{x}_{p_k} + A_k^{-1}\boldsymbol{b}_k \\ &= \bar{A}_k \boldsymbol{x}_{p_k} + \bar{\boldsymbol{b}}_k \end{aligned} \tag{11}$$

In summary, the image processor processes two consecutive images I_{k-1} and I_k, and provides a measurement set $\{\boldsymbol{x}_{tk},\ (\bar{A}_k,\ \bar{\boldsymbol{b}}_k)\}$.

IV. NAVIGATION FILTER DESIGN

This section designs an EKF-based navigation filter which estimates global position and velocity of the target as well as the UAV from the image processing outputs obtained in Section III and the onboard inertial sensor measurements.

A. Inertial Sensor Measurements

Suppose that the UAV is equipped with an inertial measurement unit (IMU) that estimates attitude angles $(\phi,\ \theta,\ \psi)$ of the UAV and also outputs its biased body acceleration. In this paper, it is assumed that the attitude angle estimates are sufficiently accurate so that the navigation filter can trust their values in calculation of the body rotation matrix L_{BI} in (2). The acceleration measurement can be modeled by

$$\boldsymbol{z}_{IMU} = L_{BI}\boldsymbol{a}_v + \Delta \boldsymbol{a}_v + \boldsymbol{\nu}_{IMU} \tag{12}$$

where $\Delta \boldsymbol{a}_v$ is the measurement bias and $\boldsymbol{\nu}_{IMU}$ is the measurement noise.

B. Process Model

An estimation state vector is chosen as follows.

$$\boldsymbol{x} = [\, \boldsymbol{X}_v^T \quad \boldsymbol{V}_v^T \quad \boldsymbol{X}_t^T \quad \boldsymbol{V}_t^T \quad \Delta \boldsymbol{a}_v^T \,]^T \tag{13}$$

A constant model is applied to dynamics of the acceleration measurement bias $\Delta \boldsymbol{a}_v$, and also to the unknown target acceleration $\boldsymbol{a}_t$. Then, from (1,5) and (12), a process model associated with the state $\boldsymbol{x}$ is derived as the following time-varying linear system.

$$\begin{aligned} \dot{\boldsymbol{x}} &= \begin{bmatrix} O & I & O & O & O \\ O & O & O & O & -L_{BI}^T \\ O & O & O & I & O \\ O & O & O & O & O \\ O & O & O & O & O \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} O \\ L_{BI}^T \\ O \\ O \\ O \end{bmatrix} \boldsymbol{z}_{IMU} + \begin{bmatrix} \boldsymbol{0} \\ -L_{BI}^T \boldsymbol{\nu}_{IMU} \\ \boldsymbol{0} \\ \boldsymbol{w}_{a_t} \\ \boldsymbol{w}_{\Delta a_v} \end{bmatrix} \\ &= F\boldsymbol{x} + G\boldsymbol{z}_{IMU} + \boldsymbol{w} \end{aligned} \tag{14}$$

where $\boldsymbol{w}$ is an augmented process noise.

C. Measurement Model

A measurement vector is composed of the image processor outputs and the target vertical velocity constraint given in (6), which is treated as a perfect measurement.

$$\begin{aligned} \boldsymbol{z}_k &= \begin{bmatrix} \boldsymbol{x}_{t_k}^T & \left[vec(\bar{A}_k) \right]^T & \bar{\boldsymbol{b}}_k^T & V_{tz\,k} \end{bmatrix}^T + \boldsymbol{\nu}_k \\ &= \boldsymbol{f}(\boldsymbol{x}_k) + \boldsymbol{\nu}_k \end{aligned} \tag{15}$$

where $\boldsymbol{\nu}_k$ is an augmented measurement noise. The target image position $\boldsymbol{x}_{tk}$ has already been modeled as a function of the estimation state $\boldsymbol{x}_k$ in (7, 8). A vertical velocity of the target V_{tz} itself is one of the state. Therefore, the measurement model $\boldsymbol{f}(\boldsymbol{x}_k)$ in (15) will be completed with an expression of the optical flow field model $(\bar{A}_k,\ \bar{\boldsymbol{b}}_k)$ in terms of the estimation state $\boldsymbol{x}_k$.

Suppose that an image position $\boldsymbol{x}_{p_k}$ on a current image I_k corresponds to a feature point $\boldsymbol{X}_p$ fixed on the ground surface $Z = Z_0$. Then, a position of the feature point in F_C at a current time step t_k can be derived as follows.

$$\boldsymbol{X}_{p_k}^c = \frac{Z_{p_k}^c}{f} \begin{bmatrix} \boldsymbol{x}_{p_k} \\ f \end{bmatrix} = \frac{Z_0 - Z_{ck}}{\boldsymbol{e}_3^T L_{CIk}^T \begin{bmatrix} \boldsymbol{x}_{p_k} \\ f \end{bmatrix}} \begin{bmatrix} \boldsymbol{x}_{p_k} \\ f \end{bmatrix}$$

where $\boldsymbol{e}_3^T = [\, 0 \quad 0 \quad 1 \,]$. From a relationship of $\boldsymbol{X}_p = \boldsymbol{X}_{ck} + L_{CIk}^T \boldsymbol{X}_{p_k}^c$, the feature point position in F_C at the previous time step t_{k-1} becomes

$$\begin{aligned} \boldsymbol{X}_{p_{k-1}}^c &= L_{CIk-1} \left(\boldsymbol{X}_p - \boldsymbol{X}_{ck-1} \right) \\ &= L_{CIk-1} L_{CIk}^T \boldsymbol{X}_{p_k}^c + L_{CIk-1} \left(\boldsymbol{X}_{ck} - \boldsymbol{X}_{ck-1} \right). \end{aligned}$$

Assuming a pin-hole camera model shown in Figure 3, the optical flow vector $\Delta \boldsymbol{x}_{p_k}$ at the image position $\boldsymbol{x}_{p_k}$ is approximated by

$$\begin{aligned} \Delta \boldsymbol{x}_{p_k} &= \boldsymbol{x}_{p_k} - \boldsymbol{x}_{p_{k-1}} = \boldsymbol{h}(\boldsymbol{X}_{p_k}^c) - \boldsymbol{h}(\boldsymbol{X}_{p_{k-1}}^c) \\ &\simeq \frac{\partial \boldsymbol{h}(\boldsymbol{X}_{p_k}^c)}{\partial \boldsymbol{X}_{p_k}^c} \left(\boldsymbol{X}_{p_k}^c - \boldsymbol{X}_{p_{k-1}}^c \right) = \begin{bmatrix} I & -\frac{\boldsymbol{x}_{p_k}}{f} \end{bmatrix} M(\boldsymbol{x}_k) \begin{bmatrix} \boldsymbol{x}_{p_k} \\ f \end{bmatrix} \end{aligned} \tag{16}$$

where the matrix function $M(\boldsymbol{x}_k)$ is given by

$$M(\boldsymbol{x}_k) = \left(\delta L_{CIk} - \frac{L_{CIk-1} \delta \boldsymbol{X}_{ck} \boldsymbol{e}_3^T}{Z_{tk} - Z_{ck}} \right) L_{CIk}^T. \tag{17}$$

The first term of (17) is a component associated with the camera's rotational motion

$$\delta L_{CIk} = L_{CIk} - L_{CIk-1} = L_{CB} \left(L_{BIk} - L_{BIk-1} \right) = L_{CB} \delta L_{BIk}$$

The second term is a component associated with the translational motion.

$$\delta \boldsymbol{X}_{ck} = \boldsymbol{X}_{ck} - \boldsymbol{X}_{ck-1} \simeq \boldsymbol{V}_{vk}(t_k - t_{k-1}) + \delta L_{BIk}^T \Delta \boldsymbol{X}_c$$

Note that the ground height Z_0 is replaced by the target height Z_{tk} in (17). The affine model of the optical flow field, $(\bar{A}_k,\ \bar{\boldsymbol{b}}_k)$, is obtained by linearizing (16) about the detected target position $\boldsymbol{x}_{tk}$.

$$\begin{cases} \bar{A}_k = \left[\, I \quad -\frac{\boldsymbol{x}_{tk}}{f} \,\right] M(\boldsymbol{x}_k) \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} - \boldsymbol{e}_3^T M(\boldsymbol{x}_k) \begin{bmatrix} \frac{\boldsymbol{x}_{tk}}{f} \\ 1 \end{bmatrix} I = \bar{A}_k\left(Z_{v_k}, \boldsymbol{V}_{v_k}, Z_{t_k}\right) \\ \bar{\boldsymbol{b}}_k = f\left[\, I \quad \mathbf{0} \,\right] M(\boldsymbol{x}_k)\boldsymbol{e}_3 + \frac{\boldsymbol{x}_{tk}}{f}\boldsymbol{e}_3^T M(\boldsymbol{x}_k) \begin{bmatrix} \boldsymbol{x}_{tk} \\ \mathbf{0} \end{bmatrix} = \bar{\boldsymbol{b}}_k\left(Z_{v_k}, \boldsymbol{V}_{v_k}, Z_{t_k}\right) \end{cases} \tag{18}$$

In summary, the measurement model $\boldsymbol{f}(\boldsymbol{x}_k)$ in (15) is given by

$$\boldsymbol{f}(\boldsymbol{x}_k) = \begin{bmatrix} \boldsymbol{h}(\boldsymbol{X}_{tk}^c) \\ vec\left(\bar{A}_k\left(Z_{v_k}, \boldsymbol{V}_{v_k}, Z_{t_k}\right)\right) \\ \bar{\boldsymbol{b}}_k\left(Z_{v_k}, \boldsymbol{V}_{v_k}, Z_{t_k}\right) \\ V_{tZ\,k} \end{bmatrix}$$

The measurement matrix associated with this model can be derived as follows.

$$\begin{aligned} H_k(\boldsymbol{x}_k) &= \frac{\partial \boldsymbol{f}(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k} \\ &= \begin{bmatrix} \frac{1}{Z_{tk}^c}\left[\, fI \quad -\boldsymbol{x}_{t_k} \,\right] L_{CI_k} & O & O & O & O \\ -\left[\, \mathbf{0} \quad \mathbf{0} \quad vec(\bar{A}_k^Z) \,\right] & \left[\, vec(\bar{A}_k^{V_1}) \quad vec(\bar{A}_k^{V_2}) \quad vec(\bar{A}_k^{V_3}) \,\right] & \left[\, \mathbf{0} \quad \mathbf{0} \quad vec(\bar{A}_k^Z) \,\right] & O & O \\ -\left[\, \mathbf{0} \quad \mathbf{0} \quad \bar{\boldsymbol{b}}_k^Z \,\right] & \left[\, \bar{\boldsymbol{b}}_k^{V_1} \quad \bar{\boldsymbol{b}}_k^{V_2} \quad \bar{\boldsymbol{b}}_k^{V_3} \,\right] & \left[\, \mathbf{0} \quad \mathbf{0} \quad \bar{\boldsymbol{b}}_k^Z \,\right] & O & O \\ O & O & O & \left[\, 0 \quad 0 \quad 1 \,\right] & O \end{bmatrix} \end{aligned} \tag{19}$$

where

$$\begin{aligned} \bar{A}_{k_Z} &= \frac{1}{Z_{tk} - Z_{ck}} \left(\left[\, I \quad -\frac{\boldsymbol{x}_{tk}}{f} \,\right] \delta L_{CI_k} L_{CI_k}^T \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} - \boldsymbol{e}_3^T \delta L_{CI_k} L_{CI_k}^T \begin{bmatrix} \frac{\boldsymbol{x}_{tk}}{f} \\ 1 \end{bmatrix} I - \bar{A}_k\left(Z_{v_k}, \boldsymbol{V}_{v_k}, Z_{t_k}\right) \right) \\ \bar{A}_k^{V_i} &= -\frac{t_k - t_{k-1}}{Z_{tk} - Z_{ck}} \left(\left[\, I \quad -\frac{\boldsymbol{x}_{tk}}{f} \,\right] L_{CIk-1} \boldsymbol{e}_i \boldsymbol{e}_3^T L_{CI_k}^T \begin{bmatrix} I \\ \mathbf{0} \end{bmatrix} - \boldsymbol{e}_3^T L_{CIk-1} \boldsymbol{e}_i \boldsymbol{e}_3^T L_{CI_k}^T \begin{bmatrix} \frac{\boldsymbol{x}_{tk}}{f} \\ 1 \end{bmatrix} I \right) \\ \bar{\boldsymbol{b}}_k^Z &= \frac{1}{Z_{tk} - Z_{ck}} \left(f\left[\, I \quad \mathbf{0} \,\right] \delta L_{CI_k} L_{CI_k}^T \boldsymbol{e}_3 + \frac{\boldsymbol{x}_{tk}}{f} \boldsymbol{e}_3^T \delta L_{CI_k} L_{CI_k}^T \begin{bmatrix} \boldsymbol{x}_{tk} \\ \mathbf{0} \end{bmatrix} - \bar{\boldsymbol{b}}_k\left(Z_{v_k}, \boldsymbol{V}_{v_k}, Z_{t_k}\right) \right) \\ \bar{\boldsymbol{b}}_k^{V_i} &= -\frac{t_k - t_{k-1}}{Z_{tk} - Z_{ck}} \left(f\left[\, I \quad \mathbf{0} \,\right] L_{CIk-1} \boldsymbol{e}_i \boldsymbol{e}_3^T L_{CI_k}^T \boldsymbol{e}_3 + \frac{\boldsymbol{x}_{tk}}{f} \boldsymbol{e}_3^T L_{CIk-1} \boldsymbol{e}_i \boldsymbol{e}_3^T L_{CI_k}^T \begin{bmatrix} \boldsymbol{x}_{tk} \\ \mathbf{0} \end{bmatrix} \right) \end{aligned}$$

D. EKF Formulation

Since the measurement model (15) is a nonlinear function of the estimation state $\boldsymbol{x}$, an extended Kalman filter (EKF) is applied. An EKF procedure includes two different steps: time propagation and measurement update. The first step is performed based on a process model, and the second one on a measurement model. Let $\hat{\boldsymbol{x}}_k^-$ and $\hat{\boldsymbol{x}}_k$ denote the propagated and updated estimates of the state $\boldsymbol{x}_k$ at time t_k, and P_k^- and P_k denote their error covariance matrices respectively. Given a process model (14) and a measurement model (15), the EKF procedure is summarized as follows.[11]

- Time Propagation :

$$\begin{cases} \hat{\boldsymbol{x}}_k^- = \Phi_k \hat{\boldsymbol{x}}_{k-1} + G_k \boldsymbol{z}_{IMUk}(t_k - t_{k-1}) \\ P_k^- = \Phi_k P_{k-1} \Phi_k^T + Q_k, \quad \Phi_k = I + F_k(t_k - t_{k-1}) \end{cases}$$

  where Q_k is a covariance matrix of the discretized process noise $\boldsymbol{w}_k$.

- Measurement Update :

$$\begin{cases} \hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_k^- + K_k\left(\boldsymbol{z}_k - \boldsymbol{f}(\hat{\boldsymbol{x}}_k^-)\right) \\ P_k = \left(I - K_k H_k(\hat{\boldsymbol{x}}_k^-)\right) P_k^- \end{cases}$$

where $H_k(\hat{\boldsymbol{x}}_k^-)$ is the measurement matrix (19) evaluated at the predicted state $\hat{\boldsymbol{x}}_k^-$. K_k is the Kalman gain defined as follows.

$$K_k = P_k^- H_k^T(\hat{\boldsymbol{x}}_k^-)\left(H_k(\hat{\boldsymbol{x}}_k^-)P_k^- H_k^T(\hat{\boldsymbol{x}}_k^-) + R_k\right)^{-1}$$

R_k is a covariance matrix of the measurement noise ν_k.

V. SIMULATIONS

The image processing algorithm developed in Section III and the navigation filter designed in Section IV are integrated, and tested on onboard camera images and vehicle state data that were synchronically recorded in actual flights of the ONERA ReSSAC autonomous helicopter.[12] The ReSSAC helicopter is based on the YAMAHA RMax industrial helicopter and its specifications are summarized in Figure 5. In this section, simulation results are presented to compare navigation performances with and without optical flow measurements.

A. Flight Test Configuration

This subsection explains a configuration of the flight test for data recording. A camera is mounted on the ReSSAC UAV helicopter underneath the body with its optical axis pointing straight down. ($\phi_c = \theta_c = 0°, \psi_c = 90°$). The camera attitude and displacement with respect to the body frame F_B are measured as

$$L_{CB} = L_1(-1.5°)L_2(-2.1°)L_3(90°), \quad \Delta\boldsymbol{X}_c = \begin{bmatrix} 0.1 \\ 0 \\ 0.499 \end{bmatrix} (m).$$

The focal length of the camera is $f = 658$ (pixels). During the data recording, the UAV flew automatically with a constant speed of 6 (m/sec) at a constant height of 40 (m) along an airstrip, whose direction is East/West. A manually-driven car played a role of a ground target in this experiment. During the flight, the car ran windingly on the airstrip while remaining in the camera's field of view. Camera images were saved at 10 (Hz) with the most recently updated UAV state data, including a navigation time, the global position and velocity (the GPS/IMU navigation solution), the body acceleration and the attitude angles (IMU outputs). The GPS measurements of the target global position and velocity were recorded at 1 (Hz) separately from the UAV state and the images.

B. EKF Initialization and Design Parameters

Suppose a situation in which GPS signals are lost when obtained the first image of the data set, say at t_1. The target tracker was manually initialized on the first image, at a position $\boldsymbol{x}_{t1}$. Since the optical flow estimation requires two consecutive images, it starts at t_2 when the second image was obtained. The EKF is also initialized at t_2 when the first complete measurement set $\{\boldsymbol{x}_{tk},\ (\bar{A}_k,\ \bar{\boldsymbol{b}}_k)\}$ was provided. The estimates of the UAV position and velocity are initialized by propagating the GPS/INS navigation result at t_1, denoted by $(\hat{\boldsymbol{X}}_{v_1}, \hat{\boldsymbol{V}}_{v_1})$.

$$\begin{cases} \hat{\boldsymbol{X}}_{v_2} = \hat{\boldsymbol{X}}_{v_1} + \hat{\boldsymbol{V}}_{v_1}(t_2 - t_1) \\ \hat{\boldsymbol{V}}_{v_2} = \hat{\boldsymbol{V}}_{v_1} + L_{BI_2}^T \boldsymbol{z}_{IMU_2}(t_2 - t_1) \end{cases} \tag{20}$$

Length	3.63 (m)
Weight	60 (kg)
Payload	20 (kg)
Sensors	GPS/INS, compass, barometer, vision

Figure 5. The ONERA ReSSAC Helicopter

The estimates of the target position and velocity are initialized by using the image processor outputs. Importantly, estimates of the target height Z_0 and the target's velocity are computed from a difference between optical flows of the target and of the ground surface.

Consider a fixed ground point $\boldsymbol{X}_p$ whose location coincides with the target position at t_2. From the target tracker output $\boldsymbol{x}_{t_2}$, this position can be written as

$$\boldsymbol{X}_p = \boldsymbol{X}_{t_2} = \frac{Z^c_{t_2}}{f} L^T_{CI_2} \begin{bmatrix} \boldsymbol{x}_{t_2} \\ f \end{bmatrix} + \boldsymbol{X}_{c_2} \tag{21}$$

with an unknown depth $Z^c_{t_2}$. From the optical flow estimation result, the position where this ground point $\boldsymbol{X}_p$ appeared on the first image can be calculated as follows.

$$\boldsymbol{x}_{p_1} = \boldsymbol{x}_{t_2} + \bar{A}_2 \boldsymbol{x}_{t_2} + \bar{\boldsymbol{b}}_2 = \left(I + \bar{A}_2\right) \boldsymbol{x}_{t_2} + \bar{\boldsymbol{b}}_2$$

Same as done in (21), $\boldsymbol{X}_p$ can be also expressed by

$$\boldsymbol{X}_p = \frac{Z^c_{p_1}}{f} L^T_{CI_1} \begin{bmatrix} \boldsymbol{x}_{p_1} \\ f \end{bmatrix} + \boldsymbol{X}_{c_1} = \frac{Z^c_{p_1}}{f} L^T_{CI_1} \begin{bmatrix} \left(I + \bar{A}_2\right) \boldsymbol{x}_{t_2} + \bar{\boldsymbol{b}}_2 \\ f \end{bmatrix} + \boldsymbol{X}_{c_1} \tag{22}$$

with an unknown depth $Z^c_{p_1}$. By subtracting (22) from (21), we obtain

$$\delta \boldsymbol{X}_{c_2} = \frac{1}{f} \left[L^T_{CI_1} \begin{bmatrix} \left(I + \bar{A}_2\right) \boldsymbol{x}_{t_2} + \bar{\boldsymbol{b}}_2 \\ f \end{bmatrix} \quad -L^T_{CI_2} \begin{bmatrix} \boldsymbol{x}_{t_2} \\ f \end{bmatrix} \right] \begin{bmatrix} Z^c_{p_1} \\ Z^c_{t_2} \end{bmatrix} = \frac{1}{f} C(\boldsymbol{x}_{t2}, \bar{A}_2, \bar{\boldsymbol{b}}_2) \begin{bmatrix} Z^c_{p_1} \\ Z^c_{t_2} \end{bmatrix}$$

This linear equation can be solved about the two unknowns $(Z^c_{p_1}, Z^c_{t_2})$ as follows.

$$\frac{1}{f} \begin{bmatrix} Z^c_{t_1} \\ Z^c_{t_2} \end{bmatrix} = (C^T C)^{-1} C^T \delta \boldsymbol{X}_{c_2}$$

Therefore, the target position estimate at t_2 is given by

$$\hat{\boldsymbol{X}}_{t_2} = L^T_{CI_2} \begin{bmatrix} \boldsymbol{x}_{t_2} \\ f \end{bmatrix} [\,0 \quad 1\,] (C^T C)^{-1} C^T \delta \hat{\boldsymbol{X}}_{c_2} + \hat{\boldsymbol{X}}_{c_2} \tag{23}$$

Similarly to (21), from the initial target tracker output $\boldsymbol{x}_{t_1}$, the target position at t_1 is calculated by

$$\boldsymbol{X}_{t_1} = \frac{Z^c_{t_1}}{f} L^T_{CI_1} \begin{bmatrix} \boldsymbol{x}_{t_1} \\ f \end{bmatrix} + \boldsymbol{X}_{c_1}$$

with an unknown depth $Z^c_{t_1}$. By using the target dynamics,

$$\hat{\boldsymbol{V}}_{t_2} = \frac{\hat{\boldsymbol{X}}_{t_2} - \boldsymbol{X}_{t_1}}{t_2 - t_1} = \frac{1}{t_2 - t_1} \left(\left(I + L^T_{CI_2} \begin{bmatrix} \boldsymbol{x}_{t_2} \\ f \end{bmatrix} [\,0 \quad 1\,] (C^T C)^{-1} C^T \right) \delta \hat{\boldsymbol{X}}_{c_2} - \frac{Z^c_{t_1}}{f} L^T_{CI_1} \begin{bmatrix} \boldsymbol{x}_{t_1} \\ f \end{bmatrix} \right) \tag{24}$$

From the ground target constraint (6), the unknown depth $Z^c_{t_1}$ can be derived.

$$\frac{Z^c_{t_1}}{f} = \frac{\boldsymbol{e}_3^T}{\boldsymbol{e}_3^T L^T_{CI_1} \begin{bmatrix} \boldsymbol{x}_{t_1} \\ f \end{bmatrix}} \left(I + L^T_{CI_2} \begin{bmatrix} \boldsymbol{x}_{t_2} \\ f \end{bmatrix} [\,0 \quad 1\,] (C^T C)^{-1} C^T \right) \delta \hat{\boldsymbol{X}}_{c_2}$$

The acceleration measurement bias is initialized as zero, i.e., $\hat{\boldsymbol{a}}_{v_2} = \boldsymbol{0}$. The initial error covariance is chosen as a block-diagonal-matrix of the following sub-matrices.

$$P_{\boldsymbol{X}_v} = 1^2 I, \quad P_{\boldsymbol{V}_v} = 1^2 I, \quad P_{\boldsymbol{X}_t} = 10^2 I, \quad P_{\boldsymbol{V}_t} = \begin{bmatrix} 10^2 & 0 & 0 \\ 0 & 10^2 & 0 \\ 0 & 0 & 1^2 \end{bmatrix}, \quad P_{\Delta \boldsymbol{a}_v} = 0.2^2 I$$

The covariance matrices of the process noise (Q_k) and the measurement error (R_k) used in the EKF design are block-diagonal matrices containing;

$$Q_{\boldsymbol{X}_v} = 0.1^2 I, \quad Q_{\boldsymbol{V}_v} = 0.1 I, \quad Q_{\boldsymbol{X}_t} = \begin{bmatrix} 10^2 & 0 & 0 \\ 0 & 10^2 & 0 \\ 0 & 0 & 1^2 \end{bmatrix}, \quad Q_{\boldsymbol{V}_t} = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}, \quad Q_{\Delta \boldsymbol{a}_v} = 0.02^2 I$$

$$R_{\boldsymbol{x}_t} = 5^2 I, \quad R_{\bar{A}} = 0.01^2 I, \quad R_{\bar{\boldsymbol{b}}} = (0.01 f)^2 I, \quad R_{V_{t_Z}} = 0.1^2$$

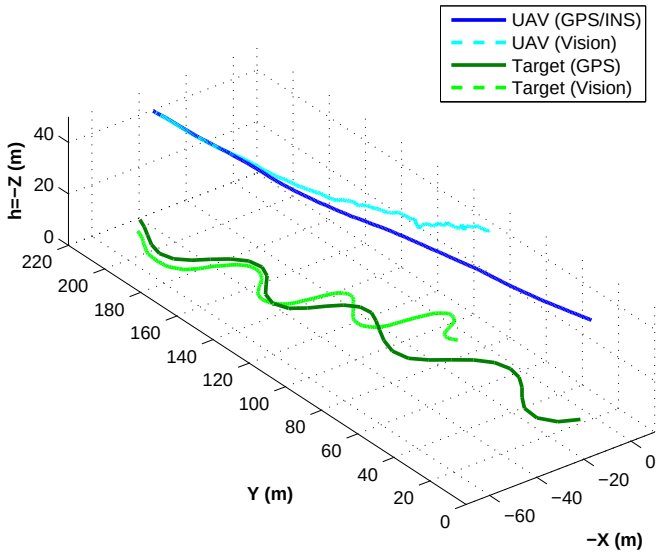


Figure 6. UAV and Target Trajectories (No Optical Flow)

C. Simulation Results

The first simulation was performed by using only the target tracker but not the optical flow estimation. An aim of this simulation is to see how quickly the INS-only estimate diverges. Figure 6 compares true and estimated trajectories of the UAV and of the target. The true UAV trajectory is given by the GPS/INS own-ship navigation solution, and the true target trajectory is the raw GPS measurement. As seen in Figure 6, a drift occurs in the UAV self-localization result in both horizontal and vertical directions. This drift leads a drift in the target localization because only the relative measurements are available.

The second simulation was performed by using simulated optical flow measurements instead of the outputs of the optical flow estimation algorithm. An objective of this simulation is to verify the theory of the navigation filter design in Section IV. Simulated optical flows were created by calculating optical flow vectors of eight ground points sampled from the area close to the target and then approximating them by an affine model. Figure 7 and 8 present the position estimation performance. Except the slight error in height, both of the UAV and the target positions are accurately estimated. This result indicates that the optical flow measurements are effectively utilized in the navigation design to complement the GPS information.

Lastly, Figure 9 and 10 illustrate the estimation performance when using the optical flow detection algorithms developed in Section III. Although the horizontal positions are estimated sufficiently accurate, there occurs a drift in height estimation. Several different factors can be considered as a cause of this drift: 1) camera calibration error; 2) measurement bias caused by image distortion; 3) error in the UAV attitude data; 4) wrong assumption of a flat ground. Further investigations are needed to assess this issue.

VI. CONCLUSION

This paper proposes a visual air-to-ground target tracking system which does not rely on GPS. It is suggested to use the optical flow field measurements in order to complement the GPS information. By applying the affine approximation to the optical flow measurements, the navigation filter effectively incorporates the flow information without an increase in computational load. Simulations have been performed with the actual flight test data, and the results demonstrated a capability of the suggested system to globally localize a ground target in a GPS-denied environment. As future work, we would like to implement the system onboard on the ReSSAC UAV and achieve a closed-loop flight of visual target tracking.

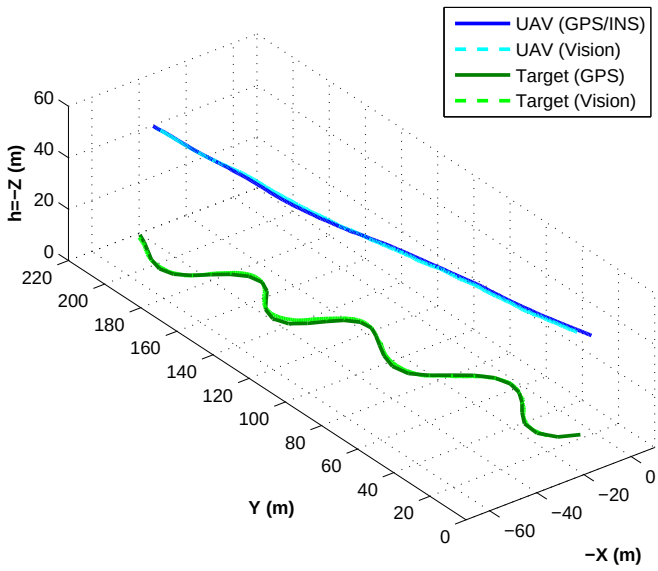
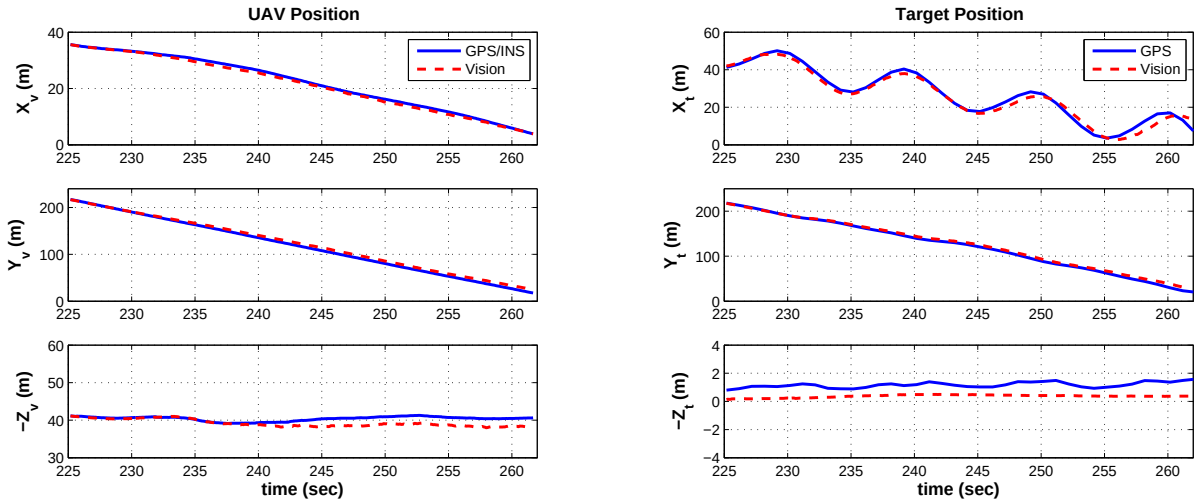


Figure 7. UAV and Target Trajectories (Simulated Optical Flow)



(a) UAV Position (b) Target Position

Figure 8. Position Estimation Results (Simulated Optical Flow)

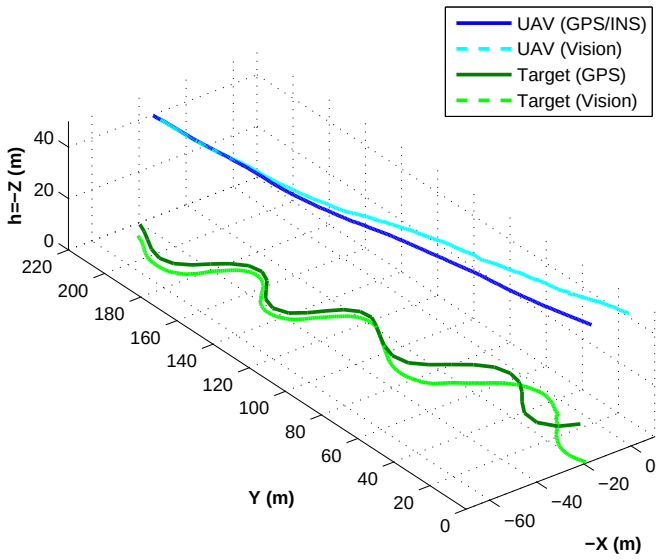
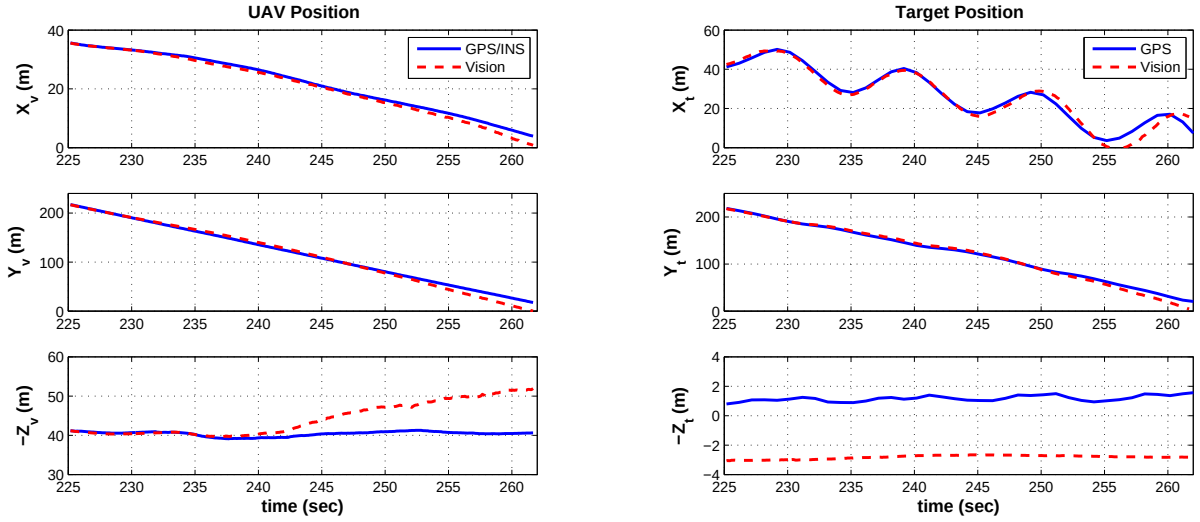


Figure 9. UAV and Target Trajectories (Measured Optical Flow)



(a) UAV Position

(b) Target Position

Figure 10. Position Estimation Results (Measured Optical Flow)

References

[1]M.V. Srinivasan, S.W. Zhang, M. Lehrer and T.S. Collett, "Honeybee Navigation *En-Route* to the Goal: Visual Flight Control and Odometry", Journal of Experimental Biology, Vol.199, No.1, 1996.

[2]L. Muratet, S. Doncieux , J.A. Meyer, "A Biomimetic Reactive Navigation System Using the Optical Flow for a Rotary-Wing UAV in Urban Environment", International Symposium on Robotics, 2004.

[3]S. Hrabar, G.S. Sukhatme, P. Corke, K. Usher and J. Roberts, "Combined Optic-Flow and Stereo-Based Navigation of Urban Canyons for a UAV", IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005.

[4]R. Smith and P. Chessman, "On the Representation of Spatial Uncertainty", International Journal of Robotics Research, Vol.5, No.4, 1987.

[5]H. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping : Part I, Part II", IEEE Robotics & Automation Magazine, Vol.13, 2006.

[6]J. Kim and S. Sukkarieh, "6 DoF SLAM Aided GNSS/INSS Navigation in GNSS Denied and Unknown Environments", Journal of Global Positioning Systems, Vol.4, No.1-2, 2005.

[7]J. Wang, M. Garratt, A. Lambert, J.J. Wang, S. Han and D. Sinclair, "Integration of GPS/INS/Vision Sensors to Navigate Unmanned Aerial Vehicles", International Society for Photogrammetry and Remote Sensing (ISPRS) Congress, 2008.

[8]C.S. Yoo and I.K.Ahn, "Low Cost GPS/INS Sensor Fusion System for UAV Navigation", Digital Avionics Systems Conference, 2003.

[9]P. Kovesi, "MATLAB and Octave Functions for Computer Vision and Image Processing", The University of Western Australia, 2000. http://www.csse.uwa.edu.au/∼pk/research/matlabfns/

[10]A. Yilmaz, O. Javed and M. Shah, "Object Tracking: A Survey", ACM Computing Surveys, Vol.38, No.4, 2005.

[11]R.G. Brown and P.Y.C. Hwang, "Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises and Solutions", John Wiley & Sons, 1997.

[12]P. Fabiani, V. Fuertes, G. Le Besnerais, R. Mampey, A. Piquereau and F. Teichteil-Königsbuch. "ReSSAC: Flying an Autonomous Helicopter in a Non-Cooperative Uncertain World", AHS Specialist Meeting on Unmanned Rotorcraft, 2007.