Error Analysis and Attitude Observability of a Monocular GPS/Visual Odometry Integrated Navigation Filter

Damien Dusha, Luis Mejias
Australian Research Centre for Aerospace Automation
Queensland University of Technology
Brisbane, Australia
{d.dusha,luis.mejias}@qut.edu.au

May 14, 2011

Abstract

In this paper, we present a method for the recovery of position and absolute attitude (including pitch, roll and yaw) using a novel fusion of monocular Visual Odometry and GPS measurements in a similar manner to a classic loosely-coupled GPS/INS error state navigation filter. The proposed filter does not require additional restrictions or assumptions such as platform-specific dynamics, map-matching, feature-tracking, visual loop-closing, gravity vector or additional sensors such as an IMU or magnetic compass.

An observability analysis of the proposed filter is performed, showing that the scale factor, position and attitude errors are fully observable under acceleration that is non-parallel to velocity vector in the navigation frame. The observability properties of the proposed filter are demonstrated using numerical simulations.

We conclude the article with an implementation of the proposed filter using real flight data collected from a Cessna 172 equipped with a downwards-looking camera and GPS, showing the feasibility of the algorithm in real-world conditions.

1 Introduction

In the last decade, researchers have expended substantial effort on vision-based navigation, including the closely-related Structure From Motion (SFM) and Simultaneous Localisation And Mapping (SLAM) problems. Much of the vision-based navigation literature has been driven by applications where Global Navigation Satellite Systems (GNSS) such as the Global Positioning System (GPS)[1] is unavailable (such as the Mars Exploration Rover [30]) or denied [5].

In the absence of a globally-referenced position such as GPS, a myriad of schemes have been proposed to limit (or at least slow) the position drift inherent

---

[1]In this paper, GPS is used synonymously with GNSS

in a vision-based navigation solution [12]. Such schemes include the SLAM problem, which seeks to bound position and attitude drift with the re-observation of landmarks [5]. Other schemes for constraining position and attitude drift include comparisons to digital elevation models [26] or pre-stored georeferenced images [9]. Other authors attempt to limit drift with the integration of additional sensors such as an Inertial Measurement Unit (IMU) [22].

Despite the near-ubiquity of GPS in outdoor applications and array of consumer devices (such as cellphones) equipped with both a camera and GPS receiver, comparatively little attention has been paid to the potential synergies between GPS and vision-based navigation. Whilst it is inherently clear that GPS can be used to constrain position drift in a vision-based navigation solution, drift in attitude is not so obviously constrained – a single-antenna GPS system cannot provide an attitude solution without exploiting additional information such as signal to noise ratio [36] or making specific assumptions about platform dynamics, such as the coordinated flight constraint used in GPS pseudo-attitude [23]. The key to constraining attitude drift in a combined GPS and vision-based navigation system lies in the observation that errors in vision-based attitude will, when integrated over time, couple into errors in position. The GPS position may then be used to correct both the position and correlated attitude errors when subjected to particular motion conditions.

In this paper, we derive the error dynamics of a monocular Visual Odometry (VO) navigation system and show how they can be combined with GPS in a similar manner to a classic GPS/INS system in order to constrain the drift in position and attitude. This is achieved without additional sensors (such as a compass or inertial sensors), platform-specific dynamics (such as coordinated flight or the non-holonomic constraint), or many of the complexities imposed on other vision-based solutions such as map-matching, visual loop-closing or explicit feature mapping. We show through an observability analysis that all attitude errors are observable in the presence of acceleration that is non-parallel to the velocity vector in the navigation frame.

Since additional sensors are not required for attitude determination in a combined GPS/VO system, this research may have future applications in determining the attitude of devices that are often not equipped with an Inertial Measurement Unit (such as a cellphone), or where the attitude of the camera is not fixed to the host platform, such as a gimballed camera. Alternatively, the combination of GPS and vision could potentially serve as a backup to a traditional IMU-based attitude solution on an Unmanned Airborne Vehicle (UAV) equipped with a GPS, camera and sufficient computing.

1.1 Related Work

Whilst vision-based navigation has been a popular topic of research over the last decade (see the survey in [4] for taxonomy of current methods), comparatively little attention has been paid to the combination of GPS and Vision. Of the research that exists combining GPS and Vision, much of it (with notable exceptions) is contained to the problem of in-car navigation where autonomous GPS positioning is not sufficiently accurate to localise a vehicle within a lane. For example, Li [27] uses lane markings, road signs and a digital map to correct for GPS position and to recover the pose of the vehicle. Similarly, Rae [31] uses multiple hypothesis tracking of road markings to reduce the error on GPS

localisation and solve for vehicle orientation.

Domain specific methods combining GPS and vision have also been proposed for aerospace applications. These include the use of runway markers to recover position and pose [10] and the use of the optical flow focus of expansion to correct GPS pseudo-attitude pitch, roll and yaw measurements [33].

Recently, several more general recursive methods of combining GPS and vision have appeared in literature that are similar to the approach presented in this paper. Agrawal [1, 2] describes a ground robot combining stereo visual odometry, inertial measurements, wheel odometry and GPS using a Kalman Filter to constrain long-term drift and to bridge GPS outages. GPS velocity-based heading is used to explicitly constrain yaw. Similarly, Wei [37] fuses stereo visual odometry and GPS to smooth GPS errors such as multipath, noting also that yaw errors are substantially reduced when GPS is fused with stereo vision. GPS has also been used to improve the accuracy of stereo-vision SLAM [34] and facilitate hierarchical separation between local and global maps.

GPS has also been used to improve vision-based methods employing batch processing. Kume [24] improved the estimation of camera extrinsic parameters by using the position accuracy of GPS to adjust the cost function of bundle adjustment. The use of GPS has also been explored in the Structure from Motion problem, where Carceroni [8] investigates its effect on pose recovery and the number of point correspondences required between views.

The bulk of the work presented has been motivated by bridging GPS outages or to correct for commonly encountered errors such as multipath with comparatively little attention paid to the relationship between attitude errors and GPS corrections, though there is substantial literature investigating pose recovery in other navigation scenarios. In the visual SLAM problem, Caballero [6] uses a Visual-Odometry loop corrected periodically by a SLAM framework. Vidal-Calleja [35] shows feature triangulation from multiple position is required for pose recovery. In visual-inertial systems, Kelly [22] shows that IMU biases, rotation between the camera and IMU, gravity vector and metric scene structure are recoverable. Similarly, Jones [21] shows the motion conditions for a visual-inertial system under which the gravity, camera to IMU rotation can be estimated online.

Of most interest to this paper is the long history of observability analysis in GPS/INS systems, given the superficial similarity between IMU and egomotion observables. Using a piecewise linear model, Rhee [32] conducted an observability analysis, showing that non-constant acceleration allowed observability of attitude angles other than angle about the jerk (change of acceleration) vector. That is, yaw is recoverable with jerk in the horizontal plane. Later, Hong [20] conducted a more extensive observability analysis using a linear time-varying observability approach on the linearised error states and similarly showed that all unobservable states in the constant acceleration case can be made observable through manoeuvring, including the lever arm between the GPS antenna and IMU.

In preliminary work [15], we utilised the methodology by Hong to analyse the observability of a GPS/Visual Odometry integrated navigation filter where the scale factor of the scene is known (i.e. the stereo vision case). In this paper, we extend this approach for a monocular Visual Odometry system where the scale factor of the scene is unknown (i.e. the monocular case) and show that similar conclusions about the observability of error states can be drawn. Furthermore,

we validate the observability analysis using real data collected using a light aircraft.

1.2 Organisation

This paper is organised as follows: We firstly perform a review of GPS/INS mechanisation and observability, which serves as a template for the integrated GPS/Visual Odometry filter introduced the subsequent section. The linearised error dynamics of Visual Odometry and GPS corrections are derived in a form suitable for estimation via a Kalman Filter. An observability analysis of the proposed filter is then presented, followed by numerical simulations demonstrating its major properties. The paper concludes with an experimental validation of the proposed filter using real data collected from a downwards-looking camera mounted to a Cessna 172.

1.3 Notation and Identities

In this paper, the following notation has been adopted:

$\mathbf{a}^{\gamma}_{\alpha\beta}$ is a vector quantity $\mathbf{a}$ of β with respect to α, resolved in the γ frame.

A vector quantity with a tilde (e.g. $\tilde{\mathbf{a}}$) denotes an estimated or measured value of the true quantity $\mathbf{a}$.

A time derivative of a vector is denoted with the dot notation, e.g. $\dot{\mathbf{a}}$

The notation $[\mathbf{a}]_{\times}$ denotes a skew-symmetric matrix constructed from a vector $\mathbf{a}$ such that when multiplied by vector $\mathbf{b}$ is equivalent to the vector cross product of $\mathbf{a}$ and $\mathbf{b}$, i.e.

$$[\mathbf{a}]_{\times} \mathbf{b} = \mathbf{a} \times \mathbf{b} \tag{1}$$

Following the anti-symmetric property of the cross-product, we may write:

$$[\mathbf{a}]_{\times} \mathbf{b} = -[\mathbf{b}]_{\times} \mathbf{a} \tag{2}$$

where $\mathbf{a} \in \mathbb{R}^3, \mathbf{b} \in \mathbb{R}^3$.

A change of reference frame can be achieved with the following identity [18]:

$$\left[\boldsymbol{\omega}^d_{ba}\right]_{\times} = \mathbf{R}^d_c \left[\boldsymbol{\omega}^c_{ba}\right]_{\times} \mathbf{R}^c_d \tag{3}$$

where a, b, c, d are arbitrary reference frames.

The following identity from [29] is used:

$$[\mathbf{R}\mathbf{v}]_{\times} = \mathbf{R} [\mathbf{v}]_{\times} \mathbf{R}^T \tag{4}$$

where $\mathbf{R} \in SO(3), \mathbf{v} \in \mathbb{R}^3$

The following reference frames are used:

- b-frame denotes the body frame, aligned with the camera axes

- i-frame denotes the Earth-Centred Inertial (ECI) frame

- n-frame denotes the local navigation frame

2 Review of GPS/INS Integration

For some time, it has been known that the errors from GPS and INS are complementary in nature [18]; GPS provides a low-bandwidth, bounded position and velocity solution based on time-of-flight measurements from satellites, whereas an INS provides a high-bandwidth, unbounded position, velocity and attitude solution based on the analytical[2] rotation and integration of specific force and angular velocity measurements. In the simplest form of GPS/INS integration – so-called *uncoupled integration* – the INS position and velocity solution is simply reset with the GPS solution on a periodic basis to prevent the accumulation errors. In this instance, the attitude solution is determined solely by the INS and therefore the yaw solution will continually drift with time.

A better solution to the GPS/INS integration problem can be found by studying how errors propagate in an INS system. It can be shown that attitude errors propagate into velocity errors which, in turn, propagate in the position errors [18]. Therefore, one may take advantage of the correlation between attitude and velocity (or position) errors to recover attitude errors in conjunction with a bounded position and velocity (GPS) solution, when the platform is subjected to particular motion conditions.

When GPS first became available to the navigation community, INS and GPS user equipment were usually sold as individual "black boxes", outputting a navigation solution without access to their internal sensors and algorithms. In the classic implementation of an integrated GPS/INS navigation filter (See, for example, Groves [18] or Farrell [16]), the integration is not directly performed on the navigation estimate, rather the *errors* of the INS states (position, velocity, attitude and sensor errors) are estimated. The estimated errors are then applied to correct the INS solution, either in a feed-forward arrangement[3] or a closed-loop arrangement as shown in Figure 1. To enable estimation using a Kalman Filter, the error dynamics are typically linearised about the current navigation solution.

The arrangement shown in Figure 1 is known as a *loosely-coupled* integration as it is integrating the outputs of two systems rather than federating the GPS and IMU observables into a single estimation engine (known as a *tightly-coupled* arrangement). The tightly-coupled arrangement has several advantages over the loosely-coupled arrangement, chief of which is that GPS observables can be used to slow drift in the INS even when there are insufficient satellites in view to calculate a position solution.

Even a loosely-coupled GPS/INS arrangement offers considerable advantages over GPS and INS alone. Generally, in addition to position, velocity and attitude, inertial sensor errors such as bias are estimated online, leading to slower drift in the INS solution during GPS outages. However, perhaps the most significant advantage is that the attitude (including yaw) has been shown to be observable in a GPS/INS solution [20].

In this paper, the structure of the classic indirect loosely-coupled GPS/INS

---

[2]Analytical (mathematical) rotation is performed when inertial sensors are in a so-called *strapdown configuration* - that is, rigidly mounted to a vehicle. In early INS systems, inertial sensors were physically rotated level and electronically integrated in the so-called *platform configuration*. Platform INS systems have all but disappeared with the passing of time.

[3]The feed-forward arrangement is sometimes used when the GPS solution must be entirely separate from the INS solution, such as integrity monitoring applications.
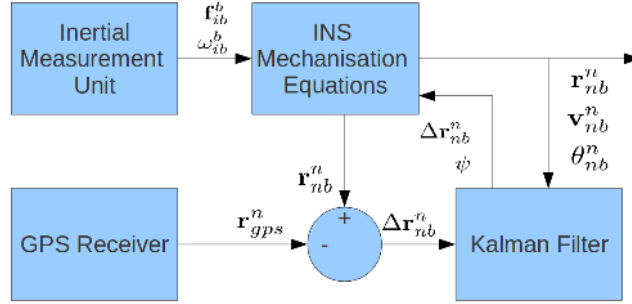
Figure 1: The schematic of a classic loosely-coupled GPS/INS error-state filter. INS errors are estimated using the Kalman Filter and the corrections fed back to the INS in a closed-loop configuration.

system serves as a template for the integration of GPS and Visual Odometry. After deriving the error dynamics, we perform an observability analysis similar to Hong to show the conditions under which attitude may be observed.

3 GPS/Visual Odometry Integration

The intuition behind the integration of GPS and Visual Odometry is that, superficially at least, camera egomotion provides similar observables to an IMU and therefore (in spite of the lack of a vertical reference) may have similar properties when integrated with GPS. Egomotion provides rotation between frames, which, as the time interval between frames approaches zero, is similar to the angular rate measurements provided by the IMU. Similarly, an IMU provides specific force measurements (i.e. acceleration in the body frame), whereas egomotion provides translation (scaled velocity) in the body frame.
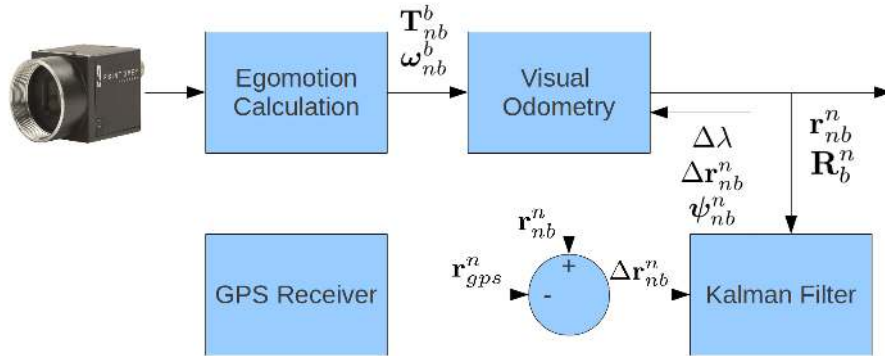

Figure 2: Schematic of the proposed loosely-coupled GPS/VO error-state filter. The egomotion effectively replaces the IMU and the Visual Odometry replaces the INS mechanisation.

The proposed GPS/Visual Odometry (GPS/VO) filter shown in Figure 2 effectively replaces the IMU from Figure 1 with egomotion estimation, the INS

mechanisation equations with the Visual Odometry equations and the INS error-state Kalman Filter with the VO error-state Kalman Filter. In the subsequent sections, we detail the Visual Odometry and derive the linearised Visual Odometry error dynamics suitable for estimation using a Kalman Filter.

3.1 Visual Odometry

There is a vast array of methods that can be used for recovering the relative motion ("egomotion") between two scenes; a recent survey by Bonin-Font [4] provides a taxonomy of methods available in literature. Classic monocular egomotion algorithms estimate the relative rotation $\mathbf{R}_\tau$ and translation $\mathbf{T}_{nb}^b$ (up to a scale factor, λ) between frames [28].

Therefore, we may calculate our attitude at time $t + \tau$:

$$\mathbf{R}_b^n(t + \tau) = \mathbf{R}_b^n(t)\mathbf{R}_\tau \tag{5}$$

And the position of the body with respect to the navigation frame may be updated as:

$$\mathbf{r}_{nb}^n(t + \tau) = \mathbf{r}_{nb}^n(t) + \lambda\mathbf{R}_b^n(t)\mathbf{T}_{nb}^n \tag{6}$$

In all practical integration-based navigation systems, the position and attitude estimates will drift with time without additional external measurements. Furthermore, for an unaided monocular Visual Odometry system, the unobservable scale factor λ will compound drift [12].

Consistent with a loosely-coupled integration approach, no specific egomotion/visual odometry implementation is mandated so long as the required quantities of rotation and translation are available for use by the error estimator.

3.2 Visual Odometry Error Dynamics

In this section, we develop a linearised model of the error dynamics of a Visual Odometry system. To do so, we assume that the motion between frames is small (e.g. a video stream) and can approximate the trajectory of the camera with a smooth (i.e. differentiable) continuous time model consisting of scaled linear velocity of the body with respect to the navigation frame and angular velocities of the body frame with respect to the navigation frame [28]. Note that no platform-specific dynamics (e.g. the non-holonomic constraint) are assumed in the error dynamics, other than that the trajectory be a differentiable function of time.

3.2.1 Attitude Error Dynamics

Since the attitude of a VO system propagates in a similar manner to an INS system, the attitude error dynamics are similar to the INS attitude error dynamics in the ECI frame. Therefore the derivation of the linearised attitude errors is similar to [18].

We define the *attitude error*[4] to be rotation between the true attitude and the estimated attitude:

---

[4]Some authors in the INS literature call this *misalignment*. However, as some authors refer to misalignment as being the non-orthogonality of an inertial sensor triad, we will avoid the use of the term in this paper.

$$\Delta \mathbf{R}_b^n \triangleq \tilde{\mathbf{R}}_b^n \mathbf{R}_n^b \tag{7}$$

If we assume the attitude errors are small, we can create a linear approximation of the attitude error by applying the small angle assumption[5]:

$$[\boldsymbol{\psi}_{nb}^n]_\times \approx [\Delta \mathbf{R}_b^n - \mathbf{I}_3] \tag{8}$$

where $\boldsymbol{\psi}_{nb}^n$ is the small angle attitude error vector [18].
Differentiating Equation 7 yields:

$$\Delta \dot{\mathbf{R}}_b^n = \dot{\tilde{\mathbf{R}}}_b^n \mathbf{R}_n^b + \tilde{\mathbf{R}}_b^n \dot{\mathbf{R}}_n^b \tag{9}$$

Differentiating Equation 8:

$$\left[\dot{\boldsymbol{\psi}}_{nb}^n \right]_\times = \Delta \dot{\mathbf{R}}_b^n \tag{10}$$

Differentiating the rotation matrix yields [18]:

$$\dot{\mathbf{R}}_n^b = \mathbf{R}_n^b \boldsymbol{\Omega}_{bn}^n \tag{11}$$

and

$$\dot{\mathbf{R}}_b^n = \mathbf{R}_b^n \boldsymbol{\Omega}_{nb}^b \tag{12}$$

where $\boldsymbol{\Omega}_{nb}^b = \left[\boldsymbol{\omega}_{nb}^b \right]_\times$ and $\boldsymbol{\Omega}_{bn}^n = [\boldsymbol{\omega}_{bn}^n]_\times$.
Substituting Equations 12, 11 and 10 into 9 gives:

$$\left[\dot{\boldsymbol{\psi}}_{nb}^n \right]_\times = \tilde{\mathbf{R}}_b^n \tilde{\boldsymbol{\Omega}}_{nb}^b \mathbf{R}_n^b + \tilde{\mathbf{R}}_b^n \mathbf{R}_n^b \boldsymbol{\Omega}_{bn}^n \tag{13}$$

Applying the identity from Equation 3:

$$\boldsymbol{\Omega}_{bn}^b = \mathbf{R}_n^b \boldsymbol{\Omega}_{bn}^n \mathbf{R}_b^n \tag{14}$$

Rearranging Equation 14, substituting into 13 and manipulating:

$$\left[\dot{\boldsymbol{\psi}}_{nb}^n \right]_\times = \tilde{\mathbf{R}}_b^n \tilde{\boldsymbol{\Omega}}_{nb}^b \mathbf{R}_n^b + \tilde{\mathbf{R}}_b^n \boldsymbol{\Omega}_{bn}^b \mathbf{R}_n^b \tag{15}$$

$$= \tilde{\mathbf{R}}_b^n \left(\tilde{\boldsymbol{\Omega}}_{nb}^b + \boldsymbol{\Omega}_{bn}^b \right) \mathbf{R}_n^b \tag{16}$$

Noting that $-\boldsymbol{\Omega}_{nb}^b = \boldsymbol{\Omega}_{bn}^b$ and substituting Equation 7 yields:

$$\left[\dot{\boldsymbol{\psi}}_{nb}^n \right]_\times = \tilde{\mathbf{R}}_b^n \left(\tilde{\boldsymbol{\Omega}}_{nb}^b - \boldsymbol{\Omega}_{nb}^b \right) \left(\tilde{\mathbf{R}}_b^n \right)^T \Delta \mathbf{R}_b^n \tag{17}$$

Defining rotation instrument error (corresponding to noise on the incremental attitude between frames) as:

$$\Delta \boldsymbol{\omega}_{nb}^b \triangleq \tilde{\boldsymbol{\omega}}_{nb}^b - \boldsymbol{\omega}_{nb}^b \tag{18}$$

---

[5]The linearised attitude error approximation can be derived from either Euler Angles or Bortz Vector using the small angle approximations $\cos(\theta) \approx 1$ and $\sin(\theta) \approx \theta$

Substituting Equations 18 and 8 into 17 and applying the identity from Equation 4:

$$\left[\dot{\boldsymbol{\psi}}_{nb}^n\right]_\times = \tilde{\mathbf{R}}_b^n \left[\Delta\boldsymbol{\omega}_{nb}^b\right]_\times \left(\tilde{\mathbf{R}}_b^n\right)^T \left([\boldsymbol{\psi}_{nb}^n]_\times + \mathbf{I}_3\right) \tag{19}$$

$$= \left[\tilde{\mathbf{R}}_b^n \Delta\boldsymbol{\omega}_{nb}^b\right]_\times \left([\boldsymbol{\psi}_{nb}^n]_\times + \mathbf{I}_3\right) \tag{20}$$

$$= \left[\tilde{\mathbf{R}}_b^n \Delta\boldsymbol{\omega}_{nb}^b\right]_\times + \left[\tilde{\mathbf{R}}_b^n \Delta\boldsymbol{\omega}_{nb}^b\right]_\times [\boldsymbol{\psi}_{nb}^n]_\times \tag{21}$$

Assuming that the product of error terms is small, the linearised attitude error dynamics may be written as:

$$\dot{\boldsymbol{\psi}}_{nb}^n = \tilde{\mathbf{R}}_b^n \Delta\boldsymbol{\omega}_{nb}^b \tag{22}$$

3.2.2   Position Error Dynamics

The velocity of the platform in the navigation frame is dependant on the scale and the translation

$$\mathbf{v}_{nb}^n = \lambda \mathbf{T}_{nb}^n \tag{23}$$

Similarly, the estimated velocity from the estimated scale and measured translation from egomotion may be written as

$$\tilde{\mathbf{v}}_{nb}^n = \tilde{\lambda} \tilde{\mathbf{T}}_{nb}^n \tag{24}$$

.

The derivative of the position in the navigation frame is the velocity

$$\dot{\mathbf{r}}_{nb}^n = \mathbf{v}_{nb}^n \tag{25}$$

And hence the estimated position derivative is

$$\dot{\tilde{\mathbf{r}}}_{nb}^n = \tilde{\mathbf{v}}_{nb}^n \tag{26}$$

We define the error in position, error in scale and error in translation respectively as:

$$\Delta\mathbf{r}_{nb}^n \triangleq \tilde{\mathbf{r}}_{nb}^n - \mathbf{r}_{nb}^n \tag{27}$$

$$\Delta\lambda \triangleq \tilde{\lambda} - \lambda \tag{28}$$

$$\Delta\mathbf{T}_{nb}^b \triangleq \tilde{\mathbf{T}}_{nb}^b - \mathbf{T}_{nb}^b \tag{29}$$

Differentiating Equation 27 yields

$$\Delta\dot{\mathbf{r}}_{nb}^n = \dot{\tilde{\mathbf{r}}}_{nb}^n - \dot{\mathbf{r}}_{nb}^n \tag{30}$$

$$= \tilde{\mathbf{v}}_{nb}^n - \mathbf{v}_{nb}^n \tag{31}$$

Defining $\boldsymbol{\Psi}_{nb}^n \triangleq [\boldsymbol{\psi}_{nb}^n]_\times$ and substituting Equations 7, 8, 23 and 24 into 30 gives

$$\Delta \dot{\mathbf{r}}_{nb}^{n} \quad = \quad \tilde{\lambda}\tilde{\mathbf{T}}_{nb}^{n} - \lambda \mathbf{T}_{nb}^{n} \tag{32}$$

$$= \quad \tilde{\lambda}\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b} - \lambda \mathbf{R}_{b}^{n}\mathbf{T}_{nb}^{b} \tag{33}$$

$$= \quad \tilde{\lambda}\left[\mathbf{I}_{3} + \boldsymbol{\Psi}_{nb}^{n}\right]\mathbf{R}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b} - \lambda \mathbf{R}_{b}^{n}\mathbf{T}_{nb}^{b} \tag{34}$$

$$= \quad \mathbf{R}_{b}^{n}\left(\tilde{\lambda}\tilde{\mathbf{T}}_{nb}^{b} - \lambda \mathbf{T}_{nb}^{b}\right) + \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\mathbf{R}_{b}^{n}\mathbf{T}_{nb}^{b} \tag{35}$$

Expanding the term $\left(\tilde{\lambda}\tilde{\mathbf{T}}_{nb}^{b} - \lambda \mathbf{T}_{nb}^{b}\right)$ and substituting Equation 27:

$$\left(\tilde{\lambda}\tilde{\mathbf{T}}_{nb}^{b} - \lambda \mathbf{T}_{nb}^{b}\right) \quad = \quad \tilde{\lambda}\tilde{\mathbf{T}}_{nb}^{b} - \left(\tilde{\lambda} - \Delta\lambda\right)\mathbf{T}_{nb}^{b} \tag{36}$$

$$= \quad \tilde{\lambda}\Delta\mathbf{T}_{nb}^{b} + \Delta\lambda\tilde{\mathbf{T}}_{nb}^{b} - \Delta\lambda\Delta\mathbf{T}_{nb}^{b} \tag{37}$$

And since the product of two error terms are small:

$$\left(\tilde{\lambda}\tilde{\mathbf{T}}_{nb}^{b} - \lambda \mathbf{T}_{nb}^{b}\right) \approx \tilde{\lambda}\Delta\mathbf{T}_{nb}^{b} + \Delta\lambda\tilde{\mathbf{T}}_{nb}^{b} \tag{38}$$

Expanding the term $\tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\mathbf{R}_{b}^{n}\mathbf{T}_{nb}^{b}$, substituting Equations 7, 8 and 27

$$\tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\mathbf{R}_{b}^{n}\mathbf{T}_{nb}^{b} \quad = \quad \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\mathbf{R}_{b}^{n}\mathbf{T}_{nb}^{b} \tag{39}$$

$$= \quad \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\tilde{\mathbf{R}}_{b}^{n}\Delta\mathbf{R}_{n}^{b}\mathbf{T}_{nb}^{b} \tag{40}$$

$$= \quad \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\tilde{\mathbf{R}}_{b}^{n}\left[\mathbf{I}_{3} - \boldsymbol{\Psi}_{nb}^{b}\right]\mathbf{T}_{nb}^{b} \tag{41}$$

$$= \quad \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b} - \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\tilde{\mathbf{R}}_{b}^{n}\Delta\mathbf{T}_{nb}^{b} - \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\tilde{\mathbf{R}}_{b}^{n}\boldsymbol{\Psi}_{nb}^{b}\mathbf{T}_{nb}^{b} \tag{42}$$

And since the product of error terms are small, we may approximate:

$$\tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\mathbf{R}_{b}^{n}\mathbf{T}_{nb}^{b} \approx \tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b} \tag{43}$$

Using the anti-symmetric property of the cross-product from Equation 2:

$$\tilde{\lambda}\boldsymbol{\Psi}_{nb}^{n}\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b} = -\tilde{\lambda}\left[\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b}\right]_{\times}\boldsymbol{\psi}_{nb}^{n} \tag{44}$$

Substituting 44, 38, into 32:

$$\Delta\dot{\mathbf{r}}_{nb}^{n} = \mathbf{R}_{b}^{n}\left(\tilde{\lambda}\Delta\mathbf{T}_{nb}^{b} + \Delta\lambda\tilde{\mathbf{T}}_{nb}^{b}\right) - \tilde{\lambda}\left[\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b}\right]_{\times}\boldsymbol{\psi}_{nb}^{n} \tag{45}$$

Manipulating and removing the product of small errors yields:

$$\Delta\dot{\mathbf{r}}_{nb}^{n} = \tilde{\mathbf{R}}_{b}^{n}\left(\tilde{\lambda}\Delta\mathbf{T}_{nb}^{b} + \Delta\lambda\tilde{\mathbf{T}}_{nb}^{b}\right) - \tilde{\lambda}\left[\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b}\right]_{\times}\boldsymbol{\psi}_{nb}^{n} \tag{46}$$

Which, when written in terms of the error states and measurement errors is:

$$\Delta\dot{\mathbf{r}}_{nb}^{n} = \tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b}\Delta\lambda - \tilde{\lambda}\left[\tilde{\mathbf{R}}_{b}^{n}\tilde{\mathbf{T}}_{nb}^{b}\right]_{\times}\boldsymbol{\psi}_{nb}^{n} + \tilde{\lambda}\tilde{\mathbf{R}}_{b}^{n}\Delta\mathbf{T}_{nb}^{b} \tag{47}$$

3.2.3 Scale Factor Error Dynamics

The translation between two frames is recoverable only up to a scale factor [28]. Whilst the scale factor may remain constant when a constant set of features remain in view, the scale factor will inevitably drift once those features fall from the field of view [12]. In the absence of further information, for a physical platform with appreciable inertia it is reasonable to assume that drift can be adequately modelled with a Gaussian random walk:

$$\Delta\dot{\lambda} = \nu \tag{48}$$

where $\nu \sim N\left(0, \sigma_\lambda\right)$.

The choice of σ_λ corresponds to magnitude of the expected drift. If the translation between frames is a unit vector, then ν corresponds to random accelerations of the platform.

3.2.4 Navigation Error State Model

From Equations 22, 47 and 48 we can now construct the system error dynamics in terms of the error states and measurement errors:

$$
\begin{bmatrix} \Delta\dot{\lambda} \\ \Delta\dot{\mathbf{r}}_{nb}^n \\ \dot{\boldsymbol{\psi}}_{nb}^n \end{bmatrix}
=
\begin{bmatrix}
0 & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\
\tilde{\mathbf{R}}_b^n \tilde{\mathbf{T}}_{nb}^b & \mathbf{0}_{3\times3} & -\tilde{\lambda}\left[\tilde{\mathbf{R}}_b^n \tilde{\mathbf{T}}_{nb}^b\right]_\times \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3}
\end{bmatrix}
\begin{bmatrix} \Delta\lambda \\ \Delta\mathbf{r}_{nb}^n \\ \boldsymbol{\psi}_{nb}^n \end{bmatrix}
+
\begin{bmatrix}
1 & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\
\mathbf{0}_{3\times1} & \tilde{\lambda}\tilde{\mathbf{R}}_b^n & \mathbf{0}_{3\times3} \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \tilde{\mathbf{R}}_b^n
\end{bmatrix}
\begin{bmatrix} \nu \\ \Delta\mathbf{T}_{nb}^b \\ \Delta\boldsymbol{\omega}_{nb}^b \end{bmatrix}
\tag{49}
$$

Neglecting the lever arm between the camera and GPS antenna, errors in position may be measured by the difference between the GPS and VO solutions:

$$\Delta\mathbf{r}_{nb}^n = \mathbf{r}_{nb,gps}^n - \mathbf{r}_{nb,vo}^n - \mathbf{v}(t) \tag{50}$$

where $\mathbf{v}(t)$ is the GPS position measurement noise, modelled as white and Gaussian with covariance $\mathbf{R_v}$.

Equation 50 may be re-written as:

$$
\mathbf{r}_{nb,gps}^n - \mathbf{r}_{nb,vo}^n =
\begin{bmatrix} \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times3} \end{bmatrix}
\begin{bmatrix} \Delta\lambda \\ \Delta\mathbf{r}_{nb}^n \\ \boldsymbol{\psi}_{nb}^n \end{bmatrix}
+ \mathbf{v}(t)
\tag{51}
$$

Equations 49 and 51 form a Linear Time-Varying (LTV) system in the form:

$$
\begin{aligned}
\dot{\mathbf{x}}(t) &= \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{w}(t) \tag{52} \\
\mathbf{z}(t) &= \mathbf{H}(t)\mathbf{x}(t) + \mathbf{v}(t) \tag{53}
\end{aligned}
$$

Since the error model is given in continuous time, a discrete-time equivalent model will need to be realised for implementation using a Kalman Filter.

3.3 Discrete-Time Equivalent Model

If the continuous time state transition matrix $\mathbf{F}(t)$ is approximately constant between image frames, it can be shown [16] that the equivalent discrete-time state transition matrix $\mathbf{A}(k)$ is given by:

$$\mathbf{A}(k) = \exp\left(\mathbf{F}(t)T\right) \tag{54}$$

where $T = t_k - t_{k-1}$ is the sampling period. The matrix exponential may be calculated numerically, or approximated using truncated Taylor-series expansion, trading accuracy for computation time.

3.4 Closed Loop Correction

The linearised error dynamics of the system assume that the magnitude of the error states are small. In particular, the attitude error dynamics rely on the small angle assumption that may be violated if the VO solution is allowed to drift over time. To prevent growth in the error states, the error estimates from the Kalman Filter may be used to correct the Visual Odometry solution.

Corrections to the VO solution are normally applied after a measurement update of the Kalman Filter. The position and attitude closed loop correction equations are the same as for a loosely-coupled GPS/INS solution [18], and the scale factor error correction follows from the definition in Equation 27. That is, the corrected values $\hat{\mathbf{R}}_b^n$, $\hat{\mathbf{r}}_{nb}^n$ and $\hat{\lambda}$ can be determined from the past Visual Odometry values $\tilde{\mathbf{R}}_b^n$, $\tilde{\mathbf{r}}_{nb}^n$ and $\tilde{\lambda}$ as follows:

$$\hat{\mathbf{R}}_b^n = \left(\Delta\mathbf{R}_b^n\right)^T \tilde{\mathbf{R}}_b^n \tag{55}$$
$$\hat{\mathbf{r}}_{nb}^n = \tilde{\mathbf{r}}_{nb}^n - \Delta\mathbf{r}_{nb}^n \tag{56}$$
$$\hat{\lambda} = \tilde{\lambda} - \Delta\lambda \tag{57}$$

where $\Delta\mathbf{r}_{nb}^n$ and $\Delta\lambda$ are obtained directly from the Kalman Filter. The attitude correction $\Delta\mathbf{R}_b^n$ is the direction cosine matrix formed from the attitude error $\boldsymbol{\psi}_{nb}^n$ estimated by the Kalman Filter using:

$$\Delta\mathbf{R}_b^n = \exp\left(\left[\boldsymbol{\psi}_{nb}^n\right]_\times\right) \tag{58}$$

Once the corrections are applied, the error states in the Kalman Filter are set to zero but the covariance matrix remains unaltered as only the mean rather than the uncertainty is changed.

4 Observability of the GPS/VO Filter

In the following section, we analyse the observability of the GPS/VO solution using a similar methodology to that used by Hong [20] when analysing GPS/INS observability. The analysis presented assumes a noise-free system, which is clearly impractical but is still of theoretical use – if states are not observable in the noise-free case then they will not be observable in the presence of noise.

4.1 Observability Definition

For this analysis, we adopt the definition for observability of a Linear Time-Varying (LTV) system from [11]. Consider an LTV in the form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) \tag{59}$$

$$\mathbf{z}(t) = \mathbf{H}(t)\mathbf{x}(t) \tag{60}$$

where $\mathbf{F}(t)$ and $\mathbf{H}(t)$ are continuous functions of time defined over the domain $[-\infty, \infty]$ and are $n-1$ times continually differentiable, where n is the length of the state vector. The LTV system in Equation 59 is observable at t_0 if there exists a finite $t_1 > t_0$ such that

$$rank \begin{bmatrix} \mathbf{N}_0(t_1) \\ \mathbf{N}_1(t_1) \\ \vdots \\ \mathbf{N}_{n-1}(t_1) \end{bmatrix} = n \tag{61}$$

where

$$\mathbf{N}_0 = \mathbf{H}(t) \tag{62}$$

$$\mathbf{N}_{m+1}(t) = \mathbf{N}_m(t)\mathbf{F}(t) + \frac{d}{dt}\mathbf{N}_m(t) \tag{63}$$

$$m = 1, 2, \ldots, n-1 \tag{64}$$

Similarly, a Linear Time-Invariant (LTI) system is observable for every initial time if and only if the rank of the observability matrix

$$\mathbf{O} = \begin{bmatrix} \mathbf{H}^T & (\mathbf{HF})^T & (\mathbf{HF}^2)^T & \ldots & (\mathbf{HF}^{n-1})^T \end{bmatrix}^T \tag{65}$$

is of rank n.

4.2 Observability Analysis

To simplify the observability analysis, we substitute $\tilde{\mathbf{R}}_b^n \tilde{\mathbf{T}}_{nb}^b = \tilde{\mathbf{T}}_{nb}^n$ into $\mathbf{F}(t)$ from Equation 49. That is, we express the translation in the navigation frame rather than the body frame. Substituting the LTV system in Equation 49 into the observability definition from Equation 61 yields the following observability matrix:

$$\mathbf{O} = \begin{bmatrix} \mathbf{0}_{3\times 1} & \mathbf{I}_3 & \mathbf{0}_{3\times 3} \\ \tilde{\mathbf{T}}_{nb}^n & \mathbf{0}_{3\times 3} & -\tilde{\lambda}\left[\tilde{\mathbf{T}}_{nb}^n\right]_\times \\ \frac{d}{dt}\tilde{\mathbf{T}}_{nb}^n & \mathbf{0}_{3\times 3} & \frac{d}{dt} - \tilde{\lambda}\left[\tilde{\mathbf{T}}_{nb}^n\right]_\times \\ \vdots & \vdots & \vdots \\ \frac{d^6}{dt^6}\tilde{\mathbf{T}}_{nb}^n & \mathbf{0}_{3\times 3} & \frac{d^6}{dt^6} - \tilde{\lambda}\left[\tilde{\mathbf{T}}_{nb}^n\right]_\times \end{bmatrix} \tag{66}$$

To gain an understanding of the physical implications of the general observation matrix in Equation 66, it is worth examining two specials cases - constant velocity in the navigation frame, and constant acceleration in the navigation frame.

4.2.1 Observability Under Constant Velocity in the Navigation Frame

Under constant velocity in the navigation frame (i.e. $\frac{d^m}{dt^m}\tilde{\lambda}\tilde{\mathbf{T}}_{nb}^n = \mathbf{0}_{3\times 1}, \forall m \geq 1$), the observability matrix becomes, after removing rows with all zero members:

$$\mathbf{O}_v = \left[\begin{array}{ccc} \mathbf{0}_{3\times 1} & \mathbf{I}_3 & \mathbf{0}_{3\times 3} \\ \tilde{\mathbf{T}}_{nb}^n & \mathbf{0}_{3\times 3} & -\tilde{\lambda}\left[\tilde{\mathbf{T}}_{nb}^n\right]_\times \end{array}\right] \tag{67}$$

When the velocity is zero, the observability matrix is of rank 3 and it is clear that only the position errors are observable. When the velocity is non-zero, the rank of the observability matrix is 6, with the nullspace given by:

$$null\left(\mathbf{O}_v\right) = \begin{bmatrix} \mathbf{0}_{4\times 1} \\ T_{nb,x}^n/T_{nb,z}^n \\ T_{nb,y}^n/T_{nb,z}^n \\ 1 \end{bmatrix} \tag{68}$$

Note that the scale factor and position errors are always observable and the non-observable component of the attitude error lies on the direction of the translation vector.

4.2.2 Observability Under Constant Acceleration in the Navigation Frame

Under constant acceleration in the navigation frame (i.e. $\frac{d^m}{dt^m}\tilde{\lambda}\tilde{\mathbf{T}}_{nb}^n = \mathbf{0}_{3\times 1}, \forall m \geq 2$), the observability matrix becomes:

$$\mathbf{O}_a = \left[\begin{array}{ccc} \mathbf{0}_{3\times 1} & \mathbf{I}_3 & \mathbf{0}_{3\times 3} \\ \tilde{\mathbf{T}}_{nb}^n & \mathbf{0}_{3\times 3} & -\tilde{\lambda}\left[\tilde{\mathbf{T}}_{nb}^n\right]_\times \\ \frac{d}{dt}\left(\tilde{\mathbf{T}}_{nb}^n\right) & \mathbf{0}_{3\times 3} & \frac{d}{dt} - \tilde{\lambda}\left[\tilde{\mathbf{T}}_{nb}^n\right]_\times \end{array}\right] \tag{69}$$

which is of full rank, except where:

1. $\| \tilde{\mathbf{T}}_{nb}^n \| = 0$ (i.e. zero velocity case)

2. $\| \dot{\tilde{\mathbf{T}}}_{nb}^n \| = 0$ (i.e. constant velocity case)

3. $\tilde{\mathbf{T}}_{nb}^n = k\dot{\tilde{\mathbf{T}}}_{nb}^n, k \neq 0$ (i.e. acceleration parallel to velocity vector)

When the acceleration is parallel to the velocity vector, the nullspace basis is

$$null\left(\mathbf{O}_a\right) = \begin{bmatrix} \mathbf{0}_{4\times 1} \\ T_{nb,x}^n/T_{nb,z}^n \\ T_{nb,y}^n/T_{nb,z}^n \\ 1 \end{bmatrix} \tag{70}$$

which is the same as for the constant velocity case.

Therefore, all error states (scale, position and attitude) are observable when the acceleration is non-parallel to translation in the navigation frame. Note that as long as both the translation and acceleration are not zero, the magnitude of the acceleration or translation plays no role in determining whether attitude errors are fully observable in a noise-free system. Of course, this analysis contains no consideration as to whether the states are stochastically observable in the presence of noise.

5 Numerical Observability Analysis

To demonstrate some of the properties determined from the analytical observability analysis, we have performed a number of numerical simulations. In this analysis, we present the following scenarios, generated using the Aerospace Blockset in Simulink:

1. Constant velocity along the x-axis, including roll about the body frame

2. Constant velocity along the x-axis, including pitch about the body frame

3. Acceleration along the x-axis, with the initial velocity vector along the x-axis

4. Acceleration along the y-axis, with the initial velocity vector along the x-axis

   In Scenarios 1-3, according to analytic observability analysis, at least one of the attitude parameters will not converge to the true value and will slowly diverge with time. In Scenario 4, all attitude components should be observable when the acceleration is applied.

   In all the scenarios presented, both the egomotion integration and the GPS measurements are performed at 20Hz[6]. White Gaussian noise is added to the body rates, velocity and GPS measurements with standard deviations of $3°/s$, $1m/s$ and $1m$ respectively. An initial random error is added to each of the states with a standard deviation of $10m$ in position and $15°$ in attitude.

5.0.3 Constant Velocity with Roll Manoeuvre

In the first scenario, the platform travels at constant velocity along the x-axis in the navigation frame and performs a rolling manoeuvre without altering the trajectory of the platform. That is, although the velocity in the body frame changes with the rolling manoeuvre, the velocity in the navigation frame is unchanged. The roll angle during manoeuvre is shown in Figure 3.
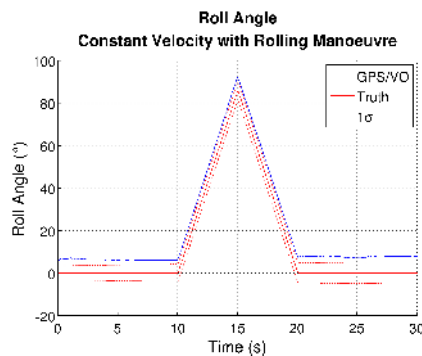
Figure 3: Roll angle during Scenario 1 (Constant velocity with rolling manoeuvre)

---

[6]For example, the NovAtel OEMV-1 GPS receiver is capable of 20Hz

The attitude errors under this manoeuvre are shown in Figure 4. Note that in under this scenario, pitch and yaw quickly converge, whereas the the roll error continually drifts with time. The errors exhibited are consistent with the analysis that shows that the component of attitude along the directory of velocity is unobservable.
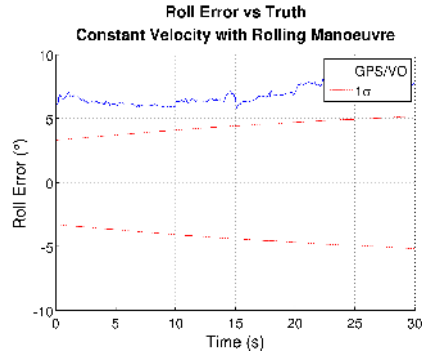
(a) Roll Error
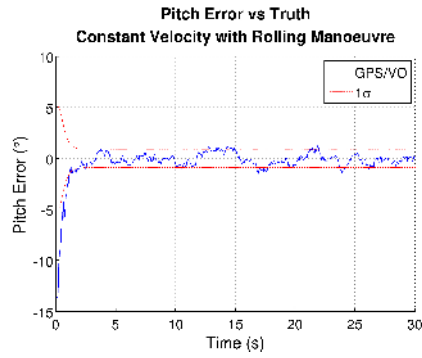
(b) Pitch Error
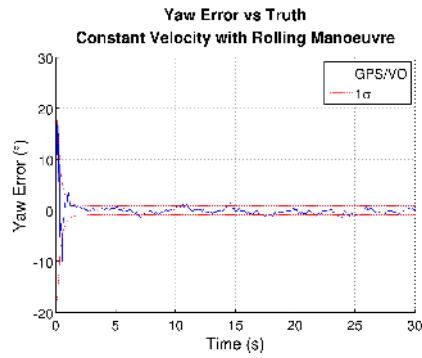
(c) Yaw Error

Figure 4: Attitude error for Scenario 1 (Constant velocity with rolling manoeuvre)

5.0.4 Constant Velocity with Pitch Manoeuvre

In the second scenario, the trajectory of the vehicle remains constant along the x-axis in the navigation frame, with a pitching manoeuvre (shown in Figure 5) performed to alter the attitude of the platform. The attitude errors for this scenario are shown in Figure 6.
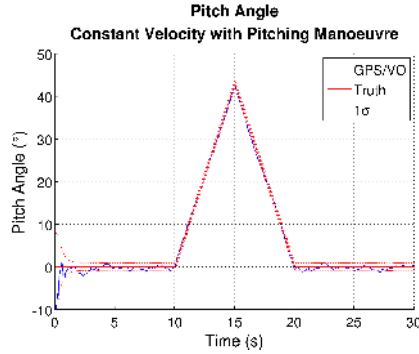
Figure 5: Pitch angle during Scenario 2 (Constant velocity with pitching manoeuvre). Note that the pitch is tracked by the GPS/VO filter throughout the manoeuvre

Similar to Scenario 1, the roll error continually grows during the sequence without the requisite manoeuvres to make the third component of attitude observable.

5.0.5 Acceleration in the X-Axis

In the third scenario, the attitude is held constant, and an acceleration manoeuvre is performed along the x-axis in the navigation frame. That is, the speed of the platform is altered, but not the attitude or direction. The velocity of the platform along the x-axis in the navigation frame is shown in Figure 7.

During the entire manoeuvre, it is clear from Figure 8 that the scale factor estimate is consistent with the speed of the platform (since the translation is normalised), and that the errors are always observable, with the large initial error rapidly converging to the true value.

However, similar to Scenarios 1 and 2, there is no observability of roll as evidenced in Figure 9, with roll diverging with time.

5.0.6 Acceleration in the Y-Axis

In the fourth scenario, the platform starts with constant speed in the x-axis then accelerates in the y-axis without changing the attitude of the platform. Whilst this manoeuvre is unrealistic for many platforms, it is useful to demonstrate the observability of the filter.

The plan view of the displacement is shown in Figure 10, with the velocity shown in Figure 12. Note there are three periods of acceleration in the y-axis – between $5 - 10s$, $15 - 20s$ and $25 - 30s$. The normalised translation in the body frame (Figure 11(a)) is not constant as a result.
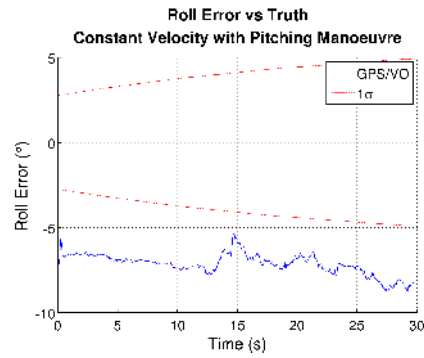
(a) Roll Error
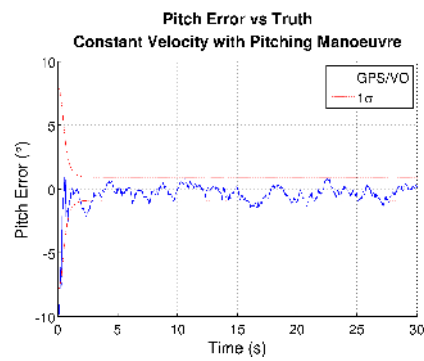
(b) Pitch Error
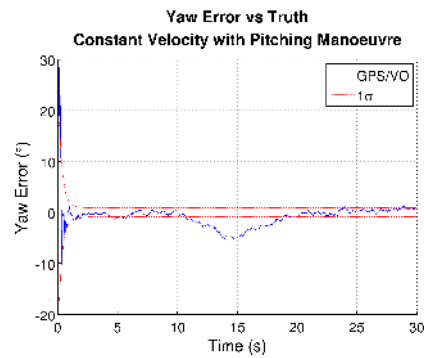
(c) Yaw Error

Figure 6: Attitude error for Scenario 2 (Constant velocity with pitching manoeuvre)

The effect of the acceleration periods are immediately apparent in the roll error in Figure 13(a). Between $5 - 10s$, it can be seen that the roll error (and standard deviation) converges towards truth, then remains constant until the acceleration period between $15 - 20s$. There is a further change between $25 - 30s$, but the error is already well within its 1σ bounds. The roll error can be seen to drift in the final $10s$ of the sequence.
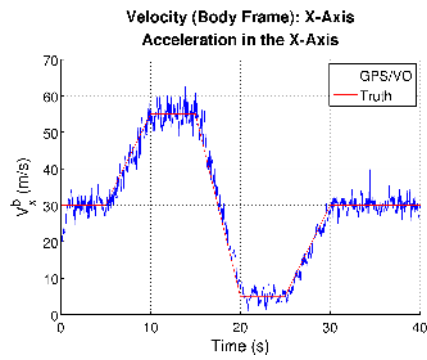
Figure 7: Velocity along the x-axis for Scenario 3 (Acceleration along the x-axis)
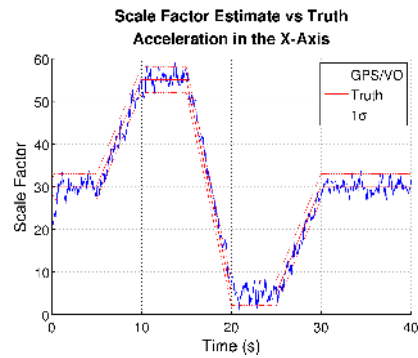
(a) Scale factor
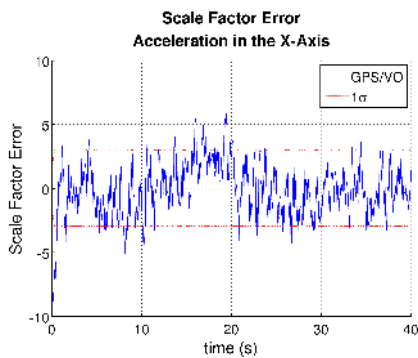
(b) Scale factor error

Figure 8: Scale factor estimate and error (speed) for Scenario 3 (Acceleration along the x-axis). Note that the initial error of approximately $10m/s$ rapidly converges.

Like scenarios 1-3, throughout the whole sequence, pitch error (Figure 13(b)), yaw error (Figure 13(c)) and scale factor (Figure 14) immediately converge and remain bounded for the entire scenario.
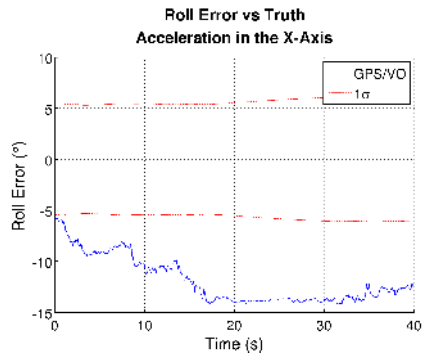
(a) Roll Error
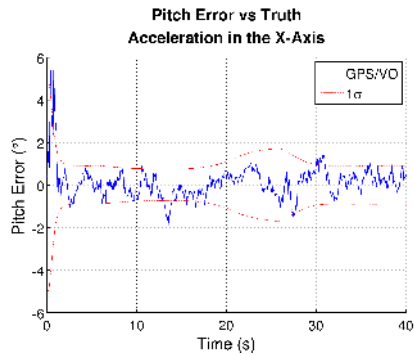
(b) Pitch Error
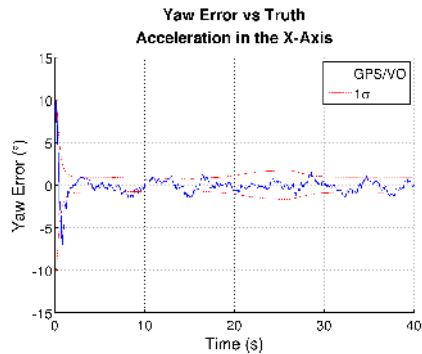
(c) Yaw Error

Figure 9: Attitude Error for Scenario 3 (Acceleration along the x-axis). Note that roll never converges, but roll and yaw rapidly converge

6   Experimental Validation on Real Data

To demonstrate the feasibility of the proposed GPS/VO algorithm on real data, we present the results from two sequences of data collected during a flight around South-East Queensland, Australia, using a downwards-looking camera mounted to a Cessna 172. In this section, we describe the data collection system, the two data sequences, the egomotion front-end and the experimental results.
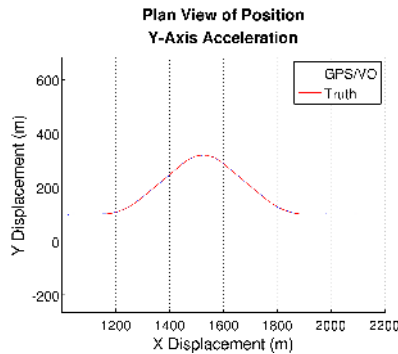
Figure 10: Plan view of Scenario 4 (Acceleration in the Y-Axis)

6.1 The Airborne Systems Laboratory (ASL)

The Airborne Systems Laboratory (ASL) is a Cessna 172 owned and operated by ARCAA, modified to carry a range of sensors and equipment and used for a variety of research activities [17]. As standard, the ASL carries a flight control computer capable of commanding the on-board autopilot and a communication payload consisting of an ISM-band radiomodem, a 3G cell modem and an Iridium modem. The position and attitude truth is provided by a NovAtel SPAN, which computes a tightly-coupled GPS/INS solution from measurements taken by a NovAtel OEMV-3 GNSS receiver and an iMAR-FSAS IMU. The claimed accuracy of the SPAN system is $0.015°(1\sigma)$ in pitch and roll and $0.06°(1\sigma)$ in heading at a rate of 100Hz.

In addition to the standard payload, the ASL may carry a client payload that can change between missions (Figure 15(b)). For the experiment detailed in this paper, the ASL was fitted with the Image Capture Payload, as described in the next section.

6.2 The ASL Image Capture Payload

The purpose of the Image Capture Payload (ICP) is to record raw, uncompressed images[7] from two IIDC cameras at a rate of up to 30Hz. The IIDC cameras are externally triggered independently by a microcontroller synchronised to the GPS 1 Pulse-Per-Second (1PPS) signal at the desired rate, allowing the captured images to be precisely timestamped. The images are recorded to multiple hard disks on a commodity x86 computer running Ubuntu with the open-source camera capture program, Videography. The x86 computer also records the 1PPS timestamps via the parallel port interrupt and records absolute time using the on-board NovAtel OEMV-1 receiver (separate to the on-board truth system).

The cameras themselves are attached to a bracket at the tie-down point on the right wing of the aircraft (Figure 16), with one forwards-looking camera[8] and one downwards looking camera. Further details on the ASL, the image capture payload and truth system are detailed by Greer [17].

---

[7]Uncompressed images were required for sense-and-avoid research as aircraft targets are often below the noise floor of the image [25].

[8]The forwards-looking camera is unused in this paper

21
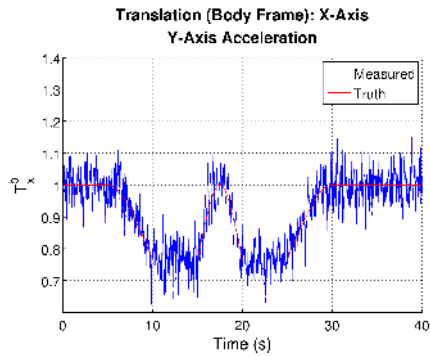
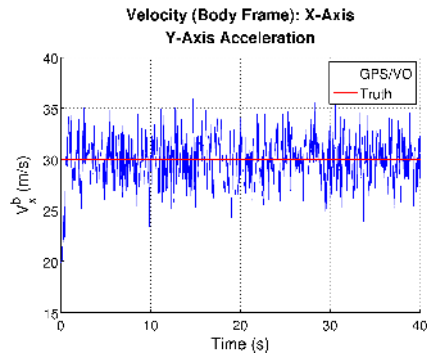(a) X-axis Translation
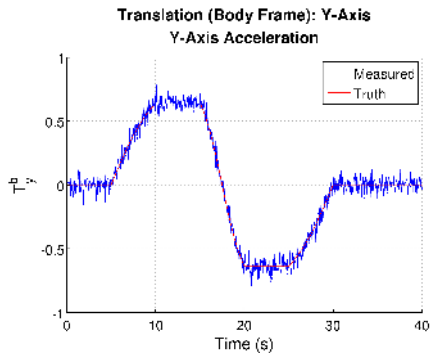
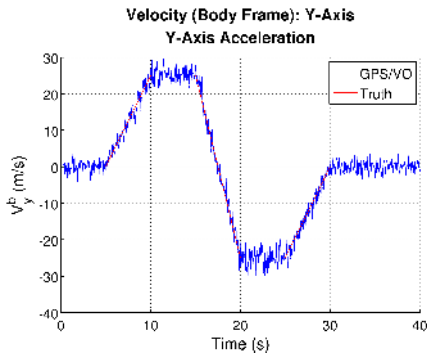(b) Y-axis Translation
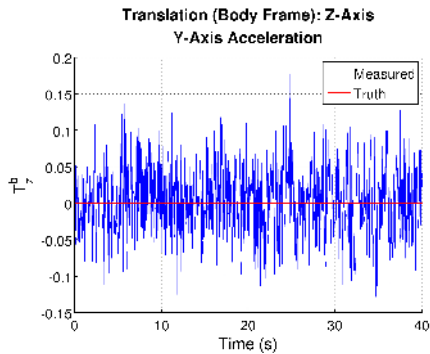
(c) Z-axis Translation

Figure 11: Translation in the b-frame for Scenario 4 (Acceleration in the Y-Axis). Note that although the translation in the x-axis varies, the velocity in this axis remains unchanged.

6.3 The ASL Dataset

In this paper, we present the results from two representative sequences from the ASL dataset, each approximately 8-9 minutes in duration and representing a different part of the flight regime. The "Archerfield Approach Sequence" consists of the transit over the western suburbs of Brisbane, manoeuvring onto final

22

(a) X-axis velocity

(b) Y-axis velocity
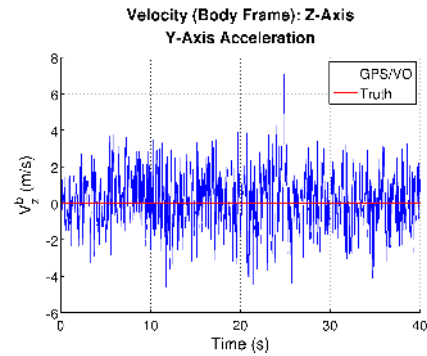
(c) Z-axis velocity

Figure 12: Velocity in the body frame for Scenario 4 (Acceleration in the Y-Axis). The GPS/VO velocity is the product of estimated translation and scale factor.

approach at Archerfield Airport. The "En-Route Sequence" consists mostly of comparatively high altitude straight and level flight between Archerfield Airport and Watt's Bridge Airfield.

A more detailed overview of the entire ASL dataset can be found in [14].
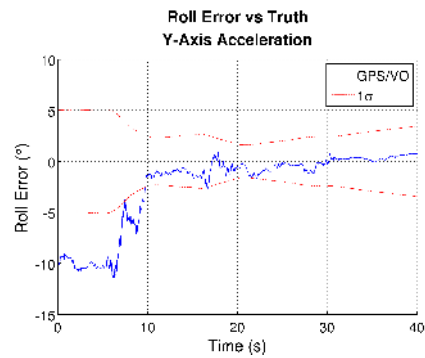
(a) Roll error
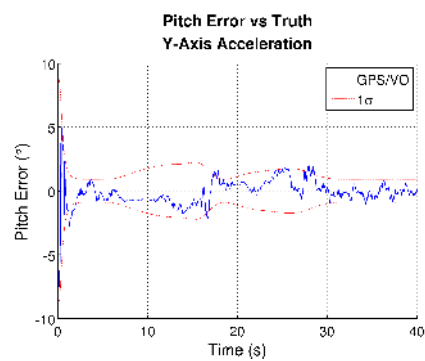
(b) Pitch error
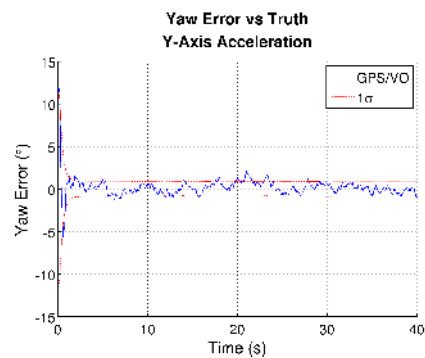
(c) Yaw error

Figure 13: Attitude Error Plots for Scenario 4 (Acceleration in the Y-Axis). Note the convergence in roll during the time periods where acceleration in the y-axis is applied.

6.3.1 Archerfield Approach Sequence

The Archerfield Approach Sequence is a sequence of 8 minutes, 20 seconds starting west of Brisbane near Samford Valley (Figure 17), using the designated southbound approach path towards Archerfield Airport, initially over hilly terrain before overflying the semi-rural (Figure 18(a)) western suburbs and subur-
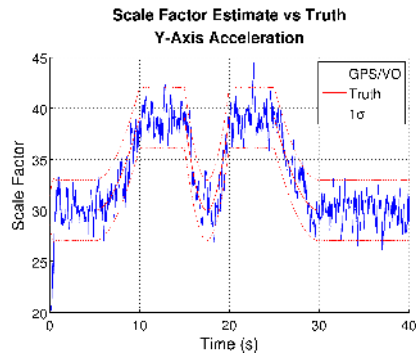
(a) Scale factor
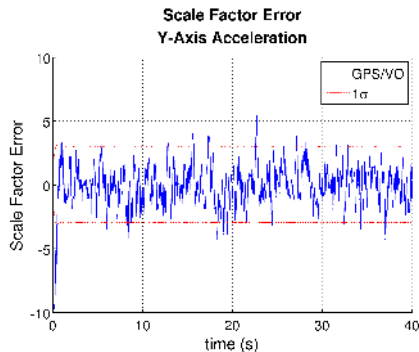
(b) Scale factor error

Figure 14: Scale factor for Scenario 4 (Acceleration in the Y-Axis). Note the near immediate convergence from an error of -10.

ban south-western suburbs at an altitude of approximately 500m above Mean Sea Level (AMSL). The sequence includes the turn from crosswind onto final approach and descent from circuit height to only a few tens of metres above the perimeter fence at Archerfield Airport (Figure 18(b)).

6.3.2 En-Route Sequence

The 9 minute En-Route Sequence represents a different flight regime to the Archerfield Approach Sequence. Whereas relatively aggressive manoeuvring is performed approaching the circuit at Archerfield Airport, the En-Route Sequence exhibits mostly straight and level flight at a near constant altitude of 1200m AMSL. The sequence commences near Cedar Creek (Figure 19), tracking north-west towards Somerset Dam. A gentle turn is performed during the dataset, changing the heading from west to north-west. At the end of the dataset, an aggressive turn is performed, marking the start of the descent to Watt's Bridge Airfield.

Since there are few manoeuvres performed in this sequence, we expect that there will be few opportunities for roll to converge towards truth during the straight and level sequence. Furthermore, the higher altitude results in smaller disparity between frames, making the translation estimates noisier compared to

(a) Airborne Systems Laboratory (ASL) Cessna 172

(b) ASL Payload rack, including the NovAtel SPAN rack and client equipment

Figure 15: The ASL and Data Collection System

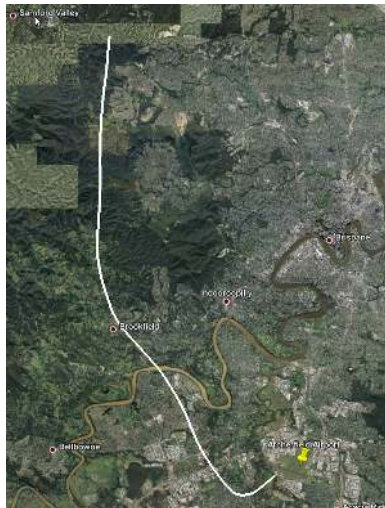Figure 16: Cameras mounted to the right wing strut of the ASL

Figure 17: Flight path taken during the Archerfield Approach Sequence, tracking south from near Samford Valley through to Archerfield Airport via Brookfield.

(a) Dataset image over the western suburbs of Brisbane, near Brookfield.

(b) Dataset image at the perimeter of Archerfield Airport on final approach.

Figure 18: Example images from the Archerfield Approach Sequence

the lower-altitude Archerfield Approach Dataset.

Figure 19: Flight path taken during the En-Route Sequence, commencing near Cedar Creek.

The visual appearance of the En-Route Sequence significantly varies across the dataset. Whilst much of the sequence is dominated by bushland over the foothills of Brisbane (Figure 20(b)), the latter part of the dataset consists of grazing land under the influence of severe drought (Figure 20(b)). Note that the grazing land near Crossdale has few salient features for feature-based motion estimation which, as will be shown in the subsequent results, have a noticeable impact on the accuracy of egomotion estimation.

6.4 Homography-Based Egomotion Estimation

The egomotion calculated during the experiments is an unsophisticated two-frame solution consisting of the following steps:

1. Image keypoints are detected using SURF [3].

27

(a) Typical dataset image of the bushland over the foothills of Brisbane captured during the En-Route Sequence
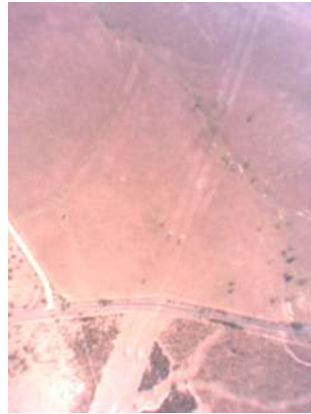
(b) Dataset image of dry grazing land towards the end of the En-Route Sequence

Figure 20: Example images from the En-Route Sequence

2. BEST Descriptors [7] for each keypoint are calculated and matched.

3. Gross errors (e.g. a feature point in one frame matching multiple points in the next or vice-versa) are discarded.

4. The homography matrix between frames is estimated using RANSAC.

5. The homography matrix is decomposed into a rotation matrix, translation and normal vector using the procedure in [28].

The egomotion estimation front-end is implemented in C++, utilising routines from the OpenCV library.

Since the variability in terrain height visible in each frame is small compared to the altitude of the camera, homography estimation is used to avoid the degeneracy that exists when attempting to estimate an essential matrix of a planar surface [28]. Information from the normal vector decomposed from the homography matrix is unused. Note that although the homography matrix was chosen in this implementation, it is not an essential choice - the loosely-coupled nature of the GPS/VO algorithm does not specify *how* the rotation and translation is to be calculated between frames.

To demonstrate that SLAM-like [13] or SFM-like [12] feature tracking is not an essential part of the algorithm, features are not tracked beyond the two frames required to estimate the homography matrix. That is, each pair of images are treated independently.

There is little doubt that better performance from the GPS/VO filter could be achieved by using a more sophisticated egomotion front-end. However, the focus of this paper is not the performance of the egomotion estimation. Rather, the aim is to demonstrate that when Visual Odometry is fused with GPS measurements, certain restrictions (such as platform-specific dynamics, map-matching, feature-tracking, loop-closing, gravity vector or additional sensors) are not necessary to recover a globally-referenced attitude solution.

6.5 Experimental Results

As previously noted, the egomotion estimation is implemented in C++, utilising routines from the OpenCV library. The remainder of the GPS/VO filter (i.e. Visual Odometry, Kalman Filter and closed-loop corrections) is implemented in MATLAB. A real-time implementation of the GPS/VO has not yet been attempted, but given the small length of the state vector, it is not hard to envisage that a real-time implementation is possible (even on modest hardware) subject to the computational requirements of the egomotion front-end.

In all experimental results, the open-loop Visual Odometry has been calculated for comparison using the same initial conditions as the GPS/VO integrated filter.

6.5.1 Archerfield Approach Sequence

The attitude profile of the Archerfield Approach Sequence is shown in Figure 21, where is can be seen that pitch and yaw of the GPS/VO filter quickly converge towards their true value, whereas the open-loop VO remains significantly in error. Examining the first few seconds of the attitude error in Figure 23 shows the pitch and yaw error converging from an initial value of $14°$ to under $5°$ within $0.5s$, consistent with the observability analysis. Furthermore, it is clear from Figure 24 that scale factor is observable throughout the entire sequence.

Roll error (Figure 22(a)) deserves further attention and need to be analysed in conjunction with the flight path in Figure 17 and measured translation in Figure 25. The slightly-curved flight path between Samford Valley and Brookfield implies weak accelerations and consequently sees only slight convergence (compared to open-loop visual odometry) towards truth during the first $200s$ of the sequence. During the manoeuvring near Brookfield at $200s$, roll converges towards truth, despite significant error in the open-loop VO solution.

For the remainder of the sequence, despite strong non-Gaussian noise[9] there are sufficient changes in acceleration from changes in heading and altitude to the allow the roll error to remain bounded without significantly drifting. The exception is between $300 - 400s$, corresponding to the straight and level flight intercepting final approach where it can be seen that the standard deviation in roll increases. The standard deviation reduces at $400s$ when the manoeuvre on final approach is performed.

6.5.2 En-Route Navigation Sequence

Owing to the higher altitude and (in the latter part of the sequence) fewer salient features, it can be clearly seen that the magnitude of the translation noise (Figure 26) is much greater than the Archerfield Approach Sequence (Figure 25). This, combined with the lack of manoeuvres in the dataset contribute to the relatively poor estimation in roll.

Examining roll (Figure 27(a)) and roll error (Figure 28(a)), the small manoeuvre at the start of the sequence reduces the standard deviation of the estimate. At $100s$, the slow turn elicits some response from the filter, but the dynamics are insufficient for the roll error to properly converge. From $200s$ to
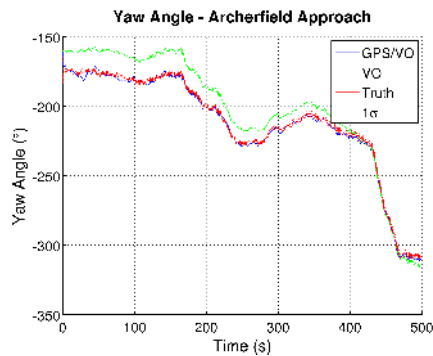
---

[9]The normalisation of translation means that the noise exhibited in the x-axis (Figure 25(a)) is clearly neither zero-mean nor Gaussian
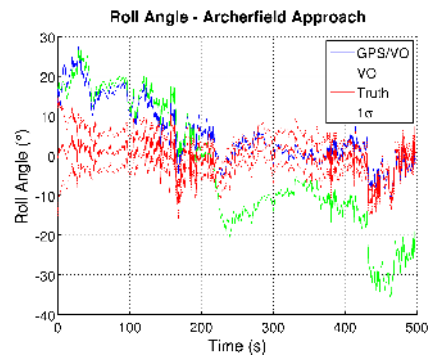
(a) Roll Angle
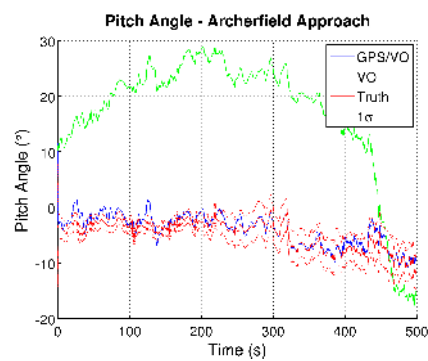
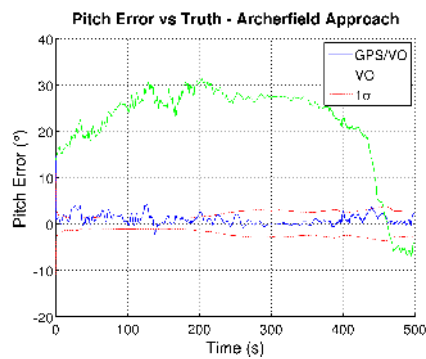
(b) Pitch Angle

(c) Yaw Angle

Figure 21: Attitude during the Archerfield Approach Sequence

near the end of the sequence, the aircraft remains straight and level without manoeuvres. During this time, the output of the GPS/VO follows the open-loop VO output which appears to reduce the error but is merely an artefact of the initial value of the error at 200s.
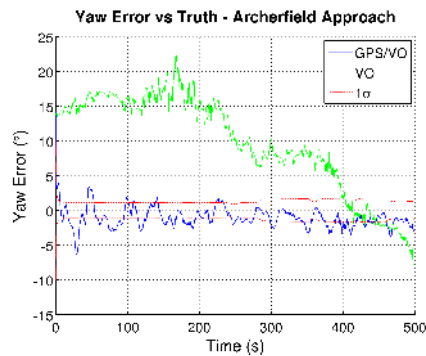
The roll error is especially interesting in the final minute of the En-Route sequence, where the aircraft performs a series of aggressive turns when it starts its descent. It can be seen that although the open-loop VO somewhat tracks the
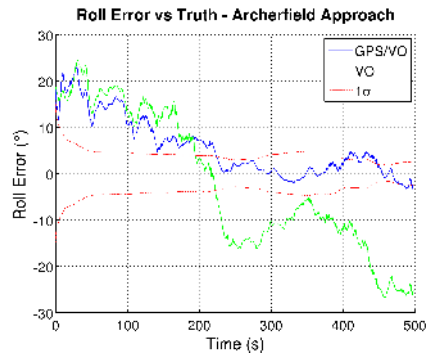
(a) Roll Error

(b) Pitch Error

(c) Yaw Error

Figure 22: Attitude error of the Archerfield Approach sequence

manoeuvre during this time, there is a significant change in error from 40° to
−20° and then returning to 40°. The GPS/VO filter, however, converges back
towards truth during this time and successfully tracks the entire manoeuvre.

Despite the poor roll observability and open-loop VO performance during
the sequence, it is clear that pitch and yaw immediately converge towards the
true value and remain about the true value for the entire sequence. Similar
comments can be made about scale factor for the sequence shown on Figure 29.

31
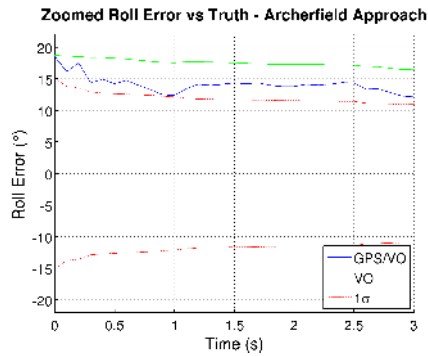
(a) Roll Error
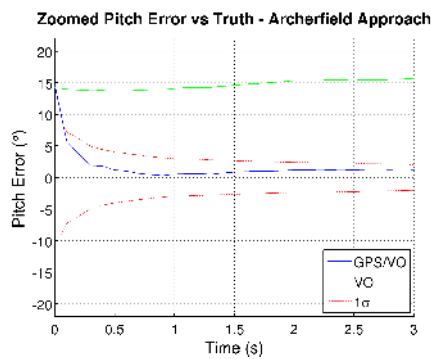
(b) Zoomed Pitch Error
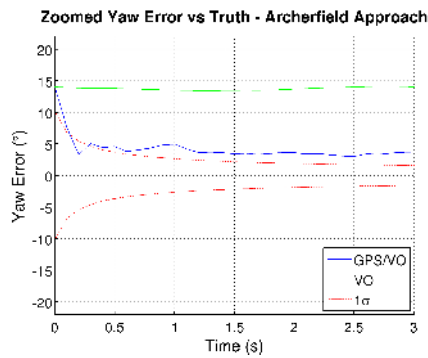
(c) Zoomed Yaw Error

Figure 23: The first few seconds of the Archerfield Approach sequence. Note the fast convergence of pitch and yaw towards their true values

Given the large amount of noise on the Visual Odometry solution and relatively small translations, it would be interesting to ascertain the difference in roll performance during the gentle turn if a better performing Visual Odometry front-end was available. We leave this question for future research.
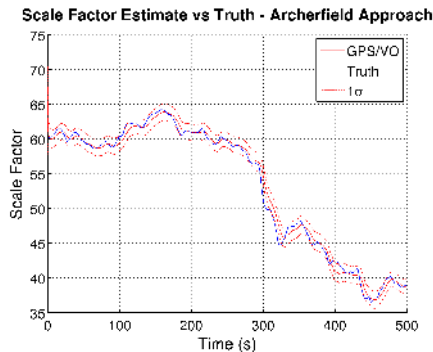
Figure 24: Scale Factor of the Archerfield Approach Sequence. Since the translation is normalised, the scale factor corresponds to the speed of the aircraft

7 Conclusion

In this paper, we presented a loosely-coupled error state GPS/Visual Odometry integrated navigation filter with a similar structure to a classic loosely-coupled error state GPS/INS navigation filter. An observability analysis of the GPS/VO filter showed that position, scale factor and globally-referenced attitude was able to be recovered under acceleration non-parallel to translation in the navigation frame. In the analysis, it was shown that the direction of translation rather than the magnitude that is important to attitude observability as the scale is always recoverable, except where the velocity is zero. This result is achieved without the assistance of additional sensors (e.g. magnetic compass or IMU), platform-specific dynamics (e.g. non-holonomic constraint or coordinated flight) or the need for explicit map management, feature tracking, landmark matching or visual loop-closing. The theoretical observability analysis was backed up by numerical simulation demonstrating the properties of the analysis.

In addition to the simulation, we demonstrated the proposed algorithm on two datasets collected using a light aircraft using an unsophisticated two-frame Visual Odometry front-end. As expected from the observability analysis, the experimental data showed that roll was difficult to estimate in the presence of straight and level flight but converged during manoeuvres. Pitch and yaw immediately converged towards the true value, even during straight and level flight. We note that this behaviour is complementary to a GPS/INS filter, where roll and pitch are bounded by the gravity vector, but yaw must be observed through manoeuvres.

In this paper, we focused on the observability properties on the GPS/VO filter, not on obtaining optimal estimation or real-time performance which presents a number of avenues for future work. A better performing Visual Odometry front-end will certainly assist in the estimation performance. Owing to the non-linear nature of the dynamics, an estimator such as the Unscented Kalman Filter or Particle Filter will likely provide better estimation and convergence performance. Tightly-coupled GPS integration (i.e. pseudorange and carrier phase) may produce interesting observability properties when there are insufficient satellites for calculating a position solution. A non-linear observability analysis, perhaps inspired by [19] or [21] may yield more comprehensive observ-

33
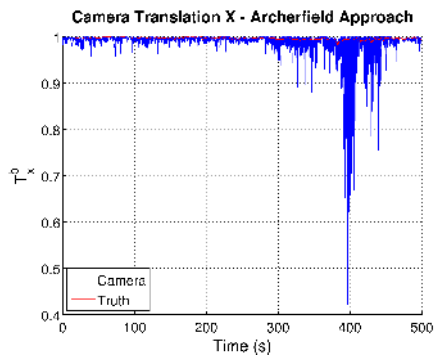
(a) X-Axis Translation
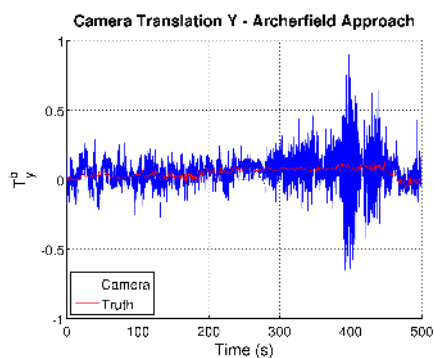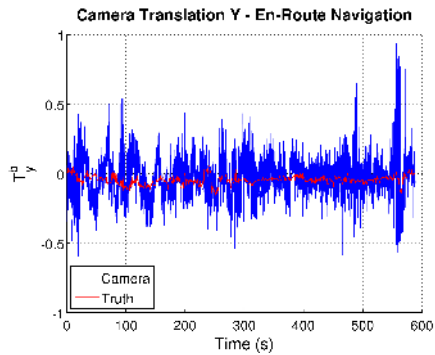

(b) Y-Axis Translation
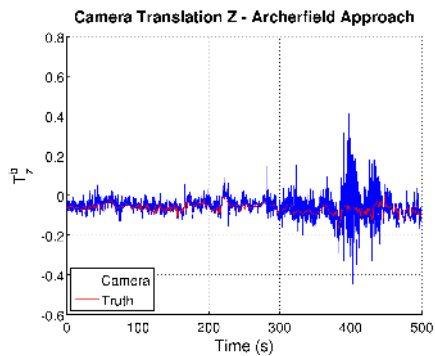
(c) Z-Axis Translation

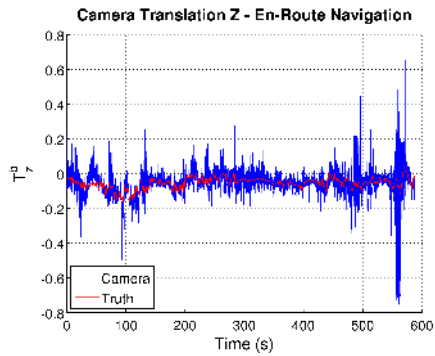Figure 25: Normalised Translation of the Archerfield Approach Sequence

ability results, especially in the presence of noise. Finally, extensions to the filter may be possible such as real-time in-flight camera self-calibration, or automatic calibration of offset between the body frame (as measured by an IMU) and the camera frame, all of which are currently laborious ground-based procedures.
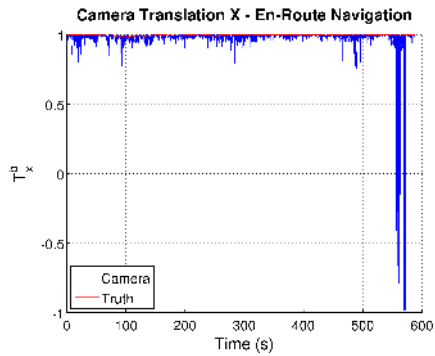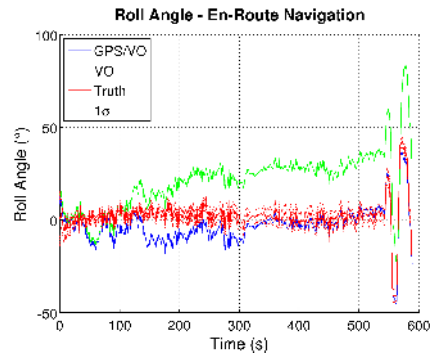
(a) X-Axis Translation

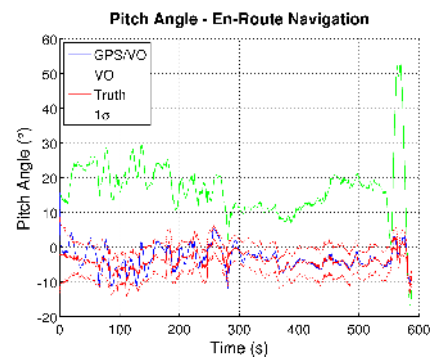(b) Y-Axis Translation

(c) Z-Axis Translation

Figure 26: Normalised Translation of the En-Route Sequence. Note the severe addition of noise near 550s, corresponding the period in Figure 20(b) where few salient features are present in the sequence.
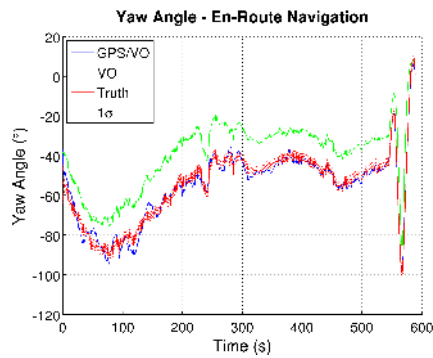
Acknowledgements

(a) Roll Angle

(b) Pitch Angle

(c) Yaw Angle

Figure 27: Attitude during the En-Route Sequence

References

[1] M. Agrawal and K. Konolige. Real-time localization in outdoor environments using stereo vision and inexpensive gps. In *International Conference on Pattern Recognition*, volume 3, pages 1063 –1068, 0-0 2006.

[2] M. Agrawal, K. Konolige, and R.C. Bolles. Localization and mapping for autonomous navigation in outdoor terrains : A stereo vision approach. In
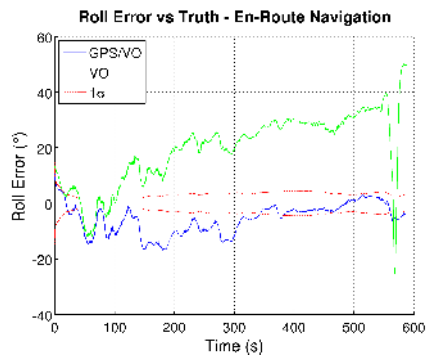
(a) Roll Error
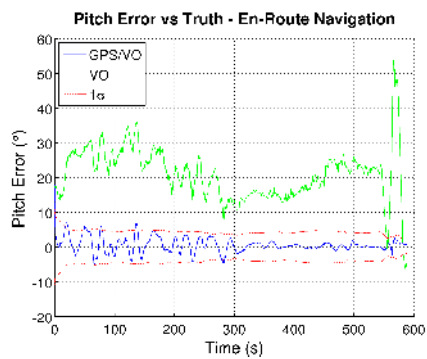
(b) Pitch Error
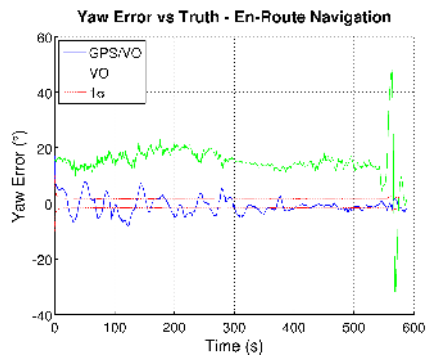
(c) Yaw Error

Figure 28: Attitude error of the En-Route Sequence

IEEE Workshop on Applications of Computer Vision, page 7, 2007.

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision*, 2006.

[4] Francisco Bonin-Font, Alberto Ortiz, and Gabriel Oliver. Visual navigation for mobile robots: A survey. *Journal of Intelligent and Robotic Systems*, 53(3):263–296, November 2008.
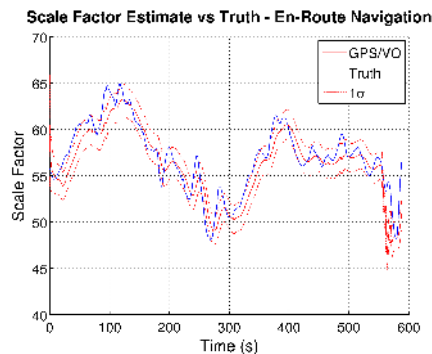
Figure 29: Scale Factor of the En-Route Sequence. Since the translation is normalised, the scale factor corresponds to the speed of the aircraft

[5] M. Bryson and S. Sukkarieh. Observability analysis and active control for airborne slam. *IEEE Transactions on Aerospace and Electronic Systems*, 44(1):261 –280, january 2008.

[6] F. Caballero, L. Merino, J. Ferruz, and A. Ollero. Vision-based odometry and slam for medium and high altitude flying uavs. *Journal of Intelligent and Robotic Systems*, 54:137–161, July 2008.

[7] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. In *European Conference on Computer Vision*, pages 778–792, 2010.

[8] R. Carceroni, A. Kumar, and K. Daniilidis. Structure from motion with known camera positions. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 477 – 484, 17-22 2006.

[9] Andrea Cesetti, Emanuele Frontoni, Adriano Mancini, Andrea Ascani, Primo Zingaretti, and Sauro Longhi. A visual global positioning system for unmanned aerial vehicles used in photogrammetric applications. *Journal of Intelligent and Robotic Systems*, 61:157–168, 2011. 10.1007/s10846-010-9489-5.

[10] G.B. Chatterji, P.K. Menon, and B. Sridhar. GPS/machine vision navigation system for aircraft. *IEEE Transactions on Aerospace and Electronic Systems*, 33(3):1012–1025, 1997.

[11] Chi-Tsong Chen. *Linear System Theory and Design*. Oxford University Press, 3rd edition, 1998.

[12] A. Chiuso, P. Favaro, Hailin Jin, and S. Soatto. Structure from motion causally integrated over time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):523–535, 2002.

[13] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052 –1067, 2007.

[14] D. Dusha, L. Mejias, and R. Walker. Fixed wing attitude estimation using temporal tracking of the horizon and optical flow. *Journal of Field Robotics*, 28(3):355–372, May/June 2011.

[15] Damien Dusha and Luis Mejias. Attitude observability of a loosely-coupled gps/visual odometry integrated navigation filter. In *Australiasian Conference on Robotics and Automation*, 2010.

[16] J.A. Farrell. *Aided Navigation: GPS with High Rate Sensors*. McGraw Hill, 2008.

[17] Duncan G. Greer, Rhys Mudford, Damien Dusha, and Rodney Walker. Airborne systems laboratory for automation research. In *International Congress of the Aeronautical Sciences*, 2010.

[18] Paul D. Groves. *Principles of GNSS, Inertial, and Multi-Sensor Integrated Navigation Systems*. Artech House, 2008.

[19] Sinpyo Hong, Ho-Hwan Chun, Sun-Hong Kwon, and Man Hyung Lee. Observability measures and their application to gps/ins. *Vehicular Technology, IEEE Transactions on*, 57(1):97 –106, jan. 2008.

[20] Sinpyo Hong, Man Hyung Lee, Ho-Hwan Chun, Sun-Hong Kwon, and J.L. Speyer. Observability of error states in gps/ins integration. *IEEE Transactions on Vehicular Technology*, 54(2):731 – 743, march 2005.

[21] Eagle S. Jones and Stefano Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, 30(4):407–430, 2011.

[22] Jonathan Kelly and Gaurav S. Sukhatme. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *International Journal of Robotics Research*, 30(1):56–79, August 2011.

[23] Richard P. Kornfeld, R. John Hansman, and John J. Deyst. Single-antenna gps-based aircraft attitude determination. *Navigation*, 45(1):51–60, 1998.

[24] H. Kume, T. Taketomi, T. Sato, and N. Yokoya. Extrinsic camera parameter estimation using video images and gps considering gps positioning accuracy. In *International Conference on Pattern Recognition*, pages 3923 –3926, 2010.

[25] John Lai, Luis Mejias, and Jason J. Ford. Airborne vision-based collision-detection system. *Journal of Field Robotics*, 28(2):137–157, 2011.

[26] R. Lerner, E. Rivlin, and H.P. Rotstein. Pose and motion recovery from feature correspondences and a digital terrain map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1404–1417, 2006.

[27] Hao Li, F. Nashashibi, and G. Toulminet. Localization for intelligent vehicle by fusing mono-camera, low-cost gps and map data. In *IEEE Conference on Intelligent Transportation Systems*, pages 1657 –1662, 2010.

[28] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3D Vision: From Images to Geometric Models*. Springer, 2004.

[29] R. Mahony, T. Hamel, and J.-M. Pflimlin. Nonlinear complementary filters on the special orthogonal group. *IEEE Transactions on Automatic Control*, 53(5):1203 –1218, 2008.

[30] M. Maimone, Y. Cheng, and L. Mattheis. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics*, 43(3):169–186, 2007.

[31] A. Rae and O. Basir. Reducing multipath effects in vehicle localization by fusing gps with machine vision. In *International Conference on Information Fusion*, pages 2099 –2106, 2009.

[32] I. Rhee, M.F. Abdel-Hafez, and J.L. Speyer. Observability of an integrated gps/ins during maneuvers. *IEEE Transactions on Aerospace and Electronic Systems*, 40(2):526 – 535, april 2004.

[33] Peter J. Roberts, Rodney A. Walker, and Peter O'Shea. Fixed wing uav navigation and control through integrated gnss and vision. In *AIAA Guidance, Navigation, and Control Conference*, San Francisco, California, 15-18 August 2005.

[34] D. Schleicher, L.M. Bergasa, M. Ocana, R. Barea, and M.E. Lopez. Realtime hierarchical outdoor slam based on stereovision and gps fusion. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):440 –452, 2009.

[35] T. Vidal-Calleja, M. Bryson, S. Sukkarieh, A. Sanfeliu, and J. Andrade-Cetto. On the observability of bearing-only slam. In *IEEE International Conference on Robotics and Automation*, pages 4114 –4119, 10-14 2007.

[36] Charles Wang, Rodney A. Walker, and Miles Moody. Single antenna attitude algorithm for non-uniform antenna gain patterns. *Journal of Spacecraft and Rockets*, 44(1):221–229, 2007.

[37] Lijun Wei, Cindy Cappelle, Yassine Ruichek, and Fredrick Zann. Gps and stereovision based visual odometry: application to urban scene mapping and intelligent vehicle localization. *International Journal of Vehicular Technology*, In Press:1–13, 2011.