Multi-Antenna Vision-and-Inertial-Aided CDGNSS for Micro Aerial Vehicle Pose Estimation

James E. Yoder, Peter A. Iannucci, Lakshay Narula, and Todd E. Humphreys
Radionavigation Laboratory
The University of Texas at Austin

BIOGRAPHIES

James Yoder is a graduate student in the department of Aerospace Engineering and Engineering Mechanics at the University of Texas at Austin, and a member of the UT Radionavigation Laboratory. He received a B.S. in Electrical and Computer Engineering from the University of Texas at Austin in 2019. His research interests currently include GNSS signal processing, estimation, and sensor fusion techniques for centimeter-accurate navigation.

Peter A. Iannucci (BS, Electrical Engineering and Computer Science and Physics, MIT; PhD, Networks and Mobile Systems, CSAIL, MIT) is a postdoctoral research fellow in the Radionavigation Laboratory at The University of Texas at Austin, and a member of the UT Wireless Networking and Communications Group (WNCG). His current research interests include collaborative navigation, multi-spectral mapping, and re-purposing broadband Internet satellites for radionavigation.

Lakshay Narula (BTech, Electronics Engineering, IIT-BHU, India; MS, Electrical and Computer Engineering, The University of Texas at Austin) is a Ph.D. student at The University of Texas at Austin, and a graduate research assistant at the UT Radionavigation Lab. His research interests include state estimation, radionavigation, localization and mapping, and secure perception for autonomous systems. He was a recipient of the 2017 Qualcomm Innovation Fellowship.

Todd Humphreys (BS, MS, Electrical Engineering, Utah State University; PhD, Aerospace Engineering, Cornell University) is an associate professor in the department of Aerospace Engineering and Engineering Mechanics at The University of Texas at Austin, where he directs the Radionavigation Laboratory. He specializes in the application of optimal detection and estimation techniques to secure and robust perception for automated systems and centimeter-accurate location. His awards include The University of Texas Regeants' Outstanding Teaching Award (2012), the National Science Foundation CAREER Award (2015), the Institute of Navigation Thurlow Award (2015), the Qualcomm Innovation Fellowship (2017), and the Presidential Early Career Award for Scientists and Engineers (PECASE, 2019). He is a Fellow of the Institute of Navigation.

ABSTRACT

A system is presented for multi-antenna carrier phase differential GNSS (CDGNSS)-based pose (position and orientation) estimation aided by monocular visual measurements and a smartphone-grade inertial sensor. The system is designed for micro aerial vehicles, but can be applied generally for low-cost, lightweight, high-accuracy, geo-referenced pose estimation. Visual and inertial measurements enable robust operation despite GNSS degradation by constraining uncertainty in the dynamics propagation, which improves fixed-integer CDGNSS availability and reliability in areas with limited sky visibility. No prior work has demonstrated an increased CDGNSS integer fixing rate when incorporating visual measurements with smartphone-grade inertial sensing. A central pose estimation filter receives measurements from separate CDGNSS position and attitude estimators, visual feature measurements based on the ROVIO measurement model, and inertial measurements. The filter's pose estimates are fed back as a prior for CDGNSS integer fixing. A performance analysis under both simulated and real-world GNSS degradation shows that visual measurements greatly increase the availability and accuracy of low-cost inertial-aided CDGNSS pose estimation.

INTRODUCTION

Micro aerial vehicles (MAVs) are increasingly being used for applications such as 3D mapping that require both (1) precise pose (position and orientation) knowledge relative to a global coordinate system fixed to the Earth's surface,

Preprint of the 2020 ION GNSS+ Conference
Online, Sept. 21-25, 2020

and (2) close-in maneuvers to ensure high resolution of the area being mapped. A global coordinate system is essential for applications such as automated infrastructure inspection [1], 3D modeling of buildings [2], disaster recovery or search and rescue [3], and open-world virtual reality [4], in which mapping data from the MAV is consumed by other, possibly automated, agents, potentially long after the initial mapping process.

Carrier-phase differential GNSS (CDGNSS) techniques such as real-time kinematic (RTK) positioning can offer centimeter-accurate positioning accuracy, and so serve as an excellent anchor for globally-referenced pose estimation. However, such accuracy is only achieved robustly and instantaneously when so-called carrier phase ambiguities are resolved to their integer values [5]. Confident ambiguity resolution depends on a large number (e.g., 12+) of participating low-multipath GNSS signals [6], or on a tight prior position estimate. But as a mapping MAV passes close to buildings, under overhanging rooftops, or around foliage, GNSS signal blockage and multipath effects become severe, limiting the availability of CDGNSS unaided by inertial sensing. Users of mapping MAVs therefore currently tend to avoid altogether areas where GNSS signals might be obstructed [7].

The MAV platform also places unique constraints on navigation systems: onboard compute is restricted by size, weight, and power limitations; the lively system dynamics of MAVs require low-latency measurement and estimation; and, in many cases, MAVs may only feature low-cost consumer-grade cameras and inertial measurement units (IMUs).

This paper describes a method for improving CDGNSS performance via tight coupling with a visual-inertial pose estimator. A CDGNSS system is defined herein as *tightly coupled* with visual and inertial sensing if the latter aid in resolving CDGNSS integer ambiguities. A *loosely coupled* CDGNSS system, in contrast, is based on a standalone CDGNSS estimator that operates without aiding from other sensors. Information in a loosely-coupled system only flows one way, from the CDGNSS estimator to the downstream estimators.

Tight coupling with inertial sensors is a widely-studied and well-understood method of increasing the robustness and availability of fixed-integer CDGNSS positioning [8]–[11]. Early efforts used high quality navigation- or tactical-grade inertial sensors to provide positioning constraints over lengthy GNSS outages. More recently, researchers have exploited lower-cost industrial-grade micro-electro-mechanical system (MEMS) inertial sensors to bridge short GNSS outages [12]–[14] or for attitude-only CDGNSS [15]. These industrial-grade MEMS sensors are significantly larger, heavier, and more expensive than the consumer-grade MEMS inertial sensors of the type commonly found in low-cost MAVs.

In a companion paper [4], tight coupling with a consumer-grade MEMS sensor is shown to improve CDGNSS performance in degraded GNSS conditions or over short complete outages. The current paper explores the addition of visual measurements to the same tightly-coupled inertial-CDGNSS system analyzed in [4].

There are reasons beyond integer ambiguity fixing for inertial sensing in precise MAV positioning: First, CDGNSS combined with an inertial sensor can provide the full pose of the vehicle. Second, inertial sensing allows the global scale of visual features to be observable when combined with visual positioning, which is important for visual-inertial positioning during GNSS outages [16].

A popular method for MAV navigation is the fusion of visual and inertial measurements [17]–[21]. These systems generally operate by tracking visual features seen by one or more camera, and taking the position of features in the camera field of view as measurements for a pose estimator [17]. In some cases, the positions of the visual features are jointly estimated along with the camera pose, in a technique known as simultaneous localization and mapping (SLAM). Absent a prior globally-referenced map, visual-inertial navigation systems are fundamentally relative positioning systems, and cannot provide a globally-referenced pose estimate. They also suffer from odometric drift except in certain cases where returning to a previously-visited location enables “loop closure.”

Tight coupling of visual-inertial sensing with CDGNSS has not been as widely studied as inertial-only coupling. The VISRTK technique proposed in [22] directly incorporates the double-difference carrier phase measurement model, including integer ambiguities, into a bundle-adjustment based SLAM problem. This approach is near-optimal, but far too computationally demanding for real-time implementation on an MAV, and does not attempt to incorporate IMU measurements.

The authors of [23] proposed tight coupling of CDGNSS with visual positioning, an inertial sensor of unstated quality, and barometric altitude measurements, but the visual positioning method used requires the collection and curation of precise aerial imagery. The use of 2-dimensional aerial maps also precludes close-in maneuvering to buildings and other obstacles.

Li et al. in [24] implemented tight coupling of single-antenna CDGNSS with a monocular visual-inertial odometry via

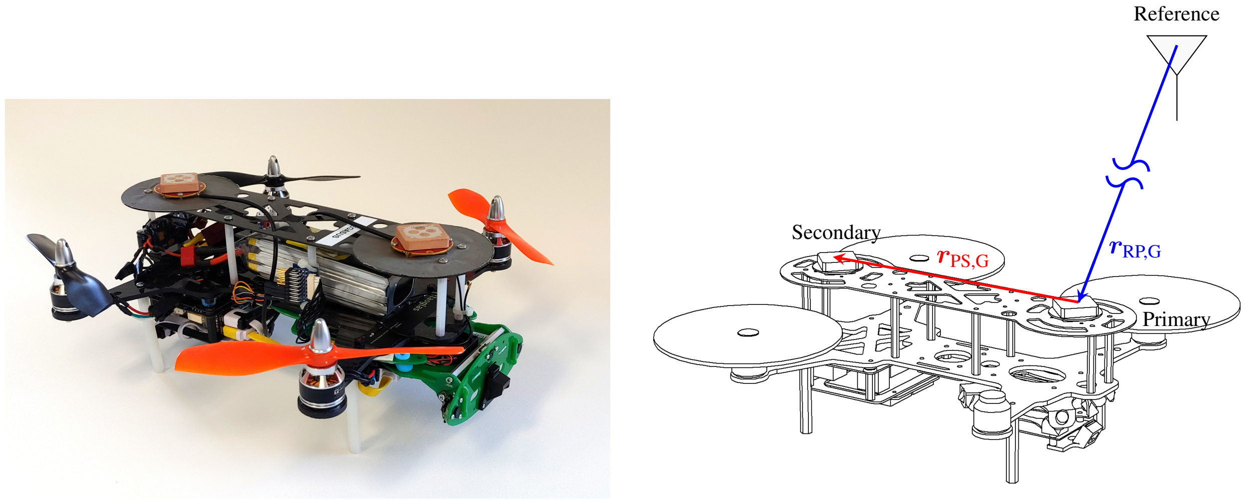

Fig. 1: The University of Texas Radionavigation Lab quadrotor MAV used for data collection. The primary and secondary GNSS antennas are seen mounted to the top plate of the MAV.

a multi state constraint Kalman filter (MSCKF) [19]. However, the system as reported depended on an industrial-grade IMU. Moreover, SLAM-type visual-inertial techniques can be advantageous over MSCKF estimation due to their ability to map visual features while an RTK fix is available, then exploit the previously-mapped features during an outage. In contrast, the MSCKF technique is fundamentally odometric: visual feature tracks are immediately marginalized when ingested into the filter.

The work perhaps most comparable to this paper is [25], which describes the tight coupling of CDGNSS, a smartphone-grade IMU, SLAM-based visual feature measurements, and a beacon-based local positioning system for a robotic lawn mower application. While [25] showed that visual measurements reduce the overall position drift during an RTK fix outage, it did not offer a convincing demonstration of an improved integer fix rate.

The primary contribution of this paper is the incorporation of visual measurements into a tightly-coupled multi-antenna CDGNSS-inertial pose estimator using a smartphone-grade IMU and camera. To best of the authors' knowledge, this paper provides the first demonstration of an increased RTK integer fix rate using visual-inertial aiding with smartphone-grade sensors.

PLATFORM, COORDINATE FRAMES, AND NOTATION

MAV Platform

The reference MAV platform used in this work, shown in Fig. 1, is a quadrotor MAV featuring two low-cost L1-only GNSS antennas connected to a custom GNSS frontend. The frontend provides raw intermediate frequency (IF) samples that are processed by an advanced software-defined GNSS receiver, described in [26], running onboard the MAV. The software-defined GNSS receiver provides GNSS observables from the GPS, Galileo, and satellite-based augmentation system (SBAS) constellations. Imagery is collected by a monocular 640x480 globally-shuttered camera running at 30 Hz. A smartphone-grade MEMS IMU (a Bosch BMX055) integrated with the GNSS frontend provides inertial measurements that are hardware timestamped with the GNSS receiver sample clock. The two GNSS antennas are referred to as the *primary* and *secondary* antennas.

Coordinate Frames

This paper's measurements and estimates are referenced to the following coordinate frames:

G: WGS-84 Earth Centered Earth Fixed (ECEF) frame.
W: "World" frame, a quasi-inertial East-North-Up (ENU) frame fixed to the Earth's surface and centered at the RTK reference antenna's phase center.

B: “Body” frame, centered at the primary GNSS antenna’s phase center and fixed to the quadrotor body.
U: “IMU” frame, centered at the IMU’s accelerometer triad and fixed to the quadrotor body.
C: “Camera” frame, centered at the optical center of the camera. Its X and Y axes are aligned with the X and Y pixel axes on the camera’s focal plane, and its Z axis points outwards towards the center of the camera’s view.

Notation

Vectors are written in lowercase and bold with a subscript indicating the frame in which the vector is expressed; e.g., $\boldsymbol{r}_\mathrm{W} \in \mathbb{R}^3$ is a vector expressed in the W frame. Matrices are non-bold and uppercase; e.g., A. Vector and matrix transpose is denoted by a superscript T; e.g., $\boldsymbol{u}^\mathsf{T}$. Subscripts for attitude (direction cosine) matrices indicate in a right-to-left sense the original and transformed frames, following a frame rotation convention; e.g., $R_\mathrm{BW} = R_\mathrm{WB}^\mathsf{T} \in SO(3)$ is the W-to-B attitude matrix: $\boldsymbol{r}_\mathrm{B} = R_\mathrm{BW}\boldsymbol{r}_\mathrm{W}$. Attitude errors are expressed as a vector of 3-1-2 Euler angles $\boldsymbol{e} = [\phi, \theta, \psi]^\mathsf{T}$ relative to a reference attitude matrix $\bar{R} \in SO(3)$. Thus, if R_BW is the true W-to-B attitude matrix, then $R_\mathrm{BW} = R(\boldsymbol{e}_\mathrm{BW})\bar{R}_\mathrm{BW}$, where $R : \mathbb{R}^3 \to SO(3)$ is a function that converts 3-1-2 Euler angles to an attitude matrix. In some cases, two-dimensional Euler angle vectors $\boldsymbol{e} = [\phi, \theta]^\mathsf{T}$ will be used to indicate the direction of a unit vector; these will be clearly specified when introduced. The skew-symmetric cross-product-equivalent matrix corresponding to a vector $\boldsymbol{u}$ is denoted by $[\boldsymbol{u}\times]$. $\boldsymbol{e}_i \in \mathbb{R}^3$ denotes a 3-vector whose ith element is unity and all others zero; e.g., $\boldsymbol{e}_3 = [0, 0, 1]^\mathsf{T}$. A vector with superscript u indicates it has been normalized to unit length; e.g., $\boldsymbol{r}^u$. The notation $[\boldsymbol{u}]_{i:j}$ denotes the vector composed of scalar elements from the ith to the jth of the vector $\boldsymbol{u}$, while $[\boldsymbol{u}]_i$ denotes the ith scalar element of $\boldsymbol{u}$.

FEDERATED ESTIMATION ARCHITECTURE

This work explores the coupling of discrete RTK estimators with a central pose estimator incorporating visual-inertial measurements. The central pose estimator is implemented as an unscented Kalman filter (UKF). As shown in Fig. 2, the central estimator receives measurements from a smartphone-grade IMU, a monocular camera, and two independent single-baseline RTK estimators, referred to as the *position* and *attitude* RTK estimators. The position RTK estimator produces a single-baseline RTK solution between the MAV’s primary GNSS antenna and a fixed reference antenna with a pre-surveyed location. The attitude RTK estimator produces an RTK solution between the MAV’s primary and secondary GNSS antennas that is constrained by the known baseline length between the two antennas.

When operating in the *loosely-coupled* mode, the RTK estimators are unaided; that is, they provide position and attitude measurements to the central pose estimator but do not ingest any inertial or vision measurements to aid in producing RTK solutions.

When operating in the *tightly-coupled* mode, the central estimator’s output is taken as the prior for the RTK estimators: transformed versions of the central estimator’s propagated pose estimate and its associated covariance matrix replace the internal state and covariance of the RTK estimators’ *a priori* estimates. In this way, the central pose estimator provides a prior constraint on the RTK solutions, aiding the integer ambiguity resolution process. This propagation and state replacement strategy is similar to the “position seeding” described in [27].

This federated architecture, in which discrete estimators pass data back and forth, is suboptimal in terms of estimation performance and consistency compared to an “all-in-one” pose estimation filter that directly takes in GNSS pseudorange and carrier phase observables. Seeding the RTK estimators with the central pose estimate risks introducing unmodeled correlations between the pose estimator’s state errors and errors in the measurements it receives from the RTK estimators. However, this “loopy inference” is mitigated by the fact that double-difference carrier phase measurements are millimeter-precise compared to the decimeter-level precision of the pose-estimator-provided prior position. Such precision asymmetry implies that, when conditioned on correct integer ambiguities, the RTK estimators’ solution errors are dominated by carrier phase multipath, which is uncorrelated with the pose estimator’s state errors. In practice, the precision asymmetry is magnified by scaling the *a priori* covariance matrices $\bar{P}_\mathrm{RP,G}$ and $\bar{P}_\mathrm{PS,G}$ sent from the pose estimator to the RTK estimators with scalar inflation factors α and β, respectively (see Fig. 2).

As noted in [27], the integer-conditioning-induced decorrelation breaks down as the number of double-difference carrier phase measurements drops below 3, at which point the 3-dimensional position state becomes truly unobservable via GNSS measurements alone. When this occurs, the returned measurements contain information solely from the prior along any unobservable directions, leading to estimator inconsistency. To avoid this situation, the pose estimator does not ingest RTK measurements made with fewer than 3 double-difference measurements. This restriction has a negligible

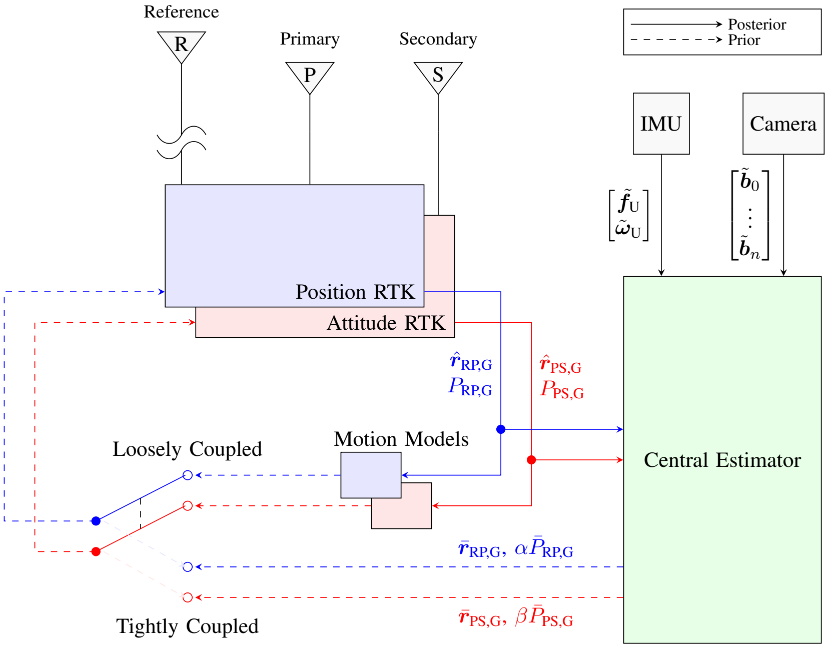
Reference
R
Primary
P
Secondary
S
Posterior
Prior
IMU
Camera
$\begin{bmatrix}\tilde{\boldsymbol{f}}_{\mathrm{U}}\\ \tilde{\boldsymbol{\omega}}_{\mathrm{U}}\end{bmatrix}$
$\begin{bmatrix}\tilde{\boldsymbol{b}}_0\\ \vdots \\ \tilde{\boldsymbol{b}}_n\end{bmatrix}$
Position RTK
Attitude RTK
$\hat{\boldsymbol{r}}_{\mathrm{RP,G}}$
$P_{\mathrm{RP,G}}$
$\hat{\boldsymbol{r}}_{\mathrm{PS,G}}$
$P_{\mathrm{PS,G}}$
Loosely Coupled
Motion Models
Central Estimator
$\bar{\boldsymbol{r}}_{\mathrm{RP,G}}$, $\alpha\bar{P}_{\mathrm{RP,G}}$
$\bar{\boldsymbol{r}}_{\mathrm{PS,G}}$, $\beta\bar{P}_{\mathrm{PS,G}}$
Tightly Coupled

Fig. 2: Estimator architecture. When loosely coupled, the state of the position and attitude RTK estimators is propagated via simple internal motion models. When tightly coupled, transformed versions of the central estimator's propagated pose estimate and its associated covariance matrix replace the internal state and covariance of the RTK estimators' *a prior* estimates.

effect in practice, as situations with only one or two double-difference measurements are typically transient in nature, and the gap is easily covered by visual-inertial positioning.

Of course, if ambiguities are incorrectly resolved, substantial correlation can build up between the RTK solution errors and the pose estimator state errors. Hence, extracting good performance from this paper's federated tight coupling scheme demands careful aperture testing to validate candidate CDGNSS integer fixes [28]. The integer fixing logic employed in this work, which is described in [26], adopts a layered approach of signal selection, pseudorange-based innovations testing, and a controlled-failed-fixing-rate integer acceptance test.

The federated estimation architecture, although suboptimal, has the virtue of being simple to implement and diagnose and can draw on existing well-tested RTK estimators [27]. Moreover, it enables switching back-and-forth between loose and tight coupling, as illustrated in Fig. 2, which allows for convenient examination of the benefits of tight coupling. Finally, it manages to deliver impressive results compared to loose coupling, as will be shown, and so serves as a valuable stepping stone to future work in more complete tight coupling.

CDGNSS ESTIMATORS

This section briefly outlines the RTK baseline definitions, integer ambiguity resolution process, and outlier rejection scheme used by the RTK estimators.

RTK Baselines

Two independent RTK solutions are maintained by the RTK estimators: a position solution, $\boldsymbol{r}_{\mathrm{RP,W}}$ representing a vector pointing from the fixed RTK reference station to the MAV's primary GNSS antenna, and an attitude solution, $\boldsymbol{r}_{\mathrm{PS,W}}$ representing a vector pointing from the MAV's primary to its secondary onboard GNSS antennas, which provides globally-referenced pitch and yaw information. The attitude solution is constrained to the known baseline length, which is approximately one wavelength at the GNSS L1 frequency. Fig. 1 illustrates these CDGNSS baselines.

Having two independent RTK estimators is suboptimal as they do not account for the correlation in measurements due to the shared antenna in the two RTK baselines. However, independent RTK estimators are simpler to develop and maintain, and the configuration parameters of each RTK estimator, such as elevation masks, integer aperture test thresholds, and outlier exclusion parameters, can be tuned independently, which is beneficial as the baseline constraint on the attitude estimator allows much looser thresholds for equivalent fixing performance.

Outlier Rejection

Multipath and diffraction effects in urban environments cause frequent corruption of GNSS observables. A two-tiered strategy is employed to mitigate these outliers. First, a set of heuristics is applied to screen incoming observables and filter out measurements likely to be outliers. This screening applies thresholds for carrier-to-noise ratio, minimum satellite elevation, and a phase lock statistic calculated by the GNSS measurement engine. Measurements which do not pass this screening are discarded. Next, the filter applies a χ^2-type innovations test to each incoming batch of double-difference measurements. If this test is failed, or if ambiguity resolution fails, the solution is iteratively re-attempted while excluding single measurements. This outlier rejection scheme has been shown to perform well in real-world urban positioning tests [26].

Integer Ambiguity Resolution

To exploit the exquisite precision of double-difference carrier phase measurements, their integer-valued ambiguities must be resolved [5]. At every measurement epoch, a real-valued (float) solution is first attempted based on the position prior in the filter state and double-difference pseudorange measurements. In the loosely-coupled mode, this position prior is propagated from the previous RTK estimator solution using a simple nearly-constant-velocity motion model [29]. In the tightly-coupled mode, this position prior is extracted from the state of the central pose estimator.

Next, integer ambiguity resolution is attempted using integer least squares (ILS) [30] and validated via an integer aperture test with a predetermined failed fixing rate [28]. For the attitude RTK estimator, the known baseline length between the two onboard GNSS antennas is used as an additional constraint in the ILS search process.

The RTK estimators apply a single-epoch integer ambiguity resolution strategy. At every measurement epoch, after performing an ILS solution, the integer components of the RTK estimator state are discarded—by marginalization in the case of a float solution, or by conditioning on the integer state in the case of a fixed solution. This makes the estimators insensitive to cycle slips, which is critical to achieving high RTK availability and reliability in an urban environment, when signal degradations and outages are extremely frequent [26]. Further details on the signal screening, integer fixing, and conditioning logic may be found in [26].

CENTRAL POSE ESTIMATOR

The central pose estimator is implemented as a UKF. Position and attitude RTK solutions and pixel intensity measurements of tracked visual features are ingested in the measurement update step, and IMU angular rate and specific force measurements are applied to propagate the filter state forward in time.

Filter Overview and State Parameterization

The central estimator state has three groups of components (“sectors”): parameters of rigid-body platform motion, parameters of an Ornstein-Uhlenbeck drift model for the IMU, and vision model parameters. The full state vector $\boldsymbol{x}$ is written in terms of these sectoral state vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ as

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \end{bmatrix}$$

Rigid-Body Sector: This sector contains 9 degrees of freedom:

$$\boldsymbol{x}_1 = \left[\begin{array}{rl} \boldsymbol{r}_{\mathrm{W}} \in \mathbb{R}^3 & \text{position of IMU's accelerometer triad} \\ \boldsymbol{v}_{\mathrm{W}} \in \mathbb{R}^3 & \text{time derivative of } \boldsymbol{r}_{\mathrm{W}} \\ \boldsymbol{e}_{\mathrm{BW}} \in \mathbb{R}^3 & \text{attitude error relative to } \bar{R}_{\mathrm{BW}} \in SO(3) \end{array} \right]$$

The reference attitude matrix $\bar{R}_{\mathrm{BW}}$ is maintained auxiliary to the estimator state vector, and absorbs the estimated attitude error after each measurement update.

IMU Bias Sector: This sector contains six degrees of freedom representing two three-dimensional Ornstein-Uhlenbeck processes, one each for the bias of the accelerometer and the gyroscope:

$$\boldsymbol{x}_2 = \begin{bmatrix} \boldsymbol{b}_{a\mathrm{U}} \in \mathbb{R}^3 \\ \boldsymbol{b}_{g\mathrm{U}} \in \mathbb{R}^3 \end{bmatrix}$$

Vision Sector: The vision sector of the central estimator assumes the measurement model and state parameterization of the filter-based visual-inertial SLAM framework developed by Bloesch et al. [18] under the name ROVIO. Many schemes for filter-based visual-inertial navigation exist, differing markedly in how measurements are ingested, what parameters are tracked in the filter state, and when parameters are dropped from the state. ROVIO defines a class of trackable external points, or "features." It ingests camera images in two phases: feature identification and feature tracking.

Feature identification is a heuristic procedure that reduces a dense grid of pixels to a discrete set of trackable features, and appends these as new components to the filter state vector. These new state components form a (nonlinear) parameterization for the estimated locations of external points. This is unlike the MSCKF [19] technique, which augments the state vector with a parameterization of past camera poses. A feature is identified by a "patch": a square bitmap extracted from the camera's field of view. Such a patch is considered high-quality if it exhibits strong gradients and therefore high trackability via optical flow [31]. Tracked features are assumed to be fixed with respect to the W frame, whereas the camera frame C is mobile.

ROVIO refines this basic method by extracting feature patches at multiple scale levels. These are stored in an image pyramid called a "multilevel patch." This paper uses pyramids of 8×8 pixels and 4 levels. Feature tracking is based on the well-known Lucas-Kanade optical flow method [32], discussed in a later section.

Camera measurements improve knowledge of a feature's position relative to the camera. By tracking correlations between the camera pose and the feature's position, the filter adjusts its MAV pose estimate based on where the features appear in each image's pixel field.

ROVIO has several advantages over competing visual measurement schemes. Unlike MSCKF, it maintains estimates of feature positions in the filter state (though the number of tracked features is constrained by the $\mathcal{O}(n^3)$ complexity of the Kalman filter). In CDGNSS aiding, this permits visual features to be mapped when GNSS is available and exploited during outages. While optimization-based SLAM achieves better accuracy than filter-based SLAM for an equivalent amount of computation [33], ROVIO offers superior robustness for real-time flight control. This is because ROVIO has predictable runtime complexity [34], naturally accommodates motion blur and rapid platform motion, and can track co-dimensional features like 1-D edges [18], which may not be trackable by the 2D feature-descriptor-based matching used by many optimization-based SLAM systems.

Tracked features are encoded in the state vector using the inverse distance parameterization (IDP) [35] relative to the instantaneous camera frame. The position $\boldsymbol{r}_{k,\mathrm{C}}$ of feature k is represented by a unit bearing vector $\boldsymbol{u}_{k,\mathrm{C}} = \boldsymbol{r}_{k,\mathrm{C}}/\|\boldsymbol{r}_{k,\mathrm{C}}\| \in S^2$ and inverse distance $\rho_k = 1/\|\boldsymbol{r}_{k,\mathrm{C}}\| \in \mathbb{R}^+$. IDP reduces linearization error compared to either a Cartesian or bearing-and-distance parameterization [36]. When a feature is first identified, its bearing $\boldsymbol{u}_{k,\mathrm{C}}$ is tightly constrained by the pixel position of the feature within the camera image, and its inverse distance ρ_k is modeled with a diffuse prior. Rather than store the degenerate (i.e., unit-norm) three-component bearing $\boldsymbol{u}_{k,\mathrm{C}}$ in the state, a two-component error state is stored in the form of two Euler angles with respect to a reference rotation $\bar{R}_{\mathrm{CT}_k}$. T_k is the "tangent" frame for feature k such that $\boldsymbol{u}_{k,\mathrm{C}} = R_{\mathrm{CT}_k}\boldsymbol{e}_3$.

Earlier implementations of IDP required additional "anchor states" to represent the pose of the camera at the time the feature was first observed [35]. But the robocentric formulation presented in [21] has been found to be free of the extraneous observable dimensions which appear when linearizing the measurement models of non-robocentric feature formulations. This "observability mismatch" has been a major cause of inconsistency in many other filtering-based visual-inertial navigation schemes [20].

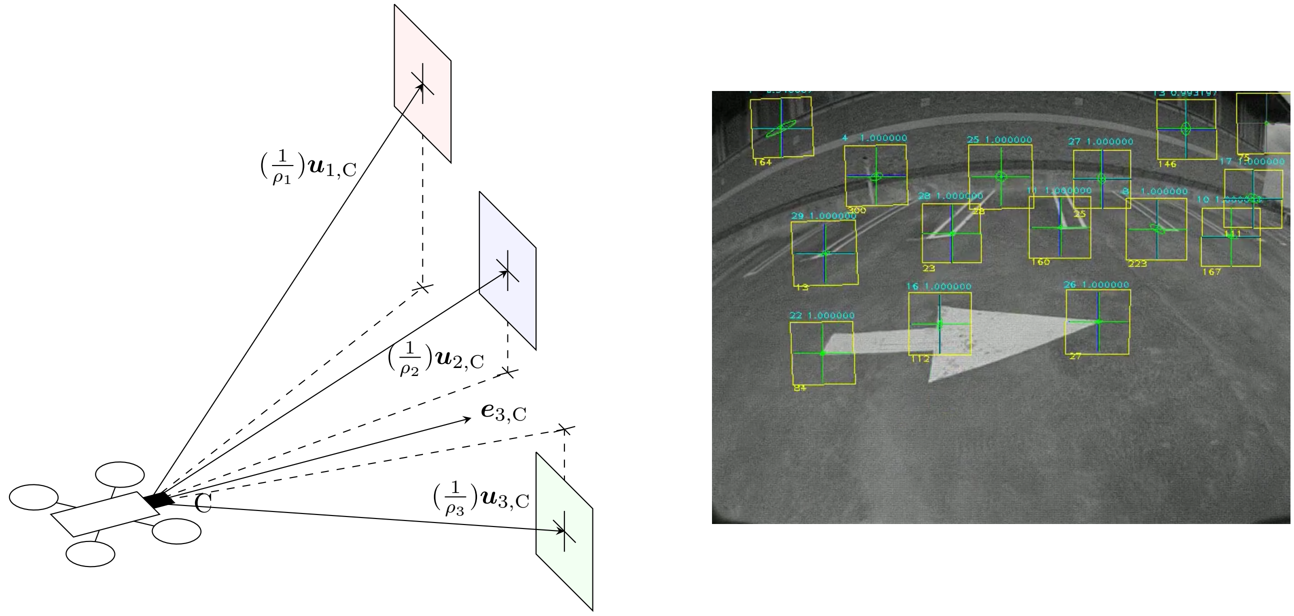
$(\frac{1}{\rho_1})\boldsymbol{u}_{1,\mathrm{C}}$
$(\frac{1}{\rho_2})\boldsymbol{u}_{2,\mathrm{C}}$
$\boldsymbol{e}_{3,\mathrm{C}}$
$(\frac{1}{\rho_3})\boldsymbol{u}_{3,\mathrm{C}}$
C

Fig. 3: Representation of feature positions in the filter state, and example camera image showing tracked features. $\boldsymbol{e}_{3,\mathrm{C}}$ represents a unit vector pointing in the direction of the camera Z axis, which is normal to the image plane.

When tracking n features, the vision sector has $3 + 3n$ degrees of freedom, expressed as

$$\boldsymbol{x}_3 = \left[\begin{array}{rll} \boldsymbol{e}_{\mathrm{CB}} & \in \mathbb{R}^3 & \text{attitude error of the camera relative to } \bar{R}_{\mathrm{CB}} \in SO(3) \\ \boldsymbol{e}_{\mathrm{CT}_1} & \in \mathbb{R}^2 & \text{Euler angles for } \boldsymbol{u}_1 \text{ orientation in the } \mathsf{T}_1 \text{ frame} \\ \rho_1 & \in \mathbb{R} & \text{inverse distance from camera optical center to feature 1} \\ \vdots & & \\ \boldsymbol{e}_{\mathrm{CT}_n} & \in \mathbb{R}^2 & \\ \rho_n & \in \mathbb{R} & \end{array}\right]$$

Here, R_{CT_k} and $\boldsymbol{e}_{\mathrm{CT}_k}$ for $k = 1, ..., n$ are to be interpreted such that

$$\boldsymbol{u}_{k,\mathrm{C}} = R_{\mathrm{CT}_k}\boldsymbol{e}_3 = \bar{R}_{\mathrm{CT}_k} R\left(\left[\begin{array}{c} \boldsymbol{e}_{\mathrm{CT}_k} \\ 0 \end{array}\right]\right)\boldsymbol{e}_3$$

where $\bar{R}_{\mathrm{CT}_k}$ is the reference T$_k$-to-C attitude matrix.

IMU Measurements

Propagation of the central pose estimator's state is based on a model replacement approach in which the IMU's specific force and angular rate measurements drive state propagation in place of a vehicle dynamics model. A 6-dimensional vector of IMU specific force and angular rate measurements

$$\boldsymbol{u} = \begin{bmatrix} \tilde{\boldsymbol{f}}_{\mathrm{U}} \in \mathbb{R}^3 \\ \tilde{\boldsymbol{\omega}}_{\mathrm{U}} \in \mathbb{R}^3 \end{bmatrix}$$

is fed to the estimator at each time update, which occurs at a rate between 70 and 120 Hz depending on the IMU configuration.

Let $R_{\mathrm{UB}}, R_{\mathrm{BW}}, R_{\mathrm{WG}} \in SO(3)$ be matrices representing attitude transformations between the indicated frames; $S_a, S_g \in \mathbb{R}^{3\times3}$ be diagonal matrices that account for scale factor errors in the primary axes of the IMU's accelerometer and gyroscope, respectively; $\boldsymbol{a}_{\mathrm{G}} \in \mathbb{R}^3$ be the acceleration of the B frame with respect to the quasi-inertial W frame; $\boldsymbol{c}_{\mathrm{G}} \in \mathbb{R}^3$ be the centripetal acceleration of the B frame due to Earth rotation; $\boldsymbol{g}_{\mathrm{G}} \in \mathbb{R}^3$ be the acceleration due to gravity at the B frame origin; $\boldsymbol{\omega}_{\mathrm{B}} \in \mathbb{R}^3$ be the angular rate of B with respect to W; $\dot{\boldsymbol{\omega}}_{\mathrm{B}} \in \mathbb{R}^3$ be the time derivative of $\boldsymbol{\omega}_{\mathrm{B}}$; $\boldsymbol{\omega}_{\mathrm{EG}}$ be the Earth angular rate with respect to the inertial frame; $\boldsymbol{r}_{\mathrm{BU,B}} \in \mathbb{R}^3$ be the (fixed) "lever arm," the vector pointing from B to U; $\boldsymbol{b}_{a0\text{U}}, \boldsymbol{b}_{g0\text{U}} \in \mathbb{R}^3$ be the static accelerometer and gyroscope biases, respectively; and $\boldsymbol{v}_{a\text{U}}, \boldsymbol{v}_{g\text{U}} \in \mathbb{R}^3$ be zero-mean white Gaussian random processes modeling accelerometer and gyroscope measurement noise, respectively.

With these preliminaries, the accelerometer measurement model can be introduced as

$$\tilde{\boldsymbol{f}}_{\text{U}} = S_a R_{\text{UB}} \left[R_{\text{BW}} R_{\text{WG}} \left(\boldsymbol{a}_{\text{G}} + \boldsymbol{c}_{\text{G}} - \boldsymbol{g}_{\text{G}} \right) + \dot{\boldsymbol{\omega}}_{\text{B}} \times \boldsymbol{r}_{\text{BU,B}} + \boldsymbol{\omega}_{\text{B}} \times \left(\boldsymbol{\omega}_{\text{B}} \times \boldsymbol{r}_{\text{BU,B}} \right) \right] + \boldsymbol{b}_{a0\text{U}} + \boldsymbol{b}_{a\text{U}} + \boldsymbol{v}_{a\text{U}}$$

This model makes two approximations. First, it treats W as an inertial frame apart from explicitly accounting for the centripetal force $\boldsymbol{c}_{\text{G}}$. Second, it neglects the Coriolis acceleration due to platform velocity and Earth rotation. These approximations are acceptable when dealing with MEMS-grade inertial sensors, whose bias variations tend to be much larger than the neglected effects.

The gyroscope measurement model is

$$\tilde{\boldsymbol{\omega}}_{\text{U}} = S_g R_{\text{UB}} \left(\boldsymbol{\omega}_{\text{B}} + R_{\text{BW}} R_{\text{WG}} \boldsymbol{\omega}_{\text{EG}} \right) + \boldsymbol{b}_{g0\text{U}} + \boldsymbol{b}_{g\text{U}} + \boldsymbol{v}_{g\text{U}}$$

The variable bias terms $\boldsymbol{b}_{a\text{U}}$ and $\boldsymbol{b}_{g\text{U}}$, which are elements of the pose estimator state, are modeled as Ornstein-Uhlenbeck random processes

$$\dot{\boldsymbol{b}}_{a\text{U}} = -\tfrac{1}{\tau_a} \boldsymbol{b}_{a\text{U}} + \tilde{\boldsymbol{v}}_{a\text{U}}, \quad \dot{\boldsymbol{b}}_{g\text{U}} = -\tfrac{1}{\tau_g} \boldsymbol{b}_{g\text{U}} + \tilde{\boldsymbol{v}}_{g\text{U}}$$

with decorrelation times $\tau_a, \tau_g > 0$ and noise processes $\boldsymbol{v}_{a2\text{U}}, \boldsymbol{v}_{g2\text{U}} \mathbb{R}^3$, which are modeled as zero-mean white Gaussian processes. The attitude matrix R_{BW} is also contained in the state in the sense that $\boldsymbol{e}_{\text{BW}}$ is a state element and $R_{\text{BW}} = R(\boldsymbol{e}_{\text{BW}}) \bar{R}_{\text{BW}}$. The static bias terms, scale factor matrices, and other attitude matrices found in the above models are determined by calibration using an offline 24-state version of the pose estimation UKF. These quantities are made observable by MAV maneuvering under conditions of good GNSS visibility [37]. Observability is further strengthened by the attitude information provided by the attitude RTK solution $\boldsymbol{r}_{\text{PS,W}}$. The lever arm $\boldsymbol{r}_{\text{BU,B}}$ is measured using a CAD model of the MAV. Note that by defining the state such that the position $\boldsymbol{r}_{\text{W}}$ is taken to be that of the IMU, the lever arm vanishes in state propagation, which eliminates the need to measure or estimate $\dot{\boldsymbol{\omega}}_{\text{B}}$.

Details about the variance and correlation properties of the various noise processes involved in the IMU models above, or their discrete-time versions, together with a mapping from standard IMU parameter values available from data sheets, are found in [38]. The specific values applied for this paper are from the Automotive IMU class in [38].

State Propagation

State propagation within the pose estimator is based on a nonlinear dynamics function $\boldsymbol{f} : \mathbb{R}^{15} \times \mathbb{R}^{6} \times \mathbb{R}^{15} \to \mathbb{R}^{15}$:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v})$$

where $\boldsymbol{x}$ and $\boldsymbol{u}$ have been previously defined, and where

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_{a\text{U}} \\ \tilde{\boldsymbol{v}}_{a\text{U}} \\ \boldsymbol{v}_{g\text{U}} \\ \tilde{\boldsymbol{v}}_{g\text{U}} \\ \boldsymbol{v}_{\text{CB}} \end{bmatrix} \in \mathbb{R}^{15}$$

is the process noise vector, with $\boldsymbol{v}_{\text{CB}}$ being a zero-mean white Gaussian process modeling the process noise associated with the B-to-C attitude error. The coupled dynamical equations in $\boldsymbol{f}$ may be grouped by state sector:

Rigid-Body Sector:

$$\begin{aligned} \dot{\boldsymbol{r}}_{\text{W}} &= \boldsymbol{v}_{\text{W}} \\ \dot{\boldsymbol{v}}_{\text{W}} &= R_{\text{WB}} R_{\text{BU}} S_a^{-1} \left(\tilde{\boldsymbol{f}}_{\text{U}} - \boldsymbol{b}_{a0\text{U}} - \boldsymbol{b}_{a\text{U}} - \boldsymbol{v}_{a\text{U}} \right) + R_{\text{WG}} \left(\boldsymbol{c}_{\text{G}} - \boldsymbol{g}_{\text{G}} \right) \\ \dot{R}_{\text{BW}} &= - \left[\left(R_{\text{BU}} S_g^{-1} \left(\tilde{\boldsymbol{\omega}}_{\text{U}} - \boldsymbol{b}_{g0\text{U}} - \boldsymbol{b}_{g\text{U}} - \boldsymbol{v}_{g\text{U}} \right) - R_{\text{BW}} R_{\text{WG}} \boldsymbol{\omega}_{\text{EG}} \right) \times \right] R_{\text{BW}} \end{aligned}$$

IMU Bias Sector:

$$\begin{aligned} \dot{\boldsymbol{b}}_{a\text{U}} &= -\tfrac{1}{\tau_a} \boldsymbol{b}_{a\text{U}} + \tilde{\boldsymbol{v}}_{a\text{U}} \\ \dot{\boldsymbol{b}}_{g\text{U}} &= -\tfrac{1}{\tau_g} \boldsymbol{b}_{g\text{U}} + \tilde{\boldsymbol{v}}_{g\text{U}} \end{aligned}$$

Vision Sector:

$$\begin{aligned}
\dot{\boldsymbol{e}}_{\text{CB}} &= \boldsymbol{v}_{\text{CB}} \\
\dot{\boldsymbol{e}}_{\text{CT}_k} &= \left[\bar{R}_{\text{CT}_k}(\boldsymbol{\omega}_{\text{C}} - \boldsymbol{u}_{k,\text{C}}\boldsymbol{u}_{k,\text{C}}^{\mathsf{T}}\boldsymbol{\omega}_{\text{C}} + \rho_k \boldsymbol{u}_{k,\text{C}} \times \boldsymbol{v}_{\text{C}})\right]_{1:2}, \quad k = 1, ..., n \\
\dot{\rho}_k &= \rho_k^2 \boldsymbol{v}_{\text{C}}^{\mathsf{T}} \boldsymbol{u}_{k,\text{C}}, \quad k = 1, ..., n
\end{aligned}$$

where $[\cdot]_{1:2}$ indicates the vector made of the first two elements of the contained vector, and where

$\boldsymbol{v}_{\text{C}} \in \mathbb{R}^3$ is the linear velocity of C with respect to W, expressed in C
$\boldsymbol{\omega}_{\text{C}} \in \mathbb{R}^3$ is the angular velocity of C with respect to W, expressed in C

For mechanization within the pose estimator's UKF, the above ordinary differential equations are time-discretized by numerical integration.

Note that, in the robocentric feature formulation, image features, which are tracked in the C frame, are modeled as fixed in the W frame, and thus must move in the C frame in order to counter the motion of the camera. The update equations are constructed so that, for the kth feature, only the two elements of $\boldsymbol{e}_{\text{CT}_k}$ need be stored, resulting in a minimal parameterization of feature positions.

CDGNSS Measurements

Position and attitude CDGNSS solutions are treated as measurements by the central pose estimator. The position CDGNSS estimator produces an estimate of $\boldsymbol{r}_{\text{RP,G}}$, the location of the primary GNSS antenna in the G frame. This estimate, transformed to the W frame, is taken as a measurement $\boldsymbol{z}_p$, modeled as

$$\boldsymbol{z}_p = R_{\text{WG}}\hat{\boldsymbol{r}}_{\text{RP,G}} = \boldsymbol{h}_p(\boldsymbol{x}) + \boldsymbol{w}_p = \boldsymbol{r}_{\text{W}} - R_{\text{WB}}\boldsymbol{r}_{\text{BU,B}} + \boldsymbol{w}_p$$

where $\boldsymbol{w}_p \in \mathbb{R}^3$ is a zero-mean white Gaussian random process with covariance $P_{\text{RP,W}}$ that models measurement noise, and $\boldsymbol{r}_{\text{BU,B}}$ is the IMU lever arm.

The attitude CDGNSS estimator produces an estimate of $\boldsymbol{r}_{\text{PS,G}}$, the known-length vector from the vehicle's primary antenna to its secondary antenna, expressed in G. This estimate, which provides pitch and yaw information to the pose estimator, is normalized to unity length and transformed to the W frame to become the measurement $\boldsymbol{z}_a$, modeled as

$$\boldsymbol{z}_a = R_{\text{WG}}\hat{\boldsymbol{r}}_{\text{PS,G}}^u = \boldsymbol{h}_a(\boldsymbol{x}) + \boldsymbol{w}_a = \boldsymbol{R}_{\text{WB}}\boldsymbol{e}_1 + \boldsymbol{w}_a$$

where $r_{\text{PS,G}}^u$ is the unit-normalized estimate vector, and $\boldsymbol{w}_p \in \mathbb{R}^3$ is a zero-mean white Gaussian random process with covariance $P_{\text{PS,W}}$ that models measurement noise. The QUEST measurement error covariance formulation is applied for $P_{\text{PS,W}}$ to account for the unity length constraint [39]. Both $P_{\text{RP,W}}$ and $P_{\text{SP,W}}$ are inflated by a factor of between 1.75 and 4 as they are ingested into the pose estimator to account for time correlation in $\boldsymbol{w}_p$ and $\boldsymbol{w}_a$ due to multipath.

The relatively weak CDGNSS geometric constraint in the vertical and the platform's short inter-antenna baseline tend to leave $\hat{\boldsymbol{r}}_{\text{PS,G}}$ with poor vertical accuracy. Accordingly, the pose estimator may be configured to optionally take in only the yaw angle from $\hat{\boldsymbol{r}}_{\text{PS,G}}$. This angle is expressed as an anomaly relative to a reference yaw angle $\bar{z}_y$ to avoid discontinuity as the yaw wraps on the range $[-\pi, \pi)$ rad. Let $\hat{\boldsymbol{r}}_{\text{PS,W}} = R_{\text{WG}}\hat{\boldsymbol{r}}_{\text{PS,G}}$. Then the scalar yaw measurement is modeled as

$$z_y = \text{atan2}\left([\hat{\boldsymbol{r}}_{\text{PS,W}}]_1, [\hat{\boldsymbol{r}}_{\text{PS,W}}]_2\right) - \bar{z}_y = h_y(\boldsymbol{x}) + w_y = \text{atan2}\left([R_{\text{WB}}\boldsymbol{r}_{\text{PS,B}}]_1, [R_{\text{WB}}\boldsymbol{r}_{\text{PS,B}}]_2\right) - \bar{z}_y + w_y$$

where $w_y \in \mathbb{R}$ is a zero-mean Gaussian random process with standard deviation of approximately 0.1 rad modeling error in z_y.

Visual Measurements

Upon the arrival of a camera frame, a direct pixel-intensity-based measurement update step is performed in sequence for each tracked visual feature expected to be entirely in the camera frame. For the kth tracked feature, a matrix of patch pixels P_k is pre-warped using an affine warping matrix to account for, to first order, the effects of the change in camera pose between when P_k was captured and the current epoch. Adopting the ROVIO direct intensity measurement formulation [18], a linearized model A_k for the expected pixel intensity error vector $\tilde{\boldsymbol{b}}_k$ is generated using the gradient

$\boldsymbol{g} \in \mathbb{R}^2 = [g_x, g_y]^\mathsf{T}$ of the camera image I, represented as a matrix. The gradient is evaluated at each of the patch's expected pixel coordinates, and the corresponding pixel intensity error is modeled as having arisen from a 2-dimensional pixel alignment error $\tilde{\boldsymbol{p}}_k = \boldsymbol{p}_k - \hat{\boldsymbol{p}}_k$, where $\boldsymbol{p}_k \in \mathbb{R}^2$ contains the true pixel coordinates (with sub-pixel resolution) of the center of the kth patch in I:

$$\underbrace{\begin{bmatrix} P_k(1,1) - I(\hat{\boldsymbol{p}} + (1,1)) \\ P_k(1,2) - I(\hat{\boldsymbol{p}} + (1,2)) \\ \vdots \end{bmatrix}}_{\tilde{\boldsymbol{b}}_k} = \underbrace{\begin{bmatrix} g_x(\hat{\boldsymbol{p}} + (1,1)) & g_y(\hat{\boldsymbol{p}} + (1,1)) \\ g_x(\hat{\boldsymbol{p}} + (1,2)) & g_y(\hat{\boldsymbol{p}} + (1,2)) \\ \vdots & \vdots \end{bmatrix}}_{A_k} \tilde{\boldsymbol{p}}_k$$

This model is essentially a first-order Taylor expansion of the pixel intensity error of the image [31]. The measurement equations for 8×8 pixel patches at each of the 4 scale levels considered are vertically concatenated to construct the final $8 \times 8 \times 4$-dimensional measurement. This measurement would be computationally intractable to directly incorporate into the pose estimator's measurement update, but it can be reduced to an equivalent 2-dimensional measurement by QR factorization, since it is modeled as having been produced by a 2-dimensional pixel alignment error $\tilde{\boldsymbol{p}}_k$ with Gaussian pixel intensity measurement noise $\boldsymbol{w}_{p,k}$. The factorization yields

$$\begin{aligned} \tilde{\boldsymbol{b}}_k &= A_k \tilde{\boldsymbol{p}}_k = Q_k R_k \tilde{\boldsymbol{p}}_k \\ &= \begin{bmatrix} Q_{1k} & Q_{2k} \end{bmatrix} \begin{bmatrix} R_{1k} \\ 0 \end{bmatrix} \tilde{\boldsymbol{p}}_k \end{aligned}$$

which leads to the following model for the patch offset measurement $\boldsymbol{z}_k \in \mathbb{R}^2$:

$$\boldsymbol{z}_k = Q_{1k}^\mathsf{T} \tilde{\boldsymbol{b}}_k = \boldsymbol{h}_k(\boldsymbol{x}) + \boldsymbol{w}_{p,k} = R_{k1} \tilde{\boldsymbol{p}}_k + \boldsymbol{w}_{p,k}$$

Each new camera frame is also searched for suitable visual features to be added to the filter state vector. A first pass is made using the FAST feature detector [40], and candidate multilevel patches are extracted from pixel locations with high FAST scores. The update step constructs the A matrix from the measurement, and a feature trackability score:

$$\text{score}(A) = \text{tr}(A^\mathsf{T} A)$$

Features with a score greater than a predefined threshold are accepted and initialized in the filter state vector. This score is essentially equivalent to the well-known Shi-Tomasi feature score [31], but applied at multiple scale levels. It also depends on the sum of the two eigenvalues of $A^\mathsf{T} A$ rather than the minimum eigenvalue, which enables the ROVIO tracking formulation to track patches that may only provide strong measurements in a single direction. A culling mechanism prevents new features from being initialized if they may overlap with existing features. If multiple new features with scores above the threshold overlap, only the one with the highest score is accepted into the estimator state.

Outlier Exclusion

A χ^2-type innovations test is applied to all incoming visual and CDGNSS measurements. The test helps to reject outlier visual measurements, which can be caused by reflective or partially-transparent surfaces, features without strong gradients, or the MAV tracking its own shadow. The innovations test also aids in rejecting false RTK fixes, especially when when the estimator is operating in the loosely-coupled mode.

EXPERIMENTAL RESULTS AND ANALYSIS

The following subsections analyze the system's performance on datasets containing simulated and real-world GNSS measurement degradations. For each of the datasets introduced below, the system was run in four modes, which are described in the figures and tables with the following abbreviations:

LC I	*loosely-coupled* mode with only RTK and inertial measurements
LC V+I	*loosely-coupled* mode with RTK, visual, and inertial measurements
TC I	*tightly-coupled* mode with only RTK and inertial measurements
TC V+I	*tightly-coupled* mode with RTK, visual, and inertial measurements

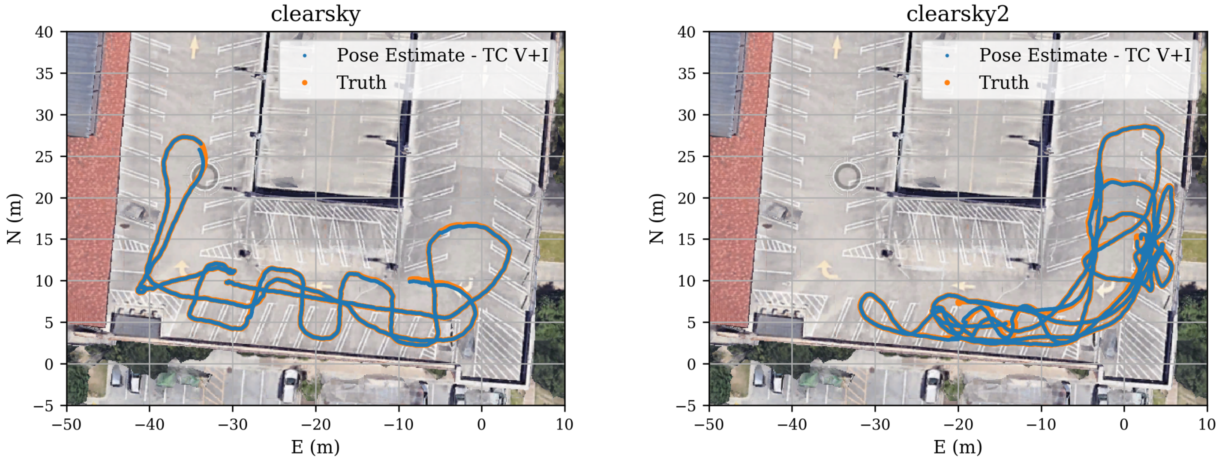

Fig. 4: Ground track of the `clearsky` and `clearsky2` datasets.

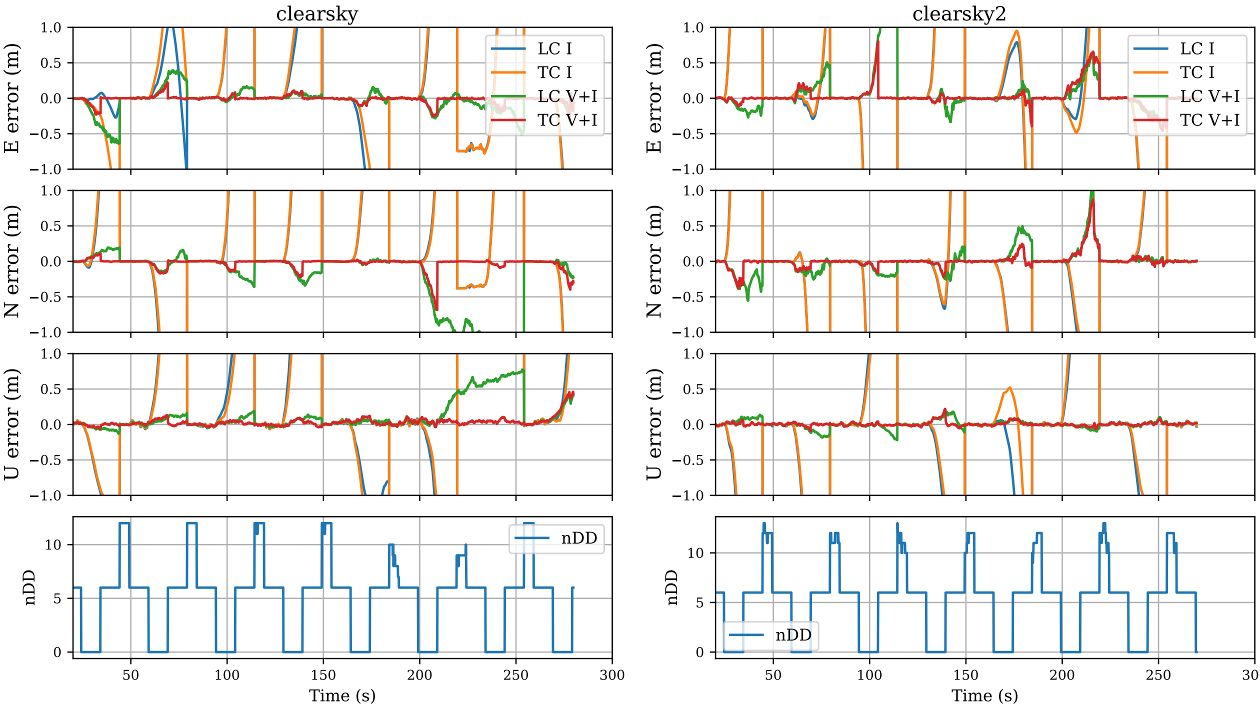

Fig. 5: Positioning errors in the ENU frame with respect to clear-sky truth data for runs with periodic simulated outages for `clearsky` and `clearsky2` datasets. `nDD` indicates the number of double-difference measurements available to the RTK position estimator. Due to the length and depth of the outages, in all of the inertial-sensing-only cases the pose estimate rapidly diverges from the truth. Tight coupling with visual-inertial pose estimation allows a confident integer fix as a few GNSS measurements return, but the loosely-coupled estimator must wait until the outage ends entirely.

Simulated GNSS Degradations

For the results presented in this section, data from the sensors on board the MAV were collected from an area with a clear view of the sky. The loosely-coupled RTK-and-inertial pose solutions from these runs were taken as “truth,” as they had a large number of high-quality double-difference measurements available with high integer aperture test statistics. Next, artificial GNSS signal outages were introduced in the processing pipeline, temporarily excluding some or all of the received GNSS signals from processing. These artificial outages were used to evaluate the system’s positioning

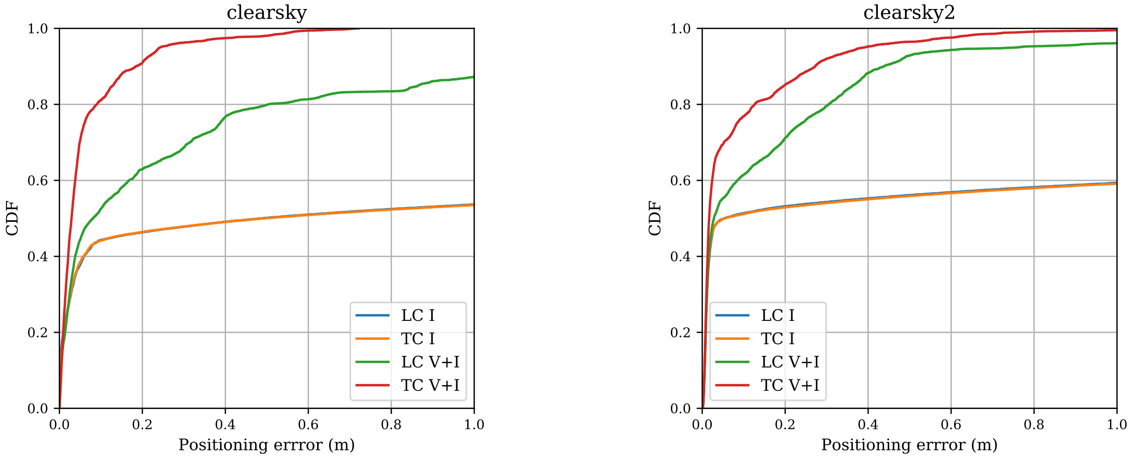

Fig. 6: Empirical CDF of positioning error with respect to truth data for runs with simulated outages on the `clearsky` and `clearsky2` datasets.

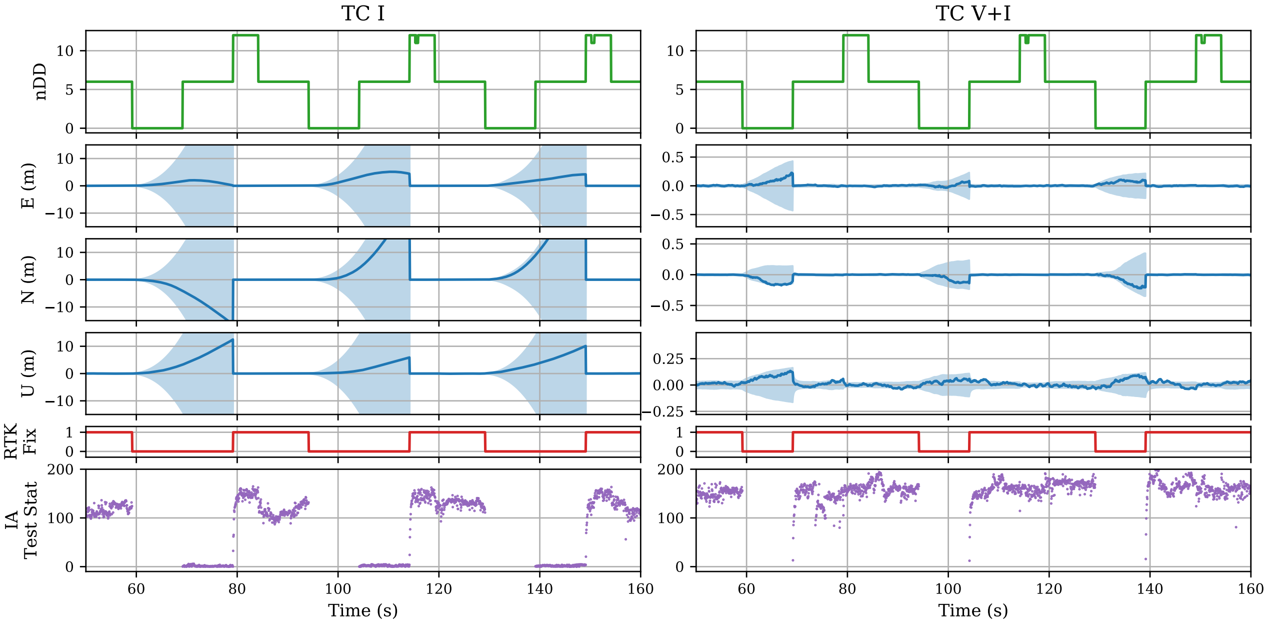

Fig. 7: Detail of 3 simulated outages on the `clearsky` dataset, when the RTK estimators are tightly coupled to inertial-only and visual-inertial pose estimation. Position errors of the pose estimate in the 3 world axes are shown, along with the estimator's reported covariance. Shading indicates the pose estimator's 3σ confidence interval. Note the change in the error y-axis between the two error plots. The precise pose estimate provided by visual-inertial positioning constrains the RTK estimator enough to produce an integer fix when emerging from a total or near-total outage into a partial outage.

accuracy and ability to retain an RTK fix in degraded GNSS conditions. The outage pattern, roughly simulating passages below an occluding structure, consisted of 10 seconds of restriction to 6 double-difference measurements, 10 seconds of zero measurements, another 10 seconds of restriction to 6 measurements, followed by 5 seconds of no degradation before repeating.

The pose estimation system was run in the four different modes on datasets named `clearsky` and `clearsky2`. The `clearsky` dataset features gentle motion, with velocities limited to $1.0\,\mathrm{m/s}$ along any axis, and angular rates limited to $45°/\mathrm{s}$. The `clearsky2` dataset featured more rapid motion and sharp turns, with velocities up to 2.0 m/s, and angular rates up to $100°/\mathrm{s}$. Fig. 5 shows the resulting positioning error in the four modes during these simulated

Dataset	Mode	Position RTK		Pose
		Availability	False Fix Rate	RMS Error
clearksy	LC I	45.9%	5.3%	9.0 m
	LC V+I	45.9%	5.3%	0.57 m
	TC I	45.9%	5.3%	8.9 m
	TC V+I	**71.2%**	**0.0%**	**0.12 m**
clearsky2	LC I	48.2%	0.0%	12.9 m
	LC V+I	48.2%	0.0%	0.36 m
	TC I	48.2%	0.0%	13.1 m
	TC V+I	**62.9%**	**0.0%**	**0.18 m**

TABLE I: Position RTK availability and false fix rate, and pose estimate RMS position error over the datasets with simulated outages. A false fix was declared when the RTK estimator reported a fixed solution more than 15cm away from the truth position. For both datasets, visual-inertial tight coupling had the highest RTK availability and lowest positioning error.

degradations in both datasets; Fig. 6 shows the corresponding empirical CDF of positioning errors.

This type of GNSS impairment—a partial outage, followed by a complete outage, followed by a partial outage—provides a clear demonstration of the advantages of visual-inertial tight coupling, as explored in Fig. 7. Inertial-only pose propagation with even a smartphone-grade IMU is perfectly capable of providing a centimeter-level accurate pose estimate over deltas of a fraction of a second, which is helpful in retaining a fix during degraded conditions with sharp antenna movements. However, once the RTK fix is lost, be it due to a complete outage, set of measurements corrupted by multipath, or a sudden shift in visible satellites, the IMU-only position estimate rapidly degrades, and after just a couple of seconds no longer provides a useful constraint for RTK integer ambiguity resolution. An IMU-only system must then wait until a sufficient (and usually large) number of satellites is reacquired before it is able to confidently provide an RTK fix again. In contrast, a visual-inertial pose estimator is often able to retain centimeter accuracy indefinitely during an outage, as its drift is primarily determined by the distance traveled rather than time elapsed. After emerging from an outage into still-degraded GNSS conditions, the visual-inertial estimator often still has a strong enough position constraint to improve the RTK fixing success rate.

In both simulated outage datasets, the tightly-coupled visual-inertial estimator was usually able to retain an integer fix during the partial outage following a complete outage, while the loosely-coupled and IMU-only estimators are never able to achieve an integer fix until the simulated degradation ended completely. This caused a large increase in both integer fix rate and positioning accuracy for the tightly-coupled visual-inertial estimator over the other modes, as can be seen in Fig. 6.

Real-World Test

Next, data from the sensors onboard the MAV were collected in challenging real-world GNSS conditions over several sessions on the roof of a parking garage next to a building with an overhanging eave.

In the Eave dataset, the rover makes several passes from an area with a clear view of the sky to underneath the eave, causing outages of all but a few GNSS signals, then moves back to an area with a clear sky view. In the Garage dataset, the rover makes several loops which take it between an area with a somewhat obstructed view of the sky and an area covered by overhanging parking garage structure, which induces essentially total GNSS outages. This dataset additionally provided a challenge for the visual-inertial navigation due to the large contrast changes encountered when entering and leaving the shadows of the overhanging garage. Ground tracks for both datasets are shown in Fig. 8.

Both datasets take place in challenging GNSS environments. Due to the lack of a clear sky view in these datasets, no truth reference is available. Therefore, they are analyzed via metrics relating to the estimators' internal consistency. Table I shows the availability of fixed-integer RTK position solutions which additionally passed the central pose estimator innovations test. In all inertial-only cases, false RTK integer fixes were seen in the positioning output. In the loosely-coupled cases, the availability decreased when visual measurements were added, due to the central pose

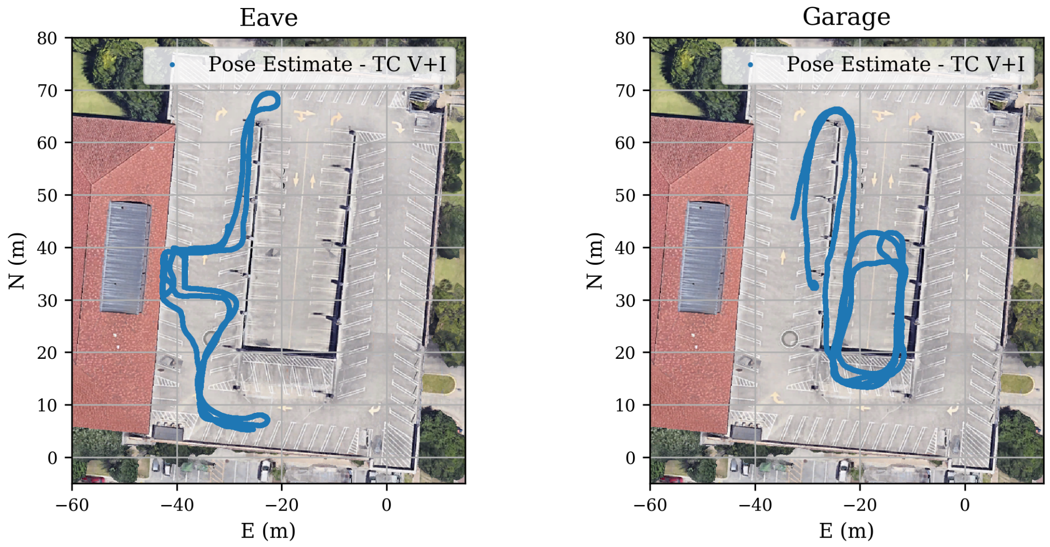

Fig. 8: Ground tracks of the `Eave` and `Garage` datasets.

Dataset	Mode	RTK Avail.	Worst-Case Discontinuity
`Eave`	LC I	<56.8%	166 m
	LC V+I	49.6%	1.11 m
	TC I	<56.8%	166 m
	TC V+I	**75.5%**	**0.57 m**
`Garage`	LC I	<59.6%	443 m
	LC V+I	27.2%	6.27 m
	TC I	<55.8%	420 m
	TC V+I	**72.7%**	**1.56 m**

TABLE II: Availability of fixed RTK solutions passing pose estimator innovations test and worst-case position discontinuity on RTK re-acquisition in the real-world datasets. "<" indicates false integer fixes were seen in these runs (as shown in Fig. 9) that the pose estimator was unable to reject. In the inertial-only modes, the pose estimator usually does not have a strong enough position prior during outages to reject false fixes.

estimator's increased ability to reject these incorrect integer fixes via innovations testing. In both datasets, integer fixing rates greatly increased when the RTK estimators ran in visual-inertial tightly-coupled mode.

Table I shows, for each run, the largest position jump that occurred on the re-acquisition of an innovations-accepted position RTK fix following an outage. Assuming the RTK integer states have been correctly fixed on re-acquisition, this is effectively a measure of the unaided pose estimator drift over the course of the RTK outage. In both the `garage` and `eave` datasets, multi-second RTK fix outages occurred in all 4 filter modes due to the challenging GNSS environment. These outages were often too long for the pose estimator to provide a useful position estimate when solely using the smartphone-grade IMU. In contrast, when visual measurements are allowed into the pose estimator, sub-meter accuracy was retained in almost every case over the course of these outages. When the estimator was run in the tightly-coupled mode, this extra constraint allowed earlier RTK fixing at the tail end of an outage, when fewer double-difference measurements were available, increasing the overall solution accuracy.

In the `garage` dataset, the RTK position estimator spent a large amount of time with an incorrect integer fix in all runs except with tightly coupled visual-inertial positioning. For all modes except tightly-coupled visual-inertial, a heuristic based on the number of DD measurements would normally prevent attempting an integer fix under these conditions. This restriction was disabled for these tests for the sake of fair comparison in environments with few available measurements. When using only the IMU, the pose estimator was unable to confidently reject these false fixes due to its accumulated

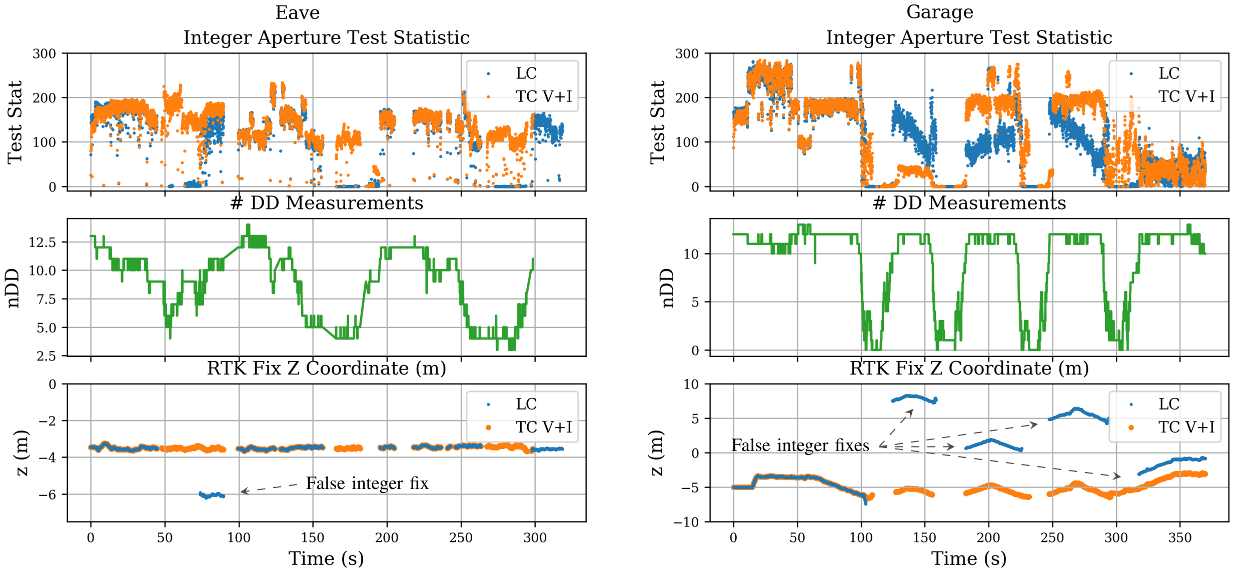

Fig. 9: Plots of the integer aperture test statistics, number of DD measurements, and Z coordinate of the `Eave` and `Garage` datasets, for the RTK position estimator as run in the unaided (loosely-coupled) and visual-inertial tightly-coupled modes. An increased availability of RTK fixes is noted when tightly coupled. As the tests generally took place with the MAV a fixed distance from the ground, false RTK fixes are readily apparent in the plots of the RTK fixed-integer position Z coordinate in both datasets. Normally, the loosely-coupled RTK estimator would be run with a restriction on the minimum number of double-difference measurements, which would prevent these false fixes. For the sake of comparison, this restriction was removed for these tests. Note the decrease of GNSS availability to effectively zero as the rover passes underneath the parking garage structure in the `Garage` dataset.

position uncertainty, and the pose estimate converged to the incorrect RTK position. When running in loosely-coupled mode with visual measurements, the pose estimator correctly rejected the incorrectly-fixed RTK position measurement as it failed the pose estimator's innovations test, but suffered from increased odometric drift as it consequently spent more time without valid RTK measurements. In contrast, when running in tightly-coupled mode with visual measurements, the enhanced position prior provided by visual-inertial pose estimate allowed the RTK estimator to produce a correctly fixed position measurement, providing increased RTK availability and therefore positioning accuracy.

CONCLUSIONS

This paper has described and evaluated a multi-antenna carrier phase differential GNSS (CDGNSS) position and orientation (pose) determination system which uses camera images and measurements from a smartphone-grade inertial sensor to aid the integer ambiguity resolution process by providing it with a pose prior. The system was tested on datasets collected by the onboard sensors of a low-cost micro aerial vehicle. Performance was evaluated over intervals of both simulated and real-world GNSS measurement degradation. Results showed that incorporating visual measurements into a tightly-coupled inertial-CDGNSS system is critical to maintaining a pose estimate accurate enough to support an integer fix when emerging from complete or near-complete GNSS measurement outages.

ACKNOWLEDGMENTS

This work was supported in part by the University of Texas Wireless Networking and Communications Group (WNCG), and in part by the Army Research Office under Cooperative Agreement W911NF-19-2-0333. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] S. Jung, S. Song, S. Kim, J. Park, J. Her, K. Roh, and H. Myung, "Toward autonomous bridge inspection: A framework and experimental results," in *2019 16th International Conference on Ubiquitous Robots (UR)*. IEEE, 2019, pp. 208–211.

[2] T. Bilis, T. Kouimtzoglou, M. Magnisali, and P. Tokmakidis, "The use of 3D scanning and photogrammetry techniques in the case study of the roman theatre of nikopolis. surveying, virtual reconstruction and restoration study." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, no. 2/W3, 2017.

[3] J. Qi, D. Song, H. Shang, N. Wang, C. Hua, C. Wu, X. Qi, and J. Han, "Search and rescue rotary-wing UAV and its application to the Lushan MS 7.0 earthquake," *Journal of Field Robotics*, vol. 33, no. 3, pp. 290–321, 2016.

[4] T. E. Humphreys, R. X. T. Kor, and P. A. Iannucci, "Open-world virtual reality headset tracking," in *Proceedings of the ION GNSS+ Meeting*, Online, 2020.

[5] M. Psiaki and S. Mohiuddin, "Global positioning system integer ambiguity resolution using factorized least-squares techniques," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 2, pp. 346–356, March-April 2007.

[6] S. Verhagen, P. J. Teunissen, and D. Odijk, "The future of single-frequency integer ambiguity resolution," in *VII Hotine-Marussi Symposium on Mathematical Geodesy*. Springer, 2012, pp. 33–38.

[7] F. Zimmermann, C. Eling, L. Klingbeil, and H. Kuhlmann, "Precise positioning of UAVs-dealing with challenging RTK-GPS measurement conditions during automated UAV flights." *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 4, 2017.

[8] M. Petovello, M. Cannon, and G. Lachapelle, "Benefits of using a tactical-grade IMU for high-accuracy positioning," *Navigation, Journal of the Institute of Navigation*, vol. 51, no. 1, pp. 1–12, 2004.

[9] B. M. Scherzinger, "Precise robust positioning with inertially aided RTK," *Navigation*, vol. 53, no. 2, pp. 73–83, 2006.

[10] H. T. Zhang, "Performance comparison on kinematic GPS integrated with different tactical-grade IMUs," Master's thesis, The University of Calgary, Jan. 2006.

[11] S. Kennedy, J. Hamilton, and H. Martell, "Architecture and system performance of SPAN—NovAtel's GPS/INS solution," in *Position, Location, And Navigation Symposium, 2006 IEEE/ION*. IEEE, 2006, p. 266.

[12] S. Godha, G. Lachapelle, and M. Cannon, "Integrated GPS/INS system for pedestrian navigation in a signal degraded environment," in *Proceedings of the ION GNSS Meeting*, vol. 2006, 2006.

[13] A. Angrisano, M. Petovello, and G. Pugliano, "Benefits of combined GPS/GLONASS with low-cost MEMS IMUs for vehicular urban navigation," *Sensors*, vol. 12, no. 4, pp. 5134–5158, 2012.

[14] T. Li, H. Zhang, Z. Gao, Q. Chen, and X. Niu, "High-accuracy positioning in urban environments using single-frequency multi-GNSS RTK/MEMS-IMU integration," *Remote Sensing*, vol. 10, no. 2, p. 205, 2018.

[15] R. Hirokawa and T. Ebinuma, "A low-cost tightly coupled GPS/INS for small uavs augmented with multiple GPS antennas," *Navigation*, vol. 56, no. 1, pp. 35–44, 2009.

[16] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, 2011.

[17] G. Huang, "Visual-inertial navigation: A concise review," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 9572–9582.

[18] M. Bloesch, M. Burri, S. Omari, M. Hutter, and R. Siegwart, "Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1053–1072, 2017.

[19] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3565–3572.

[20] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.

[21] Z. Huai and G. Huang, "Robocentric visual-inertial odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 6319–6326.

[22] D. P. Shepard and T. E. Humphreys, "High-precision globally-referenced position and attitude via a fusion of visual SLAM, carrier-phase-based GPS, and inertial measurements," in *Proceedings of the IEEE/ION PLANS Meeting*, May 2014.

[23] P. Henkel, A. Blum, and C. Günther, "Precise RTK positioning with GNSS, INS, barometer and vision," in *Proceedings of the 30th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2017)*, 2017, pp. 2290–2303.

[24] T. Li, H. Zhang, Z. Gao, X. Niu, and N. El-Sheimy, "Tight fusion of a monocular camera, mems-imu, and single-frequency multi-gnss rtk for precise navigation in GNSS-challenged environments," *Remote Sensing*, vol. 11, no. 6, p. 610, 2019.

[25] P. Henkel, A. Sperl, U. Mittmann, R. Bensch, P. Färber, and C. Günther, "Precise positioning of robots with fusion of GNSS, INS, odometry, barometer, local positioning system and visual localization," in *Proc. of the 31st Intern. Technical Meeting of The Satellite Division of the Institute of Navigation*, 2018, pp. 3078–3087.

[26] T. E. Humphreys, L. Narula, and M. J. Murrian, "Deep urban unaided precise Global Navigation Satellite System vehicle positioning," *IEEE Intelligent Transportation Systems Magazine*, 2020.

[27] B. Scherzinger, "Quasi-tightly coupled GNSS-INS integration," *Navigation: Journal of The Institute of Navigation*, vol. 62, no. 4, pp. 253–264, 2015.

[28] L. Wang and S. Verhagen, "A new ambiguity acceptance test threshold determination method with controllable failure rate," *Journal of Geodesy*, vol. 89, no. 4, pp. 361–375, 2015. [Online]. Available: http://dx.doi.org/10.1007/s00190-014-0780-2

[29] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: John Wiley and Sons, 2001.

[30] P. J. Teunissen, "The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation," *Journal of Geodesy*, vol. 70, no. 1-2, pp. 65–82, 1995.

[31] J. Shi *et al.*, "Good features to track," in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. IEEE, 1994, pp. 593–600.

[32] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJCAI'81. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1981, pp. 674–679.

[33] H. Strasdat, J. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" *Image and Vision Computing*, 2012.

[34] J. Delmerico and D. Scaramuzza, "A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2502–2509.

[35] J. M. Montiel, J. Civera, and A. J. Davison, "Unified inverse depth parametrization for monocular SLAM." Robotics: Science and Systems, 2006.

[36] J. Sola, T. Vidal-Calleja, J. Civera, and J. M. M. Montiel, “Impact of landmark parametrization on monocular EKF-SLAM with points and lines,” *International journal of computer vision*, vol. 97, no. 3, pp. 339–368, 2012.

[37] S. Hong, M. H. Lee, H.-H. Chun, S.-H. Kwon, and J. L. Speyer, “Observability of error states in GPS/INS integration,” *IEEE Transactions on Vehicular Technology*, vol. 54, no. 2, pp. 731–743, 2005.

[38] T. E. Humphreys, *Aerial Robotics*. https://gitlab.com/todd.humphreys/ar-book, 2021.

[39] M. D. Shuster, “Maximum likelihood estimation of spacecraft attitude,” *J. Astronaut. Sci.*, vol. 37, pp. 79–88, 1989.

[40] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in *European Conference on Computer Vision*. Springer, 2006, pp. 430–443.