arXiv:1906.02650v1 [cs.RO] 6 Jun 2019

VISUAL-INERTIAL NAVIGATION: A CONCISE REVIEW

Guoquan (Paul) Huang

Robot Perception and Navigation Group
University of Delaware, Newark, DE 19716
Email: ghuang@udel.edu

June 7, 2019

ABSTRACT

As inertial and visual sensors are becoming ubiquitous, visual-inertial navigation systems (VINS) have prevailed in a wide range of applications from mobile augmented reality to aerial navigation to autonomous driving, in part because of the complementary sensing capabilities and the decreasing costs and size of the sensors. In this paper, we survey thoroughly the research efforts taken in this field and strive to provide a concise but complete review of the related work – which is unfortunately missing in the literature while being greatly demanded by researchers and engineers – in the hope to accelerate the VINS research and beyond in our society as a whole.

1 Introduction

Over the years, inertial navigation systems (INS) [1, 2] have been widely used for estimating the 6DOF poses (positions and orientations) of sensing platforms (e.g., autonomous vehicles), in particular, in GPS-denied environments such as underwater, indoor, in the urban canyon, and on other planets. Most INS rely on a 6-axis inertial measurement unit (IMU) that measures the local linear acceleration and angular velocity of the platform to which it is rigidly connected. With the recent advancements of hardware design and manufacturing, low-cost light-weight micro-electro-mechanical (MEMS) IMUs have become ubiquitous [3, 4, 5], which enables high-accuracy localization for, among others, mobile devices [6] and micro aerial vehicles (MAVs) [7, 8, 9, 10, 11], holding huge implications in a wide range of emerging applications from mobile augmented reality (AR) [12, 13] and virtual reality (VR) [14] to autonomous driving [15, 16]. Unfortunately, simple integration of high-rate IMU measurements that are corrupted by noise and bias, often results in pose estimates unreliable for long-term navigation. Although a high-end tactical-grade IMU exists, it remains prohibitively expensive for widespread deployments. On the other hand, a camera that is small, light-weight, and energy-efficient, provides rich information about the environment and serves as an idea aiding source for INS, yielding visual-inertial navigation systems (VINS).

While this problem is challenging because of the lack of global information to reduce the motion drift accumulated over time (which is even exacerbated if low-cost, low-quality sensors are used), VINS have attracted significant attentions over the last decade. To date, many VINS algorithms are available for both visual-inertial SLAM [17] and visual-inertial odometry (VIO) [18, 19], such as the extended Kalman filter (EKF) [17, 18, 20, 21, 22], the unscented Kalman filter (UKF) [23, 24, 25], the batch or incremental smoother [26, 27], and (window) optimization-based approaches [28, 29, 30, 31]. Among these, the EKF-based methods remain popular because of its efficiency. For example, as a state-of-the-art solution of VINS on mobile devices, Project Tango [32] (or ARCore [12]) appears to use an EKF to fuse the visual and inertial measurements for motion tracking. Nevertheless, recent advances of preintegration have also allowed for efficient inclusion of high-rate IMU measurements in graph optimization-based formulations [29, 30, 33, 34, 35].

As evident, VINS technologies are emerging, largely due to the demanding mobile perception/navigation applications, which has given rise to a rich body of literature in this area. However, to the best of out knowledge, there is *no* contemporary literature review of VINS, although there are recent surveys broadly about SLAM [16, 36] while not specializing on VINS. This has made difficult for researchers and engineers in both academia and industry, to effectively find and understand the most important related work to their interests, which we have experienced over

the years when we are working on this problem. For this reason, we are striving to bridge this gap by: (i) offering a concise (due to space limitation) but complete review on VINS while focusing on the key aspects of state estimation, (ii) providing our understandings about the most important related work, and (iii) opening discussions about the challenges remaining to tackle. This is driven by the hope to (at least) help researchers/engineers track and understand the state-of-the-art VINS algorithms/systems, more efficiently and effectively, thus accelerating the VINS research and development in our society as a whole.

2 Visual-Inertial Navigation

In this section, we provide some basic background of canonical VINS, by describing the IMU propagation and camera measurement models within the EKF framework.

2.1 IMU Kinematic Model

The EKF uses the IMU (gyroscope and accelerometer) measurements for state propagation, and the state vector consists of the IMU states $\mathbf{x}_I$ and the feature position ${}^G\mathbf{p}_f$:

$$\begin{aligned}\mathbf{x} &= \begin{bmatrix}\mathbf{x}_I^T & {}^G\mathbf{p}_f^T\end{bmatrix}^T \\ &= \begin{bmatrix}{}^I_G\bar{\mathbf{q}}^T & \mathbf{b}_g^T & {}^G\mathbf{v}^T & \mathbf{b}_a^T & {}^G\mathbf{p}^T & {}^G\mathbf{p}_f^T\end{bmatrix}^T\end{aligned} \tag{1}$$

where ${}^I_G\bar{\mathbf{q}}$ is the unit quaternion that represents the rotation from the global frame of reference $\{G\}$ to the IMU frame $\{I\}$ (i.e., different parametrization of the rotation matrix $\mathbf{C}({}^I_G\bar{\mathbf{q}}) =: {}^I_G\mathbf{C}$); ${}^G\mathbf{p}$ and ${}^G\mathbf{v}$ are the IMU position and velocity in the global frame; and $\mathbf{b}_g$ and $\mathbf{b}_a$ denote the gyroscope and accelerometer biases, respectively.

By noting that the feature is static (with trivial dynamics), as well as using the IMU motion dynamics [37], the continuous-time dynamics of the state (1) is given by:

$$\begin{aligned}{}^I_G\dot{\bar{\mathbf{q}}}(t) &= \frac{1}{2}\mathbf{\Omega}\left({}^I\boldsymbol{\omega}(t)\right){}^I_G\bar{\mathbf{q}}(t),\ {}^G\dot{\mathbf{p}}(t) = {}^G\mathbf{v}(t),\ {}^G\dot{\mathbf{v}}(t) = {}^G\mathbf{a}(t) \\ \dot{\mathbf{b}}_g(t) &= \mathbf{n}_{wg}(t),\ \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t),\ {}^G\dot{\mathbf{p}}_f(t) = \mathbf{0}_{3\times 1}\end{aligned} \tag{2}$$

where ${}^I\boldsymbol{\omega} = \begin{bmatrix}\omega_1 & \omega_2 & \omega_3\end{bmatrix}^T$ is the rotational velocity of the IMU, expressed in $\{I\}$, ${}^G\mathbf{a}$ is the IMU acceleration in $\{G\}$, $\mathbf{n}_{wg}$ and $\mathbf{n}_{wa}$ are the white Gaussian noise processes that drive the IMU biases, and $\mathbf{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix}-\lfloor\boldsymbol{\omega}\times\rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0\end{bmatrix}$, where $\lfloor\boldsymbol{\omega}\times\rfloor$ is the skew-symmetric matrix.

A typical IMU provides gyroscope and accelerometer measurements, $\boldsymbol{\omega}_m$ and $\mathbf{a}_m$, both of which are expressed in the IMU local frame $\{I\}$ and given by:

$$\boldsymbol{\omega}_m(t) = {}^I\boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \tag{3}$$

$$\mathbf{a}_m(t) = \mathbf{C}({}^I_G\bar{\mathbf{q}}(t))\left({}^G\mathbf{a}(t) - {}^G\mathbf{g}\right) + \mathbf{b}_a(t) + \mathbf{n}_a(t) \tag{4}$$

where ${}^G\mathbf{g}$ is the gravitational acceleration expressed in $\{G\}$, and $\mathbf{n}_g$ and $\mathbf{n}_a$ are zero-mean, white Gaussian noise.

Linearization of (2) at the current state estimate yields the continuous-time state-estimate propagation model [21]:

$$\begin{aligned}{}^I_G\dot{\hat{\bar{\mathbf{q}}}}(t) &= \frac{1}{2}\mathbf{\Omega}\left({}^I\hat{\boldsymbol{\omega}}(t)\right){}^I_G\hat{\bar{\mathbf{q}}}(t),\ {}^G\dot{\hat{\mathbf{p}}}(t) = {}^G\hat{\mathbf{v}}(t),\ {}^G\dot{\hat{\mathbf{v}}}(t) = {}^G\hat{\mathbf{a}}(t) \\ \dot{\hat{\mathbf{b}}}_g(t) &= \mathbf{0}_{3\times 1},\ \dot{\hat{\mathbf{b}}}_a(t) = \mathbf{0}_{3\times 1},\ {}^G\dot{\hat{\mathbf{p}}}_f(t) = \mathbf{0}_{3\times 1}\end{aligned} \tag{5}$$

where $\hat{\mathbf{a}} = \mathbf{a}_m - \hat{\mathbf{b}}_a$ and $\hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_m - \hat{\mathbf{b}}_g$. The error state of dimension 18×1 is hence defined as follows [see (1)]:

$$\tilde{\mathbf{x}}(t) = \begin{bmatrix}{}^I\tilde{\boldsymbol{\theta}}^T(t) & \tilde{\mathbf{b}}_g^T(t) & {}^G\tilde{\mathbf{v}}^T(t) & \tilde{\mathbf{b}}_a^T(t) & {}^G\tilde{\mathbf{p}}^T(t) & {}^G\tilde{\mathbf{p}}_f^T(t)\end{bmatrix}^T \tag{6}$$

where we have employed the multiplicative error model for a quaternion [37]. That is, the error between the quaternion $\bar{\mathbf{q}}$ and its estimate $\hat{\bar{\mathbf{q}}}$ is the 3×1 angle-error vector, ${}^I\tilde{\boldsymbol{\theta}}$, implicitly defined by the error quaternion: $\delta\bar{\mathbf{q}} = \bar{\mathbf{q}}\otimes\hat{\bar{\mathbf{q}}} \simeq \begin{bmatrix}\frac{1}{2}{}^I\tilde{\boldsymbol{\theta}} \\ 1\end{bmatrix}$, where $\delta\bar{\mathbf{q}}$ describes the small rotation that causes the true and estimated attitude to coincide. The advantage of this parametrization permits a minimal representation, 3×3 covariance matrix $\mathbb{E}\left[{}^I\tilde{\boldsymbol{\theta}}\,{}^I\tilde{\boldsymbol{\theta}}^T\right]$, for the attitude uncertainty.

Now the continuous-time error-state propagation is:

$$\dot{\tilde{\mathbf{x}}}(t) = \mathbf{F}_c(t)\tilde{\mathbf{x}}(t) + \mathbf{G}_c(t)\mathbf{n}(t) \tag{7}$$

where $\mathbf{n} = \begin{bmatrix}\mathbf{n}_g^T & \mathbf{n}_{wg}^T & \mathbf{n}_a^T & \mathbf{n}_{wa}^T\end{bmatrix}^T$ is the system noise, $\mathbf{F}_c$ is the continuous-time error-state transition matrix, and $\mathbf{G}_c$ is the input noise matrix, which are given by (see [37]):

$$\mathbf{F}_c = \begin{bmatrix} -\lfloor\hat{\boldsymbol{\omega}}\times\rfloor & -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -\mathbf{C}^T({}^I_G\hat{\bar{\mathbf{q}}})\lfloor\hat{\mathbf{a}}\times\rfloor & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{C}^T({}^I_G\hat{\bar{\mathbf{q}}}) & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \tag{8}$$

$$\mathbf{G}_c = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -\mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -\mathbf{I}_3 & \mathbf{0}_3 & -\mathbf{C}^T({}^I_G\hat{\bar{\mathbf{q}}}) & \mathbf{0}_3 \\ -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \\ -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \tag{9}$$

The system noise is modelled as zero-mean white Gaussian process with autocorrelation $\mathbb{E}\left[\mathbf{n}(t)\mathbf{n}(\tau)^T\right] = \mathbf{Q}_c\delta(t-\tau)$, which depends on the IMU noise characteristics.

We have described the continuous-time propagation model using IMU measurements. However, in any practical EKF implementation, the discrete-time state-transition matrix, $\boldsymbol{\Phi}_k := \boldsymbol{\Phi}(t_{k+1}, t_k)$, is required in order to propagate the error covariance from time t_k to t_{k+1}. Typically it is found by solving the following matrix differential equation:

$$\dot{\boldsymbol{\Phi}}(t_{k+1}, t_k) = \mathbf{F}_c(t_{k+1})\boldsymbol{\Phi}(t_{k+1}, t_k) \tag{10}$$

with the initial condition $\boldsymbol{\Phi}(t_k, t_k) = \mathbf{I}_{18}$. This can be solved either numerically [21, 37] or analytically [19, 38, 39, 40]. Once it is computed, the EKF propagates the covariance as [41]:

$$\mathbf{P}_{k+1|k} = \boldsymbol{\Phi}_k\mathbf{P}_{k|k}\boldsymbol{\Phi}_k^T + \mathbf{Q}_{d,k} \tag{11}$$

where $\mathbf{Q}_{d,k}$ is the discrete-time system noise covariance matrix computed as follows:

$$\mathbf{Q}_{d,k} = \int_{t_k}^{t_{k+1}} \boldsymbol{\Phi}(t_{k+1}, \tau)\mathbf{G}_c(\tau)\mathbf{Q}_c\mathbf{G}_c^T(\tau)\boldsymbol{\Phi}^T(t_{k+1}, \tau)d\tau$$

2.2 Camera Measurement Model

The camera observes visual corner features, which are used to concurrently estimate the ego-motion of the sensing platform. Assuming a calibrated perspective camera, the measurement of the feature at time-step k is the perspective projection of the 3D point, ${}^{C_k}\mathbf{p}_f$, expressed in the current camera frame $\{C_k\}$, onto the image plane, i.e.,

$$\mathbf{z}_k = \frac{1}{z_k}\begin{bmatrix}x_k \\ y_k\end{bmatrix} + \mathbf{n}_{f_k} \tag{12}$$

$$\begin{bmatrix}x_k \\ y_k \\ z_k\end{bmatrix} = {}^{C_k}\mathbf{p}_f = \mathbf{C}({}^C_I\bar{\mathbf{q}})\mathbf{C}({}^I_G\bar{\mathbf{q}}_k)\left({}^G\mathbf{p}_f - {}^G\mathbf{p}_k\right) + {}^C\mathbf{p}_I \tag{13}$$

where $\mathbf{n}_{f_k}$ is the zero-mean, white Gaussian measurement noise with covariance $\mathbf{R}_k$. In (13), $\{{}^C_I\bar{\mathbf{q}}, {}^C\mathbf{p}_I\}$ is the rotation and translation between the camera and the IMU. This transformation can be obtained, for example, by performing camera-IMU extrinsic calibration *offline* [42]. However, in practice when the perfect calibration is unavailable, it is beneficial to VINS consistency to include these calibration parameters in the state vector and concurrently estimate them along with the IMU/camera poses [40].

For the use of EKF, linearization of (12) yields the following measurement residual [see (6)]:

$$\tilde{\mathbf{z}}_k = \mathbf{H}_k\tilde{\mathbf{x}}_{k|k-1} + \mathbf{n}_{f_k} = \mathbf{H}_{\mathbf{I}_k}\tilde{\mathbf{x}}_{I_{k|k-1}} + \mathbf{H}_{\mathbf{f}_k}{}^G\tilde{\mathbf{p}}_{f_{k|k-1}} + \mathbf{n}_{f_k} \tag{14}$$

where the measurement Jacobian $\mathbf{H}_k$ is computed as:

$$\begin{aligned}\mathbf{H}_k &= \begin{bmatrix}\mathbf{H}_{\mathbf{I}_k} & \mathbf{H}_{\mathbf{f}_k}\end{bmatrix} \\ &= \mathbf{H}_{\mathbf{proj}}\mathbf{C}({}^C_I\bar{\mathbf{q}})\begin{bmatrix}\mathbf{H}_{\boldsymbol{\theta}_k} & \mathbf{0}_{3\times 9} & \mathbf{H}_{\mathbf{p}_k} & \mathbf{C}({}^I_G\hat{\bar{\mathbf{q}}}_k)\end{bmatrix}\end{aligned} \tag{15}$$

$$\mathbf{H}_{\mathbf{proj}} = \frac{1}{\hat{z}_k^2}\begin{bmatrix}\hat{z}_k & 0 & -\hat{x}_k \\ 0 & \hat{z}_k & -\hat{y}_k\end{bmatrix} \tag{16}$$

$$\mathbf{H}_{\boldsymbol{\theta}_k} = \lfloor\mathbf{C}({}^I_G\hat{\bar{\mathbf{q}}}_k)\left({}^G\hat{\mathbf{p}}_f - {}^G\hat{\mathbf{p}}_k\right)\times\rfloor,\ \mathbf{H}_{\mathbf{p}_k} = -\mathbf{C}({}^I_G\hat{\bar{\mathbf{q}}}_k) \tag{17}$$

Once the measurement Jacobian and residual are computed, we can apply the standard EKF update equations to update the state estimates and error covariance [41].

3 State Estimation

It is clear from the preceding section that at the core of visual-inertial navigation systems (VINS) is a state estimation algorithm [see (2) and (12)], aiming to optimally fuse IMU measurements and camera images to provide motion tracking of the sensor platform. In this section, we review the VINS literature by focusing on the estimation engine.

3.1 Filtering-based vs. Optimization-based Estimation

Mourikis and Roumeliotis [18] developed one of the earliest successful VINS algorithms, known as the multi-state constraint Kalman filter (MSCKF), which later was applied to the application of spacecraft descent and landing [21] and fast UAV autonomous flight [43]. This approach uses the quaternion-based inertial dynamics [37] for state propagation tightly coupled with an efficient EKF update. Specifically, rather than adding features detected and tracked over the camera images to the state vector, their visual bearing measurements are projected onto the null space of the feature Jacobian matrix (i.e., linear marginalization [44]), thereby retaining motion constraints that only relate to the stochastically cloned camera poses in the state vector [45]. While reducing the computational cost by removing the need to co-estimate potentially hundreds and thousands of point features, this operation prevents the relinearization of the features' nonlinear measurements at later times, yielding approximations deteriorating its performance.

The standard MSCKF [18] recently has been extended and improved along different directions. In particular, by exploiting the observability-based methodology proposed in our prior work [46, 47, 48, 49], different observability-constrained (OC)-MSCKF algorithms have been developed to improve the filter consistency by enforcing the correct observability properties of the linearized VINS [19, 38, 39, 40, 50, 51, 52]. A square-root inverse version of the MSCKF, i.e., the square-root inverse sliding window filter (SR-ISWF) [6, 53] was introduced to improve the computational efficiency and numerical stability to enable VINS running on mobile devices with limited resources while not sacrificing estimation accuracy. We have introduced the optimal state constraint (OSC)-EKF [54, 55] that first optimally extracts all the information contained in the visual measurements about the relative camera poses in a sliding window and then uses these inferred relative-pose measurements in the EKF update. The (right) invariant Kalman filter [56] was recently employed to improve filter consistency [25, 57, 58, 59, 60], as well as the (iterated) EKF that was also used for VINS in robocentric formulations [22, 61, 62, 63]. On the other hand, in the EKF framework, different geometric features besides points have also been exploited to improve VINS performance, for example, line features used in [64, 65, 66, 67, 68] and plane features in [69, 70, 71, 72]. In addition, the MSCKF-based VINS was also extended to use rolling-shutter cameras with inaccurate time synchronization [64, 73], RGBD cameras [69, 74], multiple cameras [53, 75, 76] and multiple IMUs [77]. While the filtering-based VINS have shown to exhibit high-accuracy state estimation, they theoretically suffer from a limitation; that is, nonlinear measurements (12) must have a *one-time* linearization before processing, possibly introducing large linearization errors into the estimator and degrading performance.

Batch optimization methods, by contrast, solve a nonlinear least-squares (bundle adjustment or BA [78]) problem over a set of measurements, allowing for the reduction of error through relinearization [79, 80] but with high computational cost. Indelman et al. [27] employed the factor graph to represent the VINS problem and then solved it incrementally in analogy to iSAM [81, 82]. To achieve constant processing time when applied to VINS, typically a bounded-size sliding window of recent states are only considered as active optimization variables while marginalizing out past states and measurements [28, 31, 83, 84, 85]. In particular, Leutenegger et al. [28] introduced a keyframe-based optimization approach (i.e., OKVIS), whereby a set of non-sequential past camera poses and a series of recent inertial states, connected with inertial measurements, was used in nonlinear optimization for accurate trajectory estimation. Qin et al. [31] recently presented an optimization-based monocular VINS that can incorporate loop closures in a non-real time thread, while our recent VINS [86] is able to efficiently utilize loop closures in a single thread with linear complexity.

3.2 Tightly-coupled vs. Loosely-coupled Sensor Fusion

There are different schemes for VINS to fuse the visual and inertial measurements which can be broadly categorized into the loosely-coupled and the tightly-coupled. Specifically, the loosely-coupled fusion, in either filtering or optimization-based estimation, processes the visual and inertial measurements separately to infer their own motion constraints and then fuse these constraints (e.g., [27, 87, 88, 89, 90, 91]). Although this method is computationally efficient, the decoupling of visual and inertial constraints results in information loss. By contrast, the tightly-coupled

approaches directly fuse the visual and inertial measurements within a single process, thus achieving higher accuracy (e.g., [18, 28, 34, 40, 85, 92]).

3.3 VIO vs. SLAM

By jointly estimating the location of the sensor platform and the features in the surrounding environment, SLAM estimators are able to easily incorporate loop closure constraints, thus enabling bounded localization errors, which has attracted significant research efforts in the past three decades [16, 36, 93, 94]. VINS can be considered as an instance of SLAM (using particular visual and inertial sensors) and broadly include the visual-inertial (VI)-SLAM [28, 33, 85] and the visual-inertial odometry (VIO) [18, 22, 39, 40, 52, 95]. The former jointly estimates the feature positions and the camera/IMU pose that together form the state vector, whereas the latter does not include the features in the state but still utilizes the visual measurements to impose motion constraints between the camera/IMU poses. In general, by performing mapping (and thus loop closure), the VI-SLAM gains the better accuracy from the feature map and the possible loop closures while incurring higher computational complexity than the VIO, although different methods have been proposed to address this issue [18, 21, 28, 85, 96, 97]. However, VIO estimators are essentially odometry (dead reckoning) methods whose localization errors may grow unbounded unless some global information (e.g., GPS or *a priori* map) or constraints to previous locations (i.e., loop-closures) are used. Many approaches leverage feature observations from different keyframes to limit drift over the trajectory [28, 98]. Most have a two-thread system that optimizes a small window of "local" keyframes and features limiting drift in the short-term, while a background process optimizes a long-term sparse pose graph containing loop-closure constraints enforcing long-term consistency [31, 83, 99, 100]. For example, VINS-Mono [31, 100] uses loop-closure constraints in both the local sliding window and in the global batch optimization. Specifically, during the local optimization, feature observations from keyframes provide implicit loop-closure constraints, while the problem size remains small by assuming the keyframe poses are perfect (thus removing them from optimization).

In particular, whether or not performing loop closures in VINS either via mapping [83, 101, 102] and/or place recognition [103, 104, 105, 106, 107, 108] is one of the key differences between VIO and SLAM. While it is essential to utilize loop-closure information to enable bounded-error VINS performance, it is challenging due to the inability to remain computationally efficient without making inconsistent assumptions such as treating keyframe poses to be true, or reusing information. To this end, a hybrid estimator was proposed in [109] that used the MSCKF to perform real-time local estimation, and triggered global BA on loop-closure detection. This allows for the relinearization and inclusion of loop-closure constraints in a consistent manner, while requiring substantial additional overhead time where the filter waits for the BA to finish. Recently, Lynen et al. [110] developed a large-scale map-based VINS that uses a compressed prior map containing feature positions and their uncertainties and employs the matches to features in the prior map to constrain the estimates globally. DuToit et al. [102] exploited the idea of Schmidt KF [111] and developed a Cholesky-Schmidt EKF, which, however, uses *a prior* map with its full uncertainty and relaxes all the correlations between the mapped features and the current state variables; while our latest Schmidt-MSCKF [86] integrates loop closures in a single thread. Moreover, the recent point-line VIO [67] treats the 3D positions of marginalized keypoints as "true" for loop closure, which may lead to inconsistency.

3.4 Direct vs. Indirect Visual Processing

Visual processing pipeline is one of the key components to any VINS, responsible for transforming dense imagery data to motion constraints that can be incorporated into the estimation problem, whose algorithms can be categorized as either direct or indirect upon the visual residual models used. Seen as the classical technique, indirect methods [18, 28, 30, 40, 51] extract and track point features in the environment, while using geometric reprojection constraints during estimation. An example of a current state-of-the-art indirect visual SLAM is the ORB-SLAM2 [83, 112], which performs graph-based optimization of camera poses using information from 3D feature point correspondences.

In contrast, direct methods [96, 113, 114, 115] utilize raw pixel intensities in their formulation and allow for inclusion of a larger percentage of the available image information. LSD-SLAM [114] is an example of a state-of-the-art direct visual-SLAM which optimizes the transformation between pairs of camera keyframes based on minimizing their intensity error. Note that this approach also optimizes a separate graph containing keyframe constraints to allow for the incorporation of highly informative loop-closures to correct drift over long trajectories. This work was later extended from a monocular sensor to stereo and omnidirectional cameras for improved accuracy [116, 117]. Other popular direct methods include [118] and [119] which estimate keyframe depths along with the camera poses in a tightly-coupled manner, offering low-drift results. Application of direct methods to VINS has seen recent attention due to their ability to robustly track dynamic motion even in low-texture environments. For example, Bloesch et al. [61, 62] used a patch-based direct method to provide updates with an iterated EKF; Usenko et al. [96] introduced a

sliding-window VINS based on the discrete preintegration and direct image alignment; Ling et al. [9], Eckenhoff et al. [115] integrated direct image alignment with different IMU preintegration [34, 35, 85] for dynamic motion estimation.

While direct image alignments require a good initial guess and high frame rate due to the photometric consistency assumption, indirect visual tracking consumes extra computational resources on extracting and matching features. Nevertheless, indirect methods are more widely used in practical applications due to its maturity and robustness, but direct approaches have potentials in textureless scenarios.

3.5 Inertial Preintegration

Lupton and Sukkarieh [33] first developed the IMU preintegration, a computationally efficient alternative to the standard inertial measurement integration, which peforms the discrete integration of the inertial measurement dynamics in a *local* frame of reference, thus preventing the need to reintegrate the state dynamics at each optimization step. While this addresses the computational complexity issue, this method suffers from singularities due to the use of Euler angles in the orientation representation. To improve the stability of this preintegration, an on-manifold representation was introduced in [29, 34] which presents a singularity-free orientation representation on the $SO(3)$ manifold, incorporating the IMU preintegration into graph-based VINS.

While Shen et al. [85] introduced preintegration in the continuous form, they still discretely sampled the measurement dynamics without offering closed-form solutions, which left a significant gap in the theoretical completeness of preintegration theory from a continuous-time perspective. As compared to the discrete approximation of the preintegrated measurement and covariance calculations used in previous methods, in our prior work [35, 63], we have derived the closed-form solutions to both the measurement and covariance preintegration equations and showed that these solutions offer improved accuracy over the discrete methods, especially in the case of highly dynamic motion.

3.6 State Initialization

Robust, fast initialization to provide of accurate initial state estimates is crucial to bootstrap real-time VINS estimators, which is often solved in a linear closed form [7, 84, 120, 121, 122, 123, 124]. In particular, Martinelli [123] introduced a closed-form solution to the monocular visual-inertial initialization problem and later extended to the case where gyroscope bias calibration is also included [125] as well as to the cooperative scenario [126]. These approaches fail to model the uncertainty in inertial integration since they rely on the double integration of IMU measurements over an extended period of time. Faessler et al. [127] developed a re-initialization and failure recovery algorithm based on SVO [113] within a loosely-coupled estimation framework, while an additional downward-facing distance sensor is required to recover the metric scale. Mur-Artal and Tardós [83] introduced a high-latency (about 10 seconds) initializer built upon their ORB-SLAM [112], which computes initial scale, gravity direction, velocity and IMU biases with the visual-inertial full BA given a set of keyframes from ORB-SLAM. In [7, 84] a linear method was recently proposed for noise-free cases, by leveraging relative rotations obtained by short-term IMU (gyro) pre-integration but without modeling the gyroscope bias, which may be unreliable in real world in particular when distant visual features are observed.

4 Sensor Calibration

When fusing the measurements from different sensors, it is critical to determine in high precision both the *spatial* and *temporal* sensor calibration parameters. In particular, we should know accurately the *rigid-body transformation* between the camera and the IMU in order to correctly fuse motion information extracted from their measurements. In addition, due to improper hardware triggering, transmission delays, and clock synchronization errors, the timestamped sensing data of each sensor may disagree and thus, a timeline misalignment (time offset) between visual and inertial measurements might occur, which will eventually lead to unstable or inaccurate state estimates. It is therefore important that these *time offsets* should also be calibrated. The problem of sensor calibration of the spatial and/or temporal parameters has been the subject of many recent VINS research efforts [42, 128, 129, 130, 131, 132]. For example, Mirzaei and Roumeliotis [42] developed an EKF-based spatial calibration between the camera and IMU. Nonlinear observability analysis [133] for the calibration parameters was performed to show that these parameters are observable given random motion. Similarly, Jones and Soatto [128] examined the identifiability of the spatial calibration of the camera and IMU based on indistinguishable trajectory analysis and developed a filter based online calibration on an embedded platform. Kelly and Sukhatme [129] solved for the rigid-body transformation between the camera and IMU by aligning rotation curves of these two sensors via an ICP-like matching method.

Many of these research efforts have been focused on *offline* processes that often require additional calibration aids (fiducial tags) [42, 130, 134, 135, 136]. In particular, as one of the state-of-the-art approaches, the *Kalibr* calibration

toolbox [130, 134] uses a continuous-time basis function representation [137] of the sensor trajectory to calibrate both the extrinsics and intrinsics of a multi-sensor system in a batch fashion. As this B-spline representation allows for the direct computation of expected local angular velocity and local linear acceleration, the difference between the expected and measured inertial readings serve as errors in the batch optimization formulation. A downside of offline calibration is that it must be performed every time a sensor suite is reconfigured. For instance, if a sensor is removed for maintenance and returned, errors in the placement could cause poor performance, requiring a time-intensive recalibration.

Online calibration methods, by contrast, estimate the calibration parameters during every operation of the sensor suite, thereby making them more robust to and easier to use in such scenarios. Kim et al. [138] reformulated the IMU preintegration [33, 34, 35] by transforming the inertial readings from the IMU frame into a second frame. This allows for calibration between IMUs and other sensors (including other IMUs), but does not include temporal calibration and also relies on computing angular accelerations from gyroscope measurements. Li and Mourikis [131] performed navigation with simultaneous calibration of both the spatial and temporal extrinsics between a single IMU-camera pair in a filtering framework for use on mobile devices, which was later extended to include the intrinsics of both the camera and the IMU [139]. Qin and Shen [132] extended their prior work on batch-based monocular VINS [31] to include the time offset between the camera and IMU by interpolating the locations of features on the image plane. Schneider et al. [140] proposed the observability-aware online calibration utilizing the most informative motions. While we recently have also analyzed the degenerate motions of spatiotemporal calibration [141], it is not fully understood how to optimally model intrinsics and simultaneously calibrate them along with extrinsics [142, 143].

5 Observability Analysis

System observability plays an important role in the design of consistent state estimation [49], which examines whether the information provided by the available measurements is sufficient for estimating the state/parameters without ambiguity [133, 144, 145]. When a system is observable, the observability matrix is invertible, which is also closely related to the Fisher information (or covariance) matrix [146, 147]. Since this matrix describes the information available in the measurements, by studying its nullspace we can gain insights about the directions in the state space along which the estimator should acquire information. In our prior work [46, 47, 48, 52, 146, 148, 149, 150], we have been the first to design observability-constrained (OC) consistent estimators for robot localization problems. Since then, significant research efforts have been devoted to the observability analysis of VINS (e.g., [19, 38, 151, 152]).

In particular, VINS nonlinear observability analysis has been studied using different nonlinear system analysis techniques. For example, Jones and Soatto [128], Hernandez et al. [153] the system's indistinguishable trajectories [154] were examined from the observability perspective. By employing the concept of continuous symmetries as in [155], Martinelli [122] analytically derived the closed-form solution of VINS and identified that IMU biases, 3D velocity, global roll and pitch angles are observable. He has also examined the effects of degenerate motion [156], minimum available sensors [157], cooperative VIO [126] and unknown inputs [158, 159] on the system observability. Based on the Lie derivatives and observability matrix rank test [133], Hesch et al. [51] analytically showed that the monocular VINS has 4 unobservable directions, corresponding to the global yaw and the global position of the exteroceptive sensor. Guo and Roumeliotis [69] extended this method to the RGBD-camera aided INS that preserves the same unobservable directions if both point and plane measurements are available. With the similar idea, in [74, 129, 160], the observability of IMU-camera (monocular, RGBD) calibration was analytically studied, which shows that the extrinsic transformation between the IMU and camera is observable given generic motions. Additionally, in [161, 162], the system with a downward-looking camera measuring point features from horizontal planes was shown to have the observable global z position of the sensor.

As in practice VINS estimators are typically built upon the linearized system, what is practically more important is to perform observability analysis for the linearized VINS. In particular, the observability matrix [41, 163] for the linearized VINS system over the time interval $[k_o\ k]$ has the nullspace (i.e., unobservable subspace) that *ideally* spans *four* directions:

$$\mathbf{M} = \begin{bmatrix} \mathbf{H}_{k_o} \\ \mathbf{H}_{k_o+1}\mathbf{\Phi}_{k_o} \\ \vdots \\ \mathbf{H}_k\mathbf{\Phi}_{k-1}\cdots\mathbf{\Phi}_{k_o} \end{bmatrix} \overset{\mathbf{MN}=\mathbf{0}}{\Longrightarrow} \mathbf{N} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}({}^I_G\bar{\mathbf{q}}_k){}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & -\lfloor {}^G\mathbf{v}_k\times\rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{p}_k\times\rfloor {}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{p}_f\times\rfloor {}^G\mathbf{g} \end{bmatrix} \tag{18}$$

Note that the first block column of $\mathbf{N}$ in (18) corresponding to the global translation while the second block column corresponds to the and global rotation about the gravity vector [19, 38, 40, 51]. When designing a nonlinear estimator

for VINS, we would like the system model employed by the estimator to have an unobservable subspace spanned by these directions. However, this is not the case for the standard EKF as shown in [19, 38, 39, 40, 51]. In particular, the standard EKF linearized system, which linearizes system and measurement functions at the current state estimate, has an unobservable subspace of *three*, instead of four dimensions. This implies that the filter gains non-existent information from available measurements, leading to inconsistency. To address this issue, the first-estimates Jacobian (FEJ) idea [47] was adopted to improve MSCKF consistency [19, 40], and the OC methodology [48] was employed in developing the OC-VINS [38, 39, 50]. We recently have also developed the robocentric VIO (R-VIO) [22, 63] which preserves proper observability properties independent of linearization points.

6 Discussions and Conclusions

As inertial and visual sensors are becoming ubiquitous, visual-inertial navigation systems (VINS) have incurred significant research efforts and witnessed great progresses in the past decade, fostering an increasing number of innovative applications in practice. As a special instance of the well-known SLAM problem, VINS researchers have been quickly building up a rich body of literature on top of SLAM [36]. Given the growing number of papers published in this field, it has become harder (especially for practitioners) to keep up with the state of the art. Moreover, because of the particular sensor characteristics, it is not trivial to develop VINS algorithms from scratch without understanding the pros and cons of existing approaches in the literature (by noting that each method has its own particular focus and does not necessarily explain all the aspects of VINS estimation). All these have motivated us to provide this review on VINS, which, to the best of our knowledge, is unfortunately lacked in the literature and thus should be a useful reference for researchers/engineers who are working on this problem. Upon our significant prior work in this domain, we have strived to make this review concise but complete, by focusing on the key aspects about building a VINS algorithm including state estimation, sensor calibration and observability analysis.

While there are significant progresses on VINS made in the past decade, many challenges remain to cope with, and in the following we just list a few open to discuss:

- *Persistent localization:* While current VINS are able to provide accurate 3D motion tracking, but, in small-scale friendly environments, they are not robust enough for long-term, large-scale, safety-critical deployments, e.g., autonomous driving, in part due to resource constraints [95, 97, 164]. As such, it is demanding to enable persistent VINS even in challenging conditions (such as bad lighting and motions), e.g., by efficiently integrating loop closures or building and utilizing novel maps.
- *Semantic localization and mapping:* Although geometric features such as points, lines and planes [151, 165] are primarily used in current VINS for localization, these handcrafted features may not be work best for navigation, and it is of importance to be able to learn best features for VINS by leveraging recent advances of deep learning [166]. Moreover, a few recent research efforts have attempted to endow VINS with semantic understanding of environments [167, 168, 169, 170], which is only sparsely explored but holds great potentials.
- *High-dimensional object tracking:* When navigating in dynamic complex environments, besides high-precision localization, it is often necessary to detect, represent, and track moving objects that co-exist in the same space in real time, for example, 3D object tracking in autonomous navigation [92, 171, 172].
- *Distributed cooperative VINS:* Although cooperative VINS have been preliminarily studied in [126, 173], it is still challenging to develop real-time distributed VINS, e.g., for crowd sourcing operations. Recent work on cooperative mapping [174, 175] may shed some light on how to tackle this problem.
- *Extensions to different aiding sensors:* While optical cameras are seen an ideal aiding source for INS in many applications, other aiding sensors may more proper for some environments and motions, for example, acoustic sonars may be instead used in underwater [176]; low-cost light-weight LiDARs may work better in environments, e.g., with poor lighting conditions [71, 177]; and event cameras [178, 179] may better capture dynamic motions [180, 181]. Along this direction, we should investigate in-depth VINS extensions of using different aiding sources for applications at hand.

References

[1] A. B. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*. AIAA, 1997.

[2] D. Titterton and J. Weston, *Strapdown Inertial Navigation Technology*, 2nd ed. The Institution of Engineering and Technology, 2005.

[3] K. Maenaka, "MEMS inertial sensors and their applications," in *2008 5th International Conference on Networked Sensing Systems*, June 2008, pp. 71–73.

[4] N. Barbour, R. Hopkins, and A. Kourepenis, "Inertial MEMS system and applications," The Charles Stark Draper Laboratory, Tech. Rep., 2010.

[5] N. Ahmad, R. A. R. Ghazilla, N. M. Khairi, and V. Kasi, "Reviews on various inertial measurement unit (IMU) sensor applications," *International Journal of Signal Processing Systems*, vol. 1, no. 2, pp. 256–262, 2013.

[6] K. J. Wu, A. M. Ahmed, G. A. Georgiou, and S. I. Roumeliotis, "A square root inverse filter for efficient vision-aided inertial navigation on mobile devices," in *Robotics: Science and Systems Conference (RSS)*, 2015.

[7] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, "Initialization-free monocular visual-inertial estimation with application to autonomous MAVs," in *Proc. of the International Symposium on Experimental Robotics*, 2014.

[8] S. Shen, "Autonomous navigation in complex indoor and outdoor environments with micro aerial vehicles," Ph.D. dissertation, University of Pennsylvania, 2014.

[9] Y. Ling, T. Liu, and S. Shen, "Aggressive quadrotor flight using dense visual-inertial fusion," in *Proc. of the IEEE International Conference on Robotics and Automation*, 2016, pp. 1499–1506.

[10] T. Do, L. C. Carrillo-Arce, and S. I. Roumeliotis, "High-speed autonomous quadrotor navigation through visual and inertial paths," *The International Journal of Robotics Research*, 2018.

[11] J. Delmerico and D. Scaramuzza, "A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots," in *Proc. of the IEEE International Conference on Robotics and Automation*, Brisbane, Australia, May 21–25, 2018.

[12] Google, "Google ARCore," Available: https://developers.google.com/ar/.

[13] Apple, "Apple ARKit," Available: https://developer.apple.com/arkit/.

[14] Facebook, "Oculus VR," Available: https://www.oculus.com/.

[15] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, "VINS on wheels," in *Proc. of the IEEE International Conference on Robotics and Automation*, May 2017, pp. 5155–5162.

[16] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, Sept 2017.

[17] J. Kim and S. Sukkarieh, "Real-time implementation of airborne inertial-SLAM," *Robotics and Autonomous Systems*, vol. 55, no. 1, pp. 62–71, Jan. 2007.

[18] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy, Apr. 10–14, 2007, pp. 3565–3572.

[19] M. Li and A. I. Mourikis, "Improving the accuracy of EKF-based visual-inertial odometry," in *Proc. of the IEEE International Conference on Robotics and Automation*, St. Paul, MN, May 14–18, 2012, pp. 828–835.

[20] M. Bryson and S. Sukkarieh, "Observability analysis and active control for airborne SLAM," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 261–280, Jan. 2008.

[21] A. Mourikis, N. Trawny, S. Roumeliotis, A. Johnson, A. Ansar, and L. Matthies, "Vision-aided inertial navigation for spacecraft entry, descent, and landing," *IEEE Transactions on Robotics*, vol. 25, no. 2, pp. 264 –280, 2009.

[22] Z. Huai and G. Huang, "Robocentric visual-inertial odometry," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, Spain, Oct. 1-5, 2018.

[23] S. Ebcin and M. Veth, "Tightly-coupled image-aided inertial navigation using the unscented Kalman filter," Air Force Institute of Technology, Dayton, OH, Tech. Rep., 2007.

[24] G. Loianno, M. Watterson, and V. Kumar, "Visual inertial odometry for quadrotors on SE(3)," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 1544–1551.

[25] M. Brossard, S. Bonnabel, and A. Barrau, "Invariant Kalman filtering for visual inertial SLAM," in *21st International Conference on Information Fusion*, ser. 21st International Conference on Information Fusion. Cambrige, United Kingdom: University of Cambridge, Jul. 2018. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01588669

[26] D. Strelow, "Motion estimation from image and inertial measurements," Ph.D. dissertation, CMU, 2004.

[27] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing," *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 721–738, 2013.

[28] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[29] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, Feb. 2017.

[30] K. Eckenhoff, P. Geneva, and G. Huang, "Closed-form preintegration methods for graph-based visual-inertial navigation," *International Journal of Robotics Research*, vol. 38, no. 5, pp. 563–586, 2019.

[31] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[32] Google, "Google Project Tango," Available: https://www.google.com/atap/projecttango.

[33] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 61 –76, Feb. 2012.

[34] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation," in *Proc. of the Robotics: Science and Systems Conference*, Rome, Italy, Jul. 13–17, 2015.

[35] K. Eckenhoff, P. Geneva, and G. Huang, "High-accuracy preintegration for visual-inertial navigation," in *Proc. of the International Workshop on the Algorithmic Foundations of Robotics*, San Francisco, CA, Dec. 13–16, 2016.

[36] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. D. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[37] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep., Mar. 2005.

[38] J. Hesch, D. Kottas, S. Bowman, and S. Roumeliotis, "Towards consistent vision-aided inertial navigation," in *Algorithmic Foundations of Robotics X*, ser. Springer Tracts in Advanced Robotics, E. Frazzoli, T. Lozano-Perez, N. Roy, and D. Rus, Eds. Springer Berlin Heidelberg, 2013, vol. 86, pp. 559–574.

[39] ——, "Consistency analysis and improvement of vision-aided inertial navigation," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 158–176, 2013.

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

[41] P. S. Maybeck, *Stochastic Models, Estimation, and Control*. London: Academic Press, 1979, vol. 1.

[42] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1143–1156, Oct. 2008.

[43] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, April 2018.

[44] Y. Yang, J. Maley, and G. Huang, "Null-space-based marginalization: Analysis and algorithm," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vancouver, Canada, Sep. 24-28, 2017, pp. 6749–6755.

[45] S. I. Roumeliotis and J. W. Burdick, "Stochastic cloning: A generalized framework for processing relative state measurements," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington, DC, May 11-15, 2002, pp. 1788–1795.

[46] G. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Analysis and improvement of the consistency of extended Kalman filter-based SLAM," in *Proc. of the IEEE International Conference on Robotics and Automation*, Pasadena, CA, May 19-23 2008, pp. 473–479.

[47] ——, "A first-estimates Jacobian EKF for improving SLAM consistency," in *Proc. of the 11th International Symposium on Experimental Robotics*, Athens, Greece, Jul. 14–17, 2008.

[48] ——, "Observability-based rules for designing consistent EKF SLAM estimators," *International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, Apr. 2010.

[49] G. Huang, "Improving the consistency of nonlinear estimators: Analysis, algorithms, and applications," Ph.D. dissertation, Department of Computer Science and Engineering, University of Minnesota, 2012.

[50] D. G. Kottas, J. A. Hesch, S. L. Bowman, and S. I. Roumeliotis, "On the consistency of vision-aided inertial navigation," in *Proc. of the 13th International Symposium on Experimental Robotics*, Quebec City, Canada, Jun. 17–20, 2012.

[51] J. Hesch, D. Kottas, S. Bowman, and S. Roumeliotis, "Camera-IMU-based localization: Observability analysis and consistency improvement," *International Journal of Robotics Research*, vol. 33, pp. 182–201, 2014.

[52] G. Huang, M. Kaess, and J. Leonard, "Towards consistent visual-inertial navigation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Hong Kong, China, May 31-Jun. 7 2014, pp. 4926–4933.

[53] M. K. Paul, K. Wu, J. A. Hesch, E. D. Nerurkar, and S. I. Roumeliotis, "A comparative analysis of tightly-coupled monocular, binocular, and stereo VINS," in *Proc. of the IEEE International Conference on Robotics and Automation*, Singapore, Jul. 2017, pp. 165–172.

[54] G. Huang, K. Eckenhoff, and J. Leonard, "Optimal-state-constraint EKF for visual-inertial navigation," in *Proc. of the International Symposium on Robotics Research*, Sestri Levante, Italy, Sep. 12-15 2015.

[55] J. Maley, K. Eckenhoff, and G. Huang, "Generalized optimal-state-constraint extended kalman filter (OSC-EKF)," U.S. Army Research Laboratory, Tech. Rep. ARL-TR-7948, Feb. 2017.

[56] A. Barrau and S. Bonnabel, "Invariant kalman filtering," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 237–257, 2018.

[57] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, "An invariant-ekf vins algorithm for improving consistency," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2017, pp. 1578–1585.

[58] T. Zhang, K. Wu, J. Song, S. Huang, and G. Dissanayake, "Convergence and consistency analysis for a 3D invariant-ekf slam," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 733–740, April 2017.

[59] S. Heo and C. G. Park, "Consistent EKF-based visual-inertial odometry on matrix lie group," *IEEE Sensors Journal*, vol. 18, no. 9, pp. 3780–3788, May 2018.

[60] M. Brossard, S. Bonnabel, and J. Condomines, "Unscented kalman filtering on lie groups," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2017, pp. 2485–2491.

[61] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Hamburg, Germany: IEEE, September 2015, pp. 298–304.

[62] 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.

[63] Z. Huai and G. Huang, "Robocentric visual-inertial odometry," *International Journal of Robotics Research*, Apr. 2019, (to appear).

[64] H. Yu and A. I. Mourikis, "Vision-aided inertial navigation with line features and a rolling-shutter camera," in *International Conference on Robotics and Intelligent Systems*, Hamburg, Germany, October 2015, pp. 892–899.

[65] D. G. Kottas and S. I. Roumeliotis, "Efficient and consistent vision-aided inertial navigation using line observations," in *International Conference on Robotics and Automation*, Karlsruhe, Germany, May 6-10 2013, pp. 1540–1547.

[66] S. Heo, J. H. Jung, and C. G. Park, "Consistent EKF-based visual-inertial navigation using points and lines," *IEEE Sensors Journal*, 2018.

[67] F. Zheng, G. Tsai, Z. Zhang, S. Liu, C.-C. Chu, and H. Hu, "PI-VIO: Robust and Efficient Stereo Visual Inertial Odometry using Points and Lines," *ArXiv e-prints*, Mar. 2018.

[68] Y. He, J. Zhao, Y. Guo, W. He, and K. Yuan, "PL-VIO: Tightly coupled monocular visual inertial odometry using point and line features," *Sensors*, vol. 18, no. 4, 2018. [Online]. Available: http://www.mdpi.com/1424-8220/18/4/1159

[69] C. X. Guo and S. I. Roumeliotis, "IMU-RGBD camera navigation using point and plane features," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 3164–3171.

[70] M. Hsiao, E. Westman, and M. Kaess, "Dense planar-inertial slam with structural constraints," in *Proc. of the IEEE International Conference on Robotics and Automation*, Brisbane, Australia, 2018.

[71] P. Geneva, K. Eckenhoff, Y. Yang, and G. Huang, "LIPS: Lidar-inertial 3d plane slam," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, Spain, Oct. 1-5, 2018.

[72] Y. Yang, P. Geneva, X. Zuo, K. Eckenhoff, Y. Liu, and G. Huang, "Tightly-coupled aided inertial navigation with point and plane features," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[73] C. Guo, D. Kottas, R. DuToit, A. Ahmed, R. Li, and S. Roumeliotis, "Efficient visual-inertial navigation using a rolling-shutter camera with inaccurate timestamps," in *Proc. of the Robotics: Science and Systems Conference*, Berkeley, CA, Jul. 13–17, 2014.

[74] C. Guo and S. Roumeliotis, "IMU-RGBD camera 3D pose estimation and extrinsic calibration: Observability analysis and consistency improvement," in *Proc. of the IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, May 6–10, 2013.

[75] M. K. Paul and S. I. Roumeliotis, "Alternating-stereo vins: Observability analysis and performance evaluation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[76] K. Eckenhoff, P. Geneva, J. Bloecker, and G. Huang, "Multi-camera visual-inertial navigation with online intrinsic and extrinsic calibration," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[77] K. Eckenhoff, P. Geneva, and G. Huang, "Sensor-failure-resilient multi-imu visual-inertial navigation," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[78] B. Triggs, P. McLauchlan, R. Hartley, and andrew Fitzgibbon, "Bundle adjustment – A modern synthesis," in *Vision Algorithms: Theory and Practice*, ser. LNCS, W. Triggs, A. Zisserman, and R. Szeliski, Eds. Springer Verlag, 2000, pp. 298–375.

[79] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous localization and mapping via square root information smoothing," *International Journal of Robotics Research*, vol. 25, no. 12, pp. 1181–1203, Dec. 2006.

[80] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. of the IEEE International Conference on Robotics and Automation*, Shanghai, China, May 9–13, 2011, pp. 3607–3613.

[81] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[82] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *International Journal of Robotics Research*, vol. 31, pp. 217–236, Feb. 2012.

[83] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular slam with map reuse," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, April 2017.

[84] Z. Yang and S. Shen, "Monocular visual-inertial state estimation with online initialization and camera-imu extrinsic calibration," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 1, pp. 39–51, Jan 2017.

[85] S. Shen, N. Michael, and V. Kumar, "Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft mavs," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 5303–5310.

[86] P. Geneva, K. Eckenhoff, and G. Huang, "A linear-complexity EKF for visual-inertial navigation with loop closures," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[87] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. of the IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nara, Japan, Nov. 13–16, 2007, pp. 225–234.

[88] S. Weiss and R. Siegwart, "Real-time metric state estimation for modular vision-inertial systems," in *Proc. of the IEEE International Conference on Robotics and Automation*, Shanghai, China, May 9–13, 2011, pp. 4531–4537.

[89] S. Weiss, M. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments," in *Proc. of the IEEE International Conference on Robotics and Automation*, St. Paul, MN, May 14–18, 2012, pp. 957–964.

[90] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to mav navigation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 3923–3929.

[91] L. Kneip, M. Chli, and R. Y. Siegwart, "Robust real-time visual odometry with a single camera and an imu," in *British Machine Vision Conference*, September 2011.

[92] K. Eckenhoff, Y. Yang, P. Geneva, and G. Huang, "Tightly-coupled visual-inertial localization and 3D rigid-body target tracking," *IEEE Robotics and Automation Letters (RA-L)*, vol. 4, no. 2, pp. 1541–1548, 2019.

[93] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robotics Automation Magazine*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[94] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robotics Automation Magazine*, vol. 13, no. 3, pp. 108–117, 2006.

[95] M. Li, "Visual-inertial odometry on resource-constrained systems," Ph.D. dissertation, UC Riverside, 2014.

[96] V. Usenko, J. Engel, J. Stückler, and D. Cremers, "Direct visual-inertial odometry with stereo cameras," in *IEEE International Conference on Robotics and Automation*, Stockholm, Sweden, May 2016, pp. 1885–1892.

[97] M. Li and A. I. Mourikis, "Optimization-based estimator design for vision-aided inertial navigation," in *Robotics: Science and Systems*, Berlin, Germany, June 2013, pp. 241–248.

[98] E. Nerurkar, K. Wu, and S. Roumeliotis, "C-klam: Constrained keyframe-based localization and mapping," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 3638–3643.

[99] H. Liu, M. Chen, G. Zhang, H. Bao, and Y. Bao, "Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1974–1982.

[100] T. Qin, P. Li, and S. Shen, "Relocalization, global optimization and map merging for monocular visual-inertial slam," *arXiv preprint arXiv:1803.01549*, 2018.

[101] T. Schneider, M. Dymczyk, M. Fehr, K. Egger, S. Lynen, I. Gilitschenski, and R. Siegwart, "Maplab: An open framework for research in visual-inertial mapping and localization," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1418–1425, July 2018.

[102] R. C. DuToit, J. A. Hesch, E. D. Nerurkar, and S. I. Roumeliotis, "Consistent map-based 3d localization on mobile devices," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 6253–6260.

[103] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, October 2012.

[104] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual Place Recognition: A Survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1–19, Feb 2016.

[105] Y. Latif, G. Huang, J. Leonard, and J. Neira, "An online sparsity-cognizant loop-closure algorithm for visual navigation," in *Proc. of the Robotics: Science and Systems Conference*, Berkeley, CA, Jul. 12-16 2014.

[106] ——, "Sparse optimization for robust and efficient loop closing," *Robotics and Autonomous Systems*, vol. 93, pp. 13–26, Jul. 2017.

[107] F. Han, H. Wang, G. Huang, and H. Zhang, "Sequence-based sparse optimization methods for long-term loop closure detection in visual SLAM," *Autonomous Robots*, vol. 42, no. 7, pp. 1323–1335, 2018.

[108] N. Merrill and G. Huang, "Lightweight unsupervised deep loop closure," in *Proc. of Robotics: Science and Systems (RSS)*, Pittsburgh, PA, Jun. 26-30, 2018.

[109] A. I. Mourikis and S. I. Roumeliotis, "A dual-layer estimator architecture for long-term localization," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008.* IEEE, 2008, pp. 1–8.

[110] S. Lynen, T. Sattler, M. Bosse, J. A. Hesch, M. Pollefeys, and R. Siegwart, "Get out of my lab: Large-scale, real-time visual-inertial localization." in *Robotics: Science and Systems*, 2015.

[111] S. F. Schmidt, "Application of state-space methods to navigation problems," ser. Advances in Control Systems, C. LEONDES, Ed. Elsevier, 1966, vol. 3, pp. 293 – 340.

[112] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 15, no. 2, pp. 1147–1163, 2015.

[113] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *Proc. of the IEEE International Conference on Robotics and Automation*, Hong Kong, China, May 2014.

[114] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. European Conference on Computer Vision*, Zurich, Switzerland, Sep. 6–12, 2014.

[115] K. Eckenhoff, P. Geneva, and G. Huang, "Direct visual-inertial navigation with analytical preintegration," in *Proc. of the IEEE International Conference on Robotics and Automation*, Singapore, May 29–Jun.3, 2017, pp. 1429–1435.

[116] J. Engel, J. Stückler, and D. Cremers, "Large-scale direct SLAM with stereo cameras," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 1935–1942.

[117] D. Caruso, J. Engel, and D. Cremers, "Large-scale direct SLAM for omnidirectional cameras," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 141–148.

[118] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, 2018.

[119] R. Wang, M. Schwörer, and D. Cremers, "Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras," in *International Conference on Computer Vision (ICCV), Venice, Italy*, 2017.

[120] L. Kneip, S. Weiss, and R. Siegwart, "Deterministic initialization of metric state estimation filters for loosely-coupled monocular vision-inertial systems," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 2235–2241.

[121] T. Dong-Si and A. I. Mourikis, "Initialization in vision-aided inertial navigation with unknown camera-imu calibration," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura, Portugal, Oct. 2012, pp. 1064–1071.

[122] A. Martinelli, "Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 44–60, 2012.

[123] ——, "Closed-form solution of visual-inertial structure from motion," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 138–152, Jan 2014.

[124] V. Lippiello and R. Mebarki, "Closed-form solution for absolute scale velocity estimation using visual and inertial data with a sliding least-squares estimation," in *Mediterranean Conference on Control and Automation*, June 2013, pp. 1261–1266.

[125] J. Kaiser, A. Martinelli, F. Fontana, and D. Scaramuzza, "Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 18–25, Jan 2017.

[126] A. Martinelli, "Closed-form solution to cooperative visual-inertial structure from motion," *arXiv*, vol. abs/1802.08515, 2018.

[127] M. Faessler, F. Fontana, C. Forster, and D. Scaramuzza, "Automatic re-initialization and failure recovery for aggressive flight with a monocular vision-based quadrotor," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1722–1729.

[128] E. S. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach," *International Journal of Robotics Research*, vol. 30, no. 4, pp. 407–430, Apr. 2011.

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

[130] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 1280–1286.

[131] M. Li and A. I. Mourikis, "Online temporal calibration for Camera-IMU systems: Theory and algorithms," *International Journal of Robotics Research*, vol. 33, no. 7, pp. 947–964, Jun. 2014.

[132] T. Qin and S. Shen, "Online temporal calibration for monocular visual-inertial systems," *arXiv preprint arXiv:1808.00692*, 2018.

[133] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control*, vol. 22, no. 5, pp. 728–740, Oct. 1977.

[134] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 4304–4311.

[135] M. Fleps, E. Mair, O. Ruepp, M. Suppa, and D. Burschka, "Optimization based IMU camera calibration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 3297–3304.

[136] J. Nikolic, M. Burri, I. Gilitschenski, J. Nieto, and R. Siegwart, "Non-parametric extrinsic and intrinsic calibration of visual-inertial sensor systems," vol. 16, 07 2016.

[137] P. Furgale, C. H. Tong, T. D. Barfoot, and G. Sibley, "Continuous-time batch trajectory estimation using temporal basis functions," *The International Journal of Robotics Research*, vol. 34, no. 14, pp. 1688–1710, 2015.

[138] D. Kim, S. Shin, and I. S. Kweon, "On-line initialization and extrinsic calibration of an inertial navigation system with a relative preintegration method on manifold," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 3, pp. 1272–1285, July 2018.

[139] M. Li, H. Yu, X. Zheng, and A. I. Mourikis, "High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 409–416.

[140] T. Schneider, M. Li, C. Cadena, J. Nieto, and R. Siegwart, "Observability-aware self-calibration of visual and inertial sensors for ego-motion estimation," *IEEE Sensors Journal*, pp. 1–1, 2019.

[141] Y. Yang, P. Geneva, K. Eckenhoff, and G. Huang, "Degenerate motion analysis for aided INS with online spatial and temporal calibration," *IEEE Robotics and Automation Letters (RA-L)*, vol. 4, no. 2, pp. 2070–2077, 2019.

[142] J. Rehder and R. Siegwart, "Camera/imu calibration revisited," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3257–3268, June 2017.

[143] J. Nikolic, "Characterisation, calibration, and design of visual-inertial sensor systems for robot navigation," Ph.D. dissertation, ETH Zurich, 2016.

[144] W. L. Brogan, *Modern Control Theory*. Upper Saddle River, NJ: Prentice Hall, 1991.

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

[146] G. Huang, A. I. Mourikis, and S. I. Roumeliotis, "An observability constrained sliding window filter for SLAM," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, CA, Sep. 25-30 2011, pp. 65–72.

[147] G. Huang, "Towards consistent filtering for discrete-time partially-observable nonlinear systems," *Systems and Control Letters*, vol. 106, pp. 87–95, Aug. 2017.

[148] G. Huang, A. I. Mourikis, and S. I. Roumeliotis, "On the complexity and consistency of UKF-based SLAM," in *Proc. of the IEEE International Conference on Robotics and Automation*, Kobe, Japan, May 12-17 2009, pp. 4401–4408.

[149] ——, "A quadratic-complexity observability-constrained unscented Kalman filter for SLAM," *IEEE Transactions on Robotics*, vol. 29, no. 5, pp. 1226–1243, Oct. 2013.

[150] G. Huang, N. Trawny, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based consistent EKF estimators for multi-robot cooperative localization," *Autonomous Robots*, vol. 30, no. 1, pp. 99–122, Jan. 2011.

[151] Y. Yang and G. Huang, "Aided inertial navigation with geometric features: Observability analysis," in *Proc. of the IEEE International Conference on Robotics and Automation*, Brisbane, Australia, May 21–25, 2018.

[152] ——, "Observability analysis of aided ins with heterogeneous features of points, lines and planes," *IEEE Transactions on Robotics*, Jun. 2018, (accepted).

[153] J. Hernandez, K. Tsotsos, and S. Soatto, "Observability, identifiability and sensitivity of vision-aided inertial navigation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Seattle, WA, May 26–30, 2015, pp. 2319–2325.

[154] A. Isidori, *Nonlinear Control Systems*. Springer, 1995.

[155] A. Martinelli, "State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration," *IEEE Transactions on Robotics*, vol. 27, no. 2, pp. 239–255, 2011.

[156] ——, "Visual-inertial structure from motion: Observability and resolvability," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov. 2013, pp. 4235–4242.

[157] ——, "Visual-inertial structure from motion: Observability vs minimum number of sensors," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, May 2014, pp. 1020–1027.

[158] ——, *Nonlinear Unknown Input Observability: The General Analytic Solution*. arXiv:1704.03252, 2017.

[159] ——, "Nonlinear unknown input observability: Extension of the observability rank condition," *IEEE Transactions on Automatic Control*, pp. 1–1, 2018.

[160] F. M. Mirzaei and S. I. Roumeliotis, "A Kalman filter-based algorithm for IMU-camera calibration," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, Oct. 29 - Nov. 2 2007, pp. 2427–2434.

[161] G. Panahandeh, C. X. Guo, M. Jansson, and S. I. Roumeliotis, "Observability analysis of a vision-aided inertial navigation system using planar features on the ground," in *International Conference on Intelligent Robots and Systems*, Tokyo, Japan, 2013, pp. 4187–4194.

[162] G. Panahandeh, S. Hutchinson, P. Händel, and M. Jansson, "Planar-based visual inertial navigation: Observability analysis and motion estimation," *Journal of Intelligent & Robotic Systems*, vol. 82, no. 2, pp. 277–299, May 2016.

[163] Z. Chen, K. Jiang, and J. Hung, "Local observability matrix and its application to observability analyses," in *Proc. of the 16th Annual Conference of IEEE*, Pacific Grove, CA, Nov. 27–30, 1990, pp. 100–103.

[164] Z. Zhang, A. Suleiman, L. Carlone, V. Sze, and S. Karaman, "Visual-inertial odometry on chip: An algorithm-and-hardware co-design approach," in *Robotics: Science and Systems*, 2017.

[165] Y. Yang and G. Huang, "Aided inertial navigation: Unified feature representations and observability analysis," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[166] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2016.

[167] J. Dong, X. Fei, and S. Soatto, "Visual-inertial semantic scene representation for3d object detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, Jul. 2017, pp. 3567–3577.

[168] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, "Probabilistic data association for semantic SLAM," in *Proc. of the IEEE International Conference on Robotics and Automation*, May 2017, pp. 1722–1729.

[169] X. Fei and S. Soatto, "Visual-inertial object detection and mapping," *arXiv preprint arXiv:1806.08498*, 2018.

[170] K.-N. Lianos, J. L. Schönberger, M. Pollefeys, and T. Sattler, "VSO: Visual semantic odometry," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 234–250.

[171] P. Li, T. Qin, and S. Shen, "Stereo vision-based semantic 3d object and ego-motion tracking for autonomous driving," *arXiv preprint arXiv:1807.02062*, 2018.

[172] K. Eckenhoff, I. Yadav, G. Huang, and H. Tanner, "Dynamic target interception in cluttered environments," in *RT-DUNE ICRA 2018 Workshop*, Brisbane, Australia, May 21, 2018.

[173] I. Melnyk, J. Hesch, and S. Roumeliotis, "Cooperative vision-aided inertial navigation using overlapping views," in *Proc. of the IEEE International Conference on Robotics and Automation*, Saint Paul, MN, May 14–18, 2012, pp. 936–943.

[174] C. X. Guo, K. Sartipi, R. C. DuToit, G. A. Georgiou, R. Li, J. O'Leary, E. D. Nerurkar, J. A. Hesch, and S. I. Roumeliotis, "Large-scale cooperative 3d visual-inertial mapping in a manhattan world," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 1071–1078.

[175] ——, "Resource-aware large-scale cooperative three-dimensional mapping using multiple mobile devices," *IEEE Transactions on Robotics*, pp. 1–21, 2018.

[176] Y. Yang and G. Huang, "Acoustic-inertial underwater navigation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Singapore, May 29–Jun.3, 2017, pp. 4927–4933.

[177] J. A. Hesch, F. M. Mirzaei, G. L. Mariottini, and S. I. Roumeliotis, "A laser-aided inertial navigation system (l-ins) for human localization in unknown indoor environments," in *International Conference on Robotics and Automation*, Anchorage, Alaska, May. 3 - 8 2010, pp. 5376–5382.

[178] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128$\times$128 120 db 15 μs latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008.

[179] S.-C. Liu and T. Delbruck, "Neuromorphic sensory systems," *Current Opinion in Neurobiology*, vol. 20, no. 3, pp. 288–295, 2010.

[180] A. Z. Zhu, N. Atanasov, and K. Daniilidis, "Event-based visual inertial odometry," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 5816–5824.

[181] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," *IEEE Transactions on Robotics*, pp. 1–16, 2018.