2017 IEEE International Conference on Robotics and Automation (ICRA)
Singapore, May 29 - June 3, 2017

VINS on Wheels

Kejian J. Wu†, Chao X. Guo‡, Georgios Georgiou‡, and Stergios I. Roumeliotis‡

Abstract— **In this paper, we present a vision-aided inertial navigation system (VINS) for localizing wheeled robots. In particular, we prove that VINS has additional unobservable directions, such as the scale, when deployed on a ground vehicle that is constrained to move along straight lines or circular arcs. To address this limitation, we extend VINS to incorporate low-frequency wheel-encoder data, and show that the scale becomes observable. Furthermore, and in order to improve the localization accuracy, we introduce the manifold-(m)VINS that exploits the fact that the vehicle moves on an approximately planar surface. In our experiments, we first show the performance degradation of VINS due to special motions, and then demonstrate that by utilizing the additional sources of information, our system achieves significantly higher positioning accuracy, while operating in real-time on a commercial-grade mobile device.**

I. Introduction

Over the past 20 years, extensive research has focused on simultaneous localization and mapping (SLAM) with mobile robots navigating over flat terrain [1], [2]. In the absence of GPS, various exteroceptive sensors (e.g., ultrasonic, laser scanners, cameras, and, more recently, RGB-D) have been used in conjunction with 2D wheel odometry to determine, typically, the 3-degree-of-freedom (dof) position and orientation (pose) of the robot. In most cases, however, the underlying planar-motion assumption is only approximately satisfied (e.g., due to the unevenness, or roughness, of the surface, the presence of ramps, bumps, and low-height obstacles on the floor), thus significantly increasing the unmodeled part of the robot's odometry error and leading to low-accuracy estimates or, in the absence of external corrections, even divergence.

On the other hand, vision-aided inertial navigation systems (VINS), where visual observations from a camera are combined with data from an inertial measurement unit (IMU) to estimate the 6-dof pose of a platform navigating in 3D, have been shown to achieve high-accuracy localization results (e.g., [3], [4]), even on low-cost mobile devices (e.g., [5], [6]). Therefore, one would expect that it would be straightforward to deploy a VINS for localizing robots moving in 2D. Surprisingly, however, this is not the case. And one of the main reasons is that the restricted motion (approximately planar and, for the most part, along arcs or straight lines at constant speed or acceleration) that ground robots often undergo when navigating, e.g., indoors, alters the observability properties of VINS and renders certain, additional, dof unobservable.

Specifically, as proven in [7], [8], a VINS has 4 unobservable dof corresponding to 3 dof of global translation and 1 dof of rotation around the gravity vector (yaw). This result, however, holds only when the IMU-camera pair undergoes generic 3D motion. In contrast, and as shown in this paper, additional dof, such as the scale, become unobservable when the robot is restricted to move with constant acceleration.[1] In particular, under the simplifying assumption of perfectly-known gyroscope biases, [9] showed that the VINS's initial state cannot be uniquely determined for certain motions, but without specifying which are the additional unobservable directions. In this work, we consider the most general case of unknown gyroscope biases and determine these unobservable directions *analytically*.[2]

Furthermore, motivated by the key findings of our observability analysis, in this paper we focus on improving the localization accuracy of VINS when deployed on wheeled robots. Firstly, in order to ensure that information about the scale is always available (e.g., even for the periods of time when the robot moves with almost constant acceleration), we extend the VINS algorithm to incorporate wheel-odometry measurements. Since these are often noisy and of frequency significantly lower than that of the IMU, we process them in a robust manner, by first integrating the raw encoder data and then treating them as inferred displacement measurements between consecutive poses. Additionally, we take advantage of the fact that the robot moves on an *approximately* flat surface and introduce the manifold-(m)VINS, which explicitly considers the planar-motion constraint in the estimation algorithm to reduce the localization error. This is achieved by analyzing the motion profile of the robot, and its deviation from planar motion (e.g., due to terrain unevenness or vibration of the IMU-camera's mounting platform) and formulating stochastic (i.e., "soft"), instead of deterministic (i.e., "hard") constraints, that allow to properly model the vehicle's almost-planar motion. In summary, the main novel contributions of this work are:

- We analytically determine the unobservable dof of VINS under special, restrictive motions.

†K. J. Wu is with the Department of Electrical and Computer Engineering, Univ. of Minnesota, Minneapolis, USA. kejian@cs.umn.edu

‡C. X. Guo, G. Georgiou, and S. I. Roumeliotis are with the Department of Computer Science and Engineering, Univ. of Minnesota, Minneapolis, USA. {chaguo,georgiou,stergios}@cs.umn.edu

This work was supported by the University of Minnesota and the National Science Foundation (IIS-1111638, IIS-1328722).

[1]Note that although the motion constraints considered are never *exactly* satisfied, as explained later on and shown experimentally, motion profiles close to these significantly reduce the information along the unobservable directions, and hence degrade the localization accuracy.

[2]Note that observability is a fundamental property of the VINS model itself, and does not depend on the specific estimator employed for SLAM. Thus, the additional unobservable directions of monocular-VINS will negatively impact the accuracy of both batch-least-squares (e.g., [10], [11], [12]) and sliding-window filters/smoothers (e.g., [3], [4], [5], [6], [13]).

- We extend VINS to process low-frequency odometric measurements, thus rendering scale always observable.
- We introduce the mVINS which incorporates constraints about the motion of the vehicle (in this case, approximately planar) to improve the localization accuracy.
- Through experiments, we validate the key findings of our theoretical analysis, and demonstrate the increased accuracy of the proposed VINS-algorithm extensions when deployed on a tablet onboard a wheeled robot that navigates within a large-scale building.

II. Preliminaries on Vision-aided Inertial Navigation System (VINS)

In this section, we provide a brief review of the monocular-VINS which serves as the key component of our system. The VINS estimates the following state vector:

$$\mathbf{x} = \begin{bmatrix} {}^{I}\mathbf{q}_G^T & \mathbf{b}_g^T & {}^{G}\mathbf{v}_I^T & \mathbf{b}_a^T & {}^{G}\mathbf{p}_I^T \mid {}^{G}\mathbf{f}_1^T & \cdots & {}^{G}\mathbf{f}_N^T \end{bmatrix}^T \quad (1)$$

where ${}^{I}\mathbf{q}_G$ is the unit quaternion that represents the orientation of the global frame $\{G\}$ in the IMU frame $\{I\}$ at time t. ${}^{G}\mathbf{v}_I$ and ${}^{G}\mathbf{p}_I$ are the the velocity and position of $\{I\}$ in $\{G\}$, respectively, and the gyroscope and accelerometer biases are denoted by $\mathbf{b}_g$ and $\mathbf{b}_a$, respectively. Finally, the positions of point features in $\{G\}$ are denoted by ${}^{G}\mathbf{f}_j,\ j = 1, ..., N$.

The IMU provides measurements of the rotational velocity, $\boldsymbol{\omega}_m$, and the linear acceleration, $\mathbf{a}_m$, as:

$$\begin{aligned} \boldsymbol{\omega}_m(t) &= {}^{I}\boldsymbol{\omega}(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \\ \mathbf{a}_m(t) &= \mathbf{C}({}^{I}\mathbf{q}_G(t))({}^{G}\mathbf{a}(t) - {}^{G}\mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t) \end{aligned} \quad (2)$$

where the noise terms, $\mathbf{n}_g(t)$ and $\mathbf{n}_a(t)$ are modelled as zero-mean, white Gaussian noise processes, while the gravitational acceleration, ${}^{G}\mathbf{g}$, is considered a known constant. The IMU's rotational velocity ${}^{I}\boldsymbol{\omega}(t)$ and linear acceleration ${}^{G}\mathbf{a}(t)$, in (2), can be used to derive the continuous-time system equations:

$$\begin{aligned} {}^{I}\dot{\mathbf{q}}_G(t) &= \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega}_m(t) - \mathbf{b}_g(t) - \mathbf{n}_g(t)){}^{I}\mathbf{q}_G(t) \\ \dot{\mathbf{b}}_g(t) &= \mathbf{n}_{wg}(t) \\ {}^{G}\dot{\mathbf{v}}_I(t) &= \mathbf{C}({}^{I}\mathbf{q}_G(t))^T(\mathbf{a}_m(t) - \mathbf{b}_a(t) - \mathbf{n}_a(t)) + {}^{G}\mathbf{g} \\ \dot{\mathbf{b}}_a(t) &= \mathbf{n}_{wa}(t) \\ {}^{G}\dot{\mathbf{p}}_I(t) &= {}^{G}\mathbf{v}_I(t) \\ {}^{G}\dot{\mathbf{f}}_j(t) &= \mathbf{0}, \quad j = 1, \ldots, N \end{aligned} \quad (3)$$

where, $\boldsymbol{\Omega}(\boldsymbol{\omega}) \triangleq \begin{bmatrix} -\lfloor\boldsymbol{\omega}\rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix}$ for $\boldsymbol{\omega} \in \mathbb{R}^3$, $\lfloor\cdot\rfloor$ denotes the skew-symmetric matrix, while the IMU biases are modelled as random walks driven by white, zero-mean Gaussian noise processes $\mathbf{n}_{wg}(t)$ and $\mathbf{n}_{wa}(t)$, respectively.

As the camera-IMU pair moves, the camera provides measurements of point features extracted from the images. Each such measurement, $\mathbf{z}_j$, is modeled as the perspective projection of the point feature $\mathbf{f}_j$, expressed in the current IMU frame[3] $\{I\}$, onto the image plane:

$$\mathbf{z}_j = \frac{1}{z}\begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{n}_j, \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} \triangleq {}^{I}\mathbf{f}_j = \mathbf{C}({}^{I}\mathbf{q}_G)({}^{G}\mathbf{f}_j - {}^{G}\mathbf{p}_I) \quad (4)$$

[3]For clarity of presentation, we assume that the IMU-camera frames coincide. In practice, we estimate the IMU-camera extrinsics online.

where the measurement noise, $\mathbf{n}_j$, is modeled as zero mean, white Gaussian. For modeling the IMU propagation [see (3)] and camera observations [see (4)], including their error equations and analytical Jacobians, we follow [8].

III. VINS: Observability Analysis Under Specific Motion Profiles

Observability is a fundamental property of a dynamic system and provides important insights. Previous works have studied the observability properties of VINS, and employed the results of their analysis to improve the consistency of the estimator [8]. Specifically, in [7], [8], it was shown that, *for generic motions*, a VINS has four unobservable directions (three for global translation and one for global yaw).

In this paper, we are interested in the case when the VINS is deployed on a ground vehicle, whose motion is approximately planar, and, for the most part, along a straight line (e.g., when moving forward) or a circular arc (e.g., when turning). In particular, we are interested in the impact that such motions have on the VINS's observability properties, and hence the accuracy of the corresponding estimator.

A. Constant Acceleration

Consider that the platform moves with constant *local* linear acceleration (e.g., on a circle), i.e.,

$${}^{I}\mathbf{a}(t) \triangleq \mathbf{C}({}^{I}\mathbf{q}_G(t)){}^{G}\mathbf{a}(t) \equiv {}^{I}\mathbf{a}, \quad \forall\ t \geq t_0 \quad (5)$$

where ${}^{I}\mathbf{a}$ is a constant vector with respect to time, we prove the following theorem:

Theorem 1: The linearized monocular-VINS model of (3) - (4) has the following *additional* unobservable direction, besides the global translation and yaw, *if and only if* condition (5) is satisfied:

$$\mathbf{N}_s = \begin{bmatrix} \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & {}^{G}\mathbf{v}_{I_0}^T & -{}^{I}\mathbf{a}^T & {}^{G}\mathbf{p}_{I_0}^T & {}^{G}\mathbf{f}_1^T & \cdots & {}^{G}\mathbf{f}_N^T \end{bmatrix}^T \quad (6)$$

Proof: See Appendix I.

Remark: The unobservable direction in (6) corresponds to the scale, as shown in [14].

The physical interpretation of Thm. 1 is that, when the local acceleration is non-varying, one *cannot* distinguish the magnitude of the true body acceleration from that of the accelerometer bias, as both of them are, at least temporarily, constant. As a consequence, the magnitude of the true body acceleration can be arbitrary, leading to scale ambiguity.

At this point, we should note that in most cases in practice, a ground vehicle moves on a plane with (almost) constant acceleration, such as when following a straight line path with constant speed or acceleration, or when making turns along a circular arc with constant speed, etc. Based on Thm. 1, these motions render the scale estimated by the VINS inaccurate.

B. No Rotation

Consider that the platform has no rotational motion, i.e., the orientation remains the same across time:

$${}^{I_t}_{G}\mathbf{C} \triangleq \mathbf{C}({}^{I}\mathbf{q}_G(t)) \equiv {}^{I_0}_{G}\mathbf{C}, \quad \forall\ t \geq t_0 \quad (7)$$

where I_t denotes the IMU frame at time t. Then, the following theorem regarding observability holds:

Theorem 2: The linearized VINS model of (3) - (4) has the following *additional* unobservable directions, besides the global translation, *if and only if* condition (7) is satisfied:

$$\mathbf{N}_o = \begin{bmatrix} {}^{I_0}_{G}\mathbf{C}^T & \mathbf{0}_{3\times 3} & \lfloor {}^{G}\mathbf{v}_{I_0} \rfloor & -\lfloor {}^{G}\mathbf{g} \rfloor {}^{I_0}_{G}\mathbf{C}^T & \lfloor {}^{G}\mathbf{p}_{I_0} \rfloor \\ & & & \lfloor {}^{G}\mathbf{f}_1 \rfloor & \cdots & \lfloor {}^{G}\mathbf{f}_N \rfloor \end{bmatrix}^T \quad (8)$$

Proof: See Appendix II.

Remark: The unobservable directions in (8) correspond to all 3 dof of global orientation instead of only yaw [14].

The physical interpretation of Thm. 2 is that, when there is no rotational motion, one *cannot* distinguish the direction of the local gravitational acceleration from that of the accelerometer bias, as both of them are, at least temporarily, constant. As a consequence, the roll and pitch angles become ambiguous.

The motion profile considered in Thm. 2 is the case typically followed by a robot moving on a straight line, or (for a holonomic vehicle) sliding sideways. In such cases, due to the lack of observability, the orientation estimates generated by the VINS become inaccurate.

In summary, moving with constant acceleration or without rotating can introduce extra unobservable directions to the VINS model. At this point, we should reiterate that, although, in practice, these specific motion constraints are never *exactly* satisfied all the time, when the robot (even temporarily) *approximately* follows them, it acquires very limited information along the unobservable directions. This will cause the information (Hessian) matrix estimated by the VINS to be severely ill-conditioned, or even numerically rank-deficient, and hence degrades the localization performance. The impact of such motion on the VINS accuracy is demonstrated experimentally in Sect. V.

Among the two cases of unobservability, that of global orientation (see Thm. 2) is the one that can be easily alleviated by allowing the robot to deviate from its straight-line path. On the other hand, rendering scale observable is quite challenging as it would require the robot to constantly change its acceleration, which would increase the wear and tear of its mobility system. Instead, in what follows, we propose to address this issue and ensure scale observability by extending the VINS to incorporate measurements provided by the robot's wheel odometer.

IV. VINS: Incorporating Extra Information

In order to improve the performance of VINS for wheeled vehicles, we hereafter present our methodology for incorporating two additional sources of information: (i) Odometry measurements and (ii) Planar-motion constraints.

A. VINS with Odometer

Most ground vehicles are equipped with wheel encoders that provide low-frequency, often noisy, and maybe only intermittently, reliable measurements of the motion of each wheel. On the other hand, these measurements contain scale information necessary for improving the accuracy of VINS

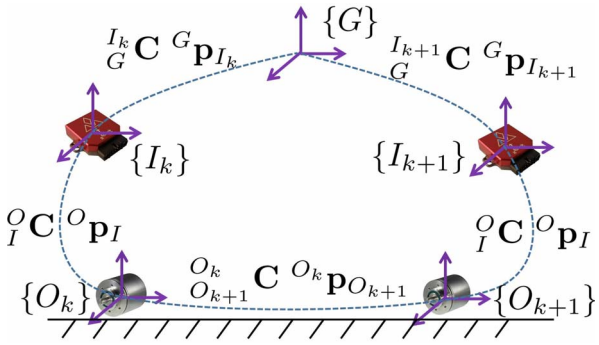

Fig. 1. Geometric relation between the IMU, $\{I\}$, and odometer, $\{O\}$, frames when the robot moves from time step k to $k+1$.

under constant-acceleration motions. In particular, the wheel-encoder data can be transformed into local 2D linear and rotational velocity measurements by employing the odometer intrinsics,[4] i.e.,

$$v = \frac{r_l w_l + r_r w_r}{2}, \quad w = \frac{r_r w_r - r_l w_l}{a} \quad (9)$$

where w_l, w_r are the rotational velocities of the left and right wheels, respectively, r_l, r_r are their corresponding radii, and a denotes the vehicle's baseline.

First, we show that adding these odometric measurements makes the scale of VINS observable:

Theorem 3: Given the odometry measurements of (9), the scale direction in (6) of the linearized VINS model [see (3) - (4)] becomes observable.

Proof: See Appendix III.

In particular, the odometer's linear velocity measurements contain the absolute scale information. Thus, an odometric sensor improves the localization accuracy of VINS not only by recording additional motion measurements, but primarily by providing critical information along the VINS's scale direction which often becomes unobservable due to the vehicle's motion.

In order to process the noisy odometer data in a robust manner, instead of using the velocity measurements in (9), we propose to integrate them and fuse the resulting 2D displacement estimates into the 3D VINS. We start by deriving the measurement model, where we assume that, between consecutive odometer readings, the motion is planar. Hence, the transformation between two consecutive odometer frames, $\{O_k\}$ and $\{O_{k+1}\}$, involves a rotation around only the z axis by an angle ${}^{O_k}\phi_{O_{k+1}}$:

$${}^{O_k}_{O_{k+1}}\mathbf{C} = \mathbf{C}_z({}^{O_k}\phi_{O_{k+1}}) \quad (10)$$

and a translation within the x-y plane, i.e., the first two elements of the translation vector ${}^{O_k}\mathbf{p}_{O_{k+1}}$. Integrating the linear and rotational velocities obtained from the odometer

[4] In this work, we compute offline the batch least-squares estimates of the wheel encoder intrinsics, including the baseline and the radius of each wheel, based on visual, inertial, and odometry data. Subsequently, we consider them as known quantities. Note that these intrinsic states are observable within VINS. The proof is omitted due to lack of space.

provides measurements to these 3-dof quantities, i.e.,

$$\phi_k = {}^{O_k}\phi_{O_{k+1}} + n_\phi \tag{11}$$

$$\mathbf{d}_k = \Lambda {}^{O_k}\mathbf{p}_{O_{k+1}} + \mathbf{n}_d, \quad \Lambda = \begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_2\end{bmatrix}^T \tag{12}$$

where $\begin{bmatrix} n_\phi & \mathbf{n}_d^T \end{bmatrix}^T$ is a 3×1 zero-mean Gaussian noise vector.

Furthermore, from the geometric constraints, depicted in Fig. 1, the transformation between two consecutive odometer frames, at time steps k and $k+1$, can be written as:

$${}^{O_k}_{O_{k+1}}\mathbf{C} = {}^{O}_{I}\mathbf{C}\ {}^{I_k}_{G}\mathbf{C}\ {}^{I_{k+1}}_{G}\mathbf{C}^T\ {}^{O}_{I}\mathbf{C}^T \tag{13}$$

$${}^{O_k}\mathbf{p}_{O_{k+1}} = {}^{O}\mathbf{p}_I + {}^{O}_{I}\mathbf{C}\ {}^{I_k}_{G}\mathbf{C}({}^{G}\mathbf{p}_{I_{k+1}} - {}^{G}\mathbf{p}_{I_k} - {}^{I_{k+1}}_{G}\mathbf{C}^T\ {}^{O}_{I}\mathbf{C}^T {}^{O}\mathbf{p}_I) \tag{14}$$

where ${}^{O}_{I}\mathbf{C}$ and ${}^{O}\mathbf{p}_I$ are the rotation and translation of the odometer-IMU extrinsics,[5] respectively.

Next, we employ (11)-(14) to derive the Jacobians and residuals of the corresponding measurement models to be used by the VINS estimator.

1) Rotational Component: By equating (10) to (13), and employing small-angle approximations in the rotation matrices involved, we obtain the following error equation:

$$\delta\boldsymbol{\phi} = {}^{O}_{I}\mathbf{C}\ \delta\boldsymbol{\theta}_{I_k} - {}^{O}_{I}\mathbf{C}\,{}^{I_k}_{G}\hat{\mathbf{C}}\,{}^{I_{k+1}}_{G}\hat{\mathbf{C}}^T\delta\boldsymbol{\theta}_{I_{k+1}} - n_\phi\mathbf{e}_3 \tag{15}$$

$$\text{with} \quad \lfloor\delta\boldsymbol{\phi}\rfloor = \mathbf{I}_3 - \mathbf{C}_z(\phi_k)\ {}^{O_k}_{O_{k+1}}\hat{\mathbf{C}}^T$$

$${}^{O_k}_{O_{k+1}}\hat{\mathbf{C}} = {}^{O}_{I}\mathbf{C}\ {}^{I_k}_{G}\hat{\mathbf{C}}\ {}^{I_{k+1}}_{G}\hat{\mathbf{C}}^T\ {}^{O}_{I}\mathbf{C}^T$$

where $\hat{\mathbf{C}}$ denotes the estimate of the rotation matrix $\mathbf{C}$, and $\delta\boldsymbol{\theta}$ is the error state of the corresponding quaternion parameterization. The third element of the vector $\delta\boldsymbol{\phi}$ represents the angular error between the measured and the estimated in-plane rotation. Multiplying both sides of (15) with $\mathbf{e}_3^T$ yields the Jacobians and residual:

$$\mathbf{H}_{\delta\theta_{I_k}} = \mathbf{e}_3^T\ {}^{O}_{I}\mathbf{C}\ , \quad \mathbf{H}_{\delta\theta_{I_{k+1}}} = -\mathbf{e}_3^T\ {}^{O}_{I}\hat{\mathbf{C}}\,{}^{I_k}_{G}\hat{\mathbf{C}}\,{}^{I_{k+1}}_{G}\hat{\mathbf{C}}^T$$

$$\mathbf{r} = \mathbf{e}_3^T\ \delta\boldsymbol{\phi} \tag{16}$$

2) Translational Component: By substituting (14) into (12) and linearizing, it is straightforward to obtain the following Jacobians and residual:

$$\mathbf{H}_{\delta\theta_{I_k}} = \Lambda\,{}^{O}_{I}\mathbf{C}\lfloor\boldsymbol{\xi}\rfloor,\ \mathbf{H}_{\delta\theta_{I_{k+1}}} = \Lambda\,{}^{O}_{I}\mathbf{C}\,{}^{I_k}_{G}\hat{\mathbf{C}}\,{}^{I_{k+1}}_{G}\hat{\mathbf{C}}^T\lfloor{}^{O}_{I}\mathbf{C}^T {}^{O}\mathbf{p}_I\rfloor$$

$$\mathbf{H}_{p_{I_k}} = -\Lambda\,{}^{O}_{I}\mathbf{C}\,{}^{I_k}_{G}\hat{\mathbf{C}}\ , \quad \mathbf{H}_{p_{I_{k+1}}} = \Lambda\,{}^{O}_{I}\mathbf{C}\,{}^{I_k}_{G}\hat{\mathbf{C}}$$

$$\mathbf{r} = \mathbf{d}_k - \Lambda({}^{O}\mathbf{p}_I + {}^{O}_{I}\mathbf{C}\,\boldsymbol{\xi}) \tag{17}$$

$$\text{with} \quad \boldsymbol{\xi} \triangleq {}^{I_k}_{G}\hat{\mathbf{C}}({}^{G}\hat{\mathbf{p}}_{I_{k+1}} - {}^{G}\hat{\mathbf{p}}_{I_k} - {}^{I_{k+1}}_{G}\hat{\mathbf{C}}^T\,{}^{O}_{I}\mathbf{C}^T {}^{O}\mathbf{p}_I).$$

Finally, (16) and (17) represent stochastic constraints between the poses of the platform, and can be combined in a tightly-coupled manner into standard VINS estimators.

B. mVINS: VINS within a Manifold

In many cases in practice, the trajectory of a moving object often lies within some manifold. Ground vehicles, for example, travel mostly on a plane, especially when navigating indoors. The knowledge of this specific motion manifold can provide additional information for improving the localization accuracy of VINS.

[5]In this work, we compute offline the batch least-squares estimates of the odometer-IMU extrinsics, based on visual, inertial, and odometry data. Subsequently, we consider them as known quantities. The observable directions of the VINS with odometer extrinsics are presented in [15].

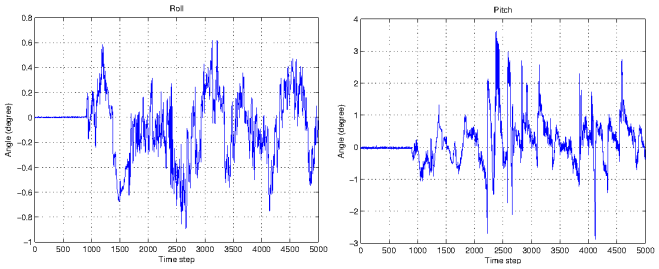

Fig. 2. The roll (left) and pitch (right) angles in degrees across time, when the robot is moving on a flat surface. The mean is -0.08 and 0.2 degree, and the standard deviation is 0.3 and 0.7 degree, respectively.

A motion manifold can be described mathematically as geometric constraints, $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, where $\mathbf{g}$ is, in general, a nonlinear function of the state $\mathbf{x}$. There are two approaches for incorporating such information into a VINS:

1) Deterministic Constraints: A standard VINS estimator (e.g., a filter or a smoother) optimizes a cost function $\mathcal{C}(\mathbf{x})$ arising from the information in the sensor (visual, inertial, and potentially odometric) data (e.g., [4], [6], [10], [12]), while the motion manifold is described as a deterministic constraint of the optimization problem, i.e.,

$$\begin{aligned} \min \quad & \mathcal{C}(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}) = \mathbf{0} \end{aligned} \tag{18}$$

For VINS, the cost function $\mathcal{C}(\mathbf{x})$ typically takes the form of nonlinear least squares, and (18) can be solved by employing iterative Gauss-Newton minimization [16].

2) Stochastic Constraints: In practice, the motion manifold is never exactly satisfied. Fig. 2 depicts the platform's roll and pitch angles across time, when a ground robot (a Pioneer 3 in our case) is moving on a flat surface. During an ideal planar motion, the roll and pitch angles would have remained constant. As evident, however, this is not the case in practice due to the vibrations of the moving platform and the unevenness of the surface. To account for such deviations, we model the manifold as a stochastic constraint $\mathbf{g}(\mathbf{x}) = \mathbf{n}$, where $\mathbf{n}$ is assumed to be a zero-mean Gaussian process with covariance $\mathbf{R}$, and incorporate this information as an additional cost term:

$$\min \quad \mathcal{C}(\mathbf{x}) + ||\mathbf{g}(\mathbf{x})||^2_{\mathbf{R}} \tag{19}$$

Note that (19) can be solved by employing standard VINS estimators. Moreover, this stochastic approach (as compared to the deterministic one) provides more flexibility for rejecting false information due to outliers. Specifically, we employ the Mahalanobis distance test to detect and temporally remove the constraints when they are least likely (in the probabilistic sense) to be satisfied (e.g., when the robot goes over a bump).

In what follows, we focus on a specific manifold, the one corresponding to planar motion, and present in detail

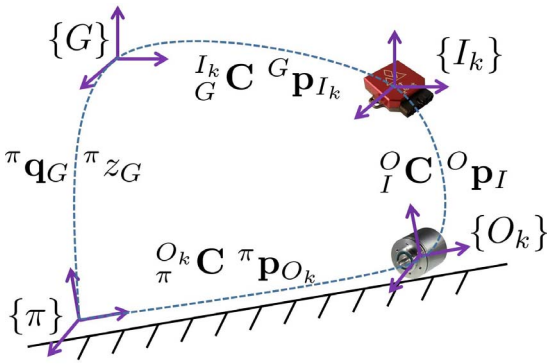

Fig. 3. Geometric relationship between the IMU, $\{I\}$, odometer, $\{O\}$, and plane, $\{\pi\}$, frames when the robot moves on the plane, at time step k.

how to employ this information in VINS. The frame of the plane, $\{\pi\}$, is defined so that the $x-y$ plane coincides with the physical plane. We parameterize the plane with a 2-dof quaternion, ${}^{\pi}\mathbf{q}_G$, representing the orientation between the plane frame and the global frame, and a scalar ${}^{\pi}z_G$, denoting the perpendicular distance from the origin of the global frame to the plane. The error state of the quaternion ${}^{\pi}\mathbf{q}_G$ is defined as a 2×1 vector $\delta\boldsymbol{\theta}_\pi$ so that the error quaternion is given by $\delta\mathbf{q} \triangleq \begin{bmatrix}\frac{1}{2}\delta\boldsymbol{\theta}_\pi^T & 0 & 1\end{bmatrix}^T$. Note that our parameterization agrees with the fact that a plane in the 3D space has 3 dof. As depicted in Fig. 3, we can express the geometric constraint of the odometer frame, $\{O\}$, moving within the plane, as:

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} \Lambda_I^O \mathbf{C}_G^{I_k} \mathbf{C}_G^{\pi} \mathbf{C}^T \mathbf{e}_3 \\ {}^{\pi}z_G + \mathbf{e}_3^T {}_G^{\pi}\mathbf{C}({}^G\mathbf{p}_{I_k} - {}_G^{I_k}\mathbf{C}\,{}_I^O\mathbf{C}^T\,{}^O\mathbf{p}_I) \end{bmatrix} = \mathbf{0} \quad (20)$$

where the first block element (2×1 vector) corresponds to the planar rotational constraint, i.e., that the roll and pitch angles are zero between $\{\pi\}$ and $\{O\}$, while the second block element (scalar) corresponds to the planar translational constraint, i.e., that the position displacement along the z-axis is zero between $\{\pi\}$ and $\{O\}$.

Lastly, we provide the Jacobians of the planar model, derived from (20), employed by the VINS estimator:

i) Rotational component:

$$\begin{aligned} \mathbf{H}_{\delta\theta_{I_k}} &= \Lambda_I^O \mathbf{C} \lfloor {}_G^{I_k}\hat{\mathbf{C}}\, {}_G^{\pi}\hat{\mathbf{C}}^T \mathbf{e}_3 \rfloor \\ \mathbf{H}_{\delta\theta_\pi} &= \Lambda_I^O \mathbf{C}_G^{I_k} \hat{\mathbf{C}}_G^{\pi} \hat{\mathbf{C}}^T \begin{bmatrix} -\mathbf{e}_2 & \mathbf{e}_1 \end{bmatrix} \end{aligned} \quad (21)$$

ii) Translational component:

$$\begin{aligned} \mathbf{H}_{\delta\theta_{I_k}} &= \mathbf{e}_3^T {}_G^{\pi}\hat{\mathbf{C}}\, {}_G^{I_k}\hat{\mathbf{C}}^T \lfloor {}_I^O\mathbf{C}^T\,{}^O\mathbf{p}_I \rfloor \\ \mathbf{H}_{\delta\theta_\pi} &= \left({}^G\hat{\mathbf{p}}_{I_k} - {}_G^{I_k}\hat{\mathbf{C}}\, {}_I^O\mathbf{C}^T\,{}^O\mathbf{p}_I\right)^T {}_G^{\pi}\hat{\mathbf{C}}^T \begin{bmatrix} -\mathbf{e}_2 & \mathbf{e}_1 \end{bmatrix} \\ \mathbf{H}_{p_{I_k}} &= \mathbf{e}_3^T {}_G^{\pi}\hat{\mathbf{C}}\,, \quad \mathbf{H}_{z_G} = 1. \end{aligned} \quad (22)$$

V. Experimental Results

We aim to examine the impact of different motions on the localization accuracy of VINS, as well as to validate the proposed methods for incorporating information from the odometer and the motion manifold. Note that our observability findings and the proposed methods are generic and not restricted to any particular VINS estimator. In our experiments, we chose the square-root inverse sliding window filter (SR-ISWF) [6] that is implemented with single-precision data types, in order to obtain highly-efficient localization results on mobile devices.[6]

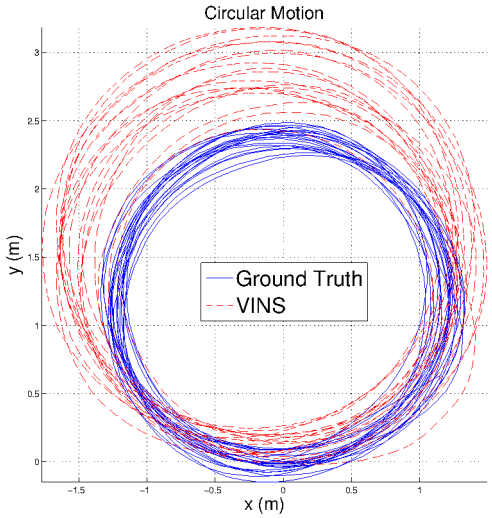

Fig. 4. $x-y$ overview of the Pioneer robot's trajectory during the circular-motion experiment: The ground truth is shown in blue solid line, while the VINS estimate is shown in red dashed line.

Our testing platform involves commercial-grade sensors and CPU: A Pioneer 3 DX robot,[7] with a Project Tango tablet [17] mounted on it for visual and inertial sensing, as well as for processing. This tablet has a 2.3 GHz quad-core NVIDIA Tegra K1 CPU and 4 GB on-chip RAM, and is able to record: (i) MEMS-based IMU data, at 100 Hz, and (ii) Grayscale images from its wide field-of-view camera, with a resolution of 640×480, at 30 Hz. Around 200 FAST corners [18] are extracted from each image and tracked using the Kanade-Lucas-Tomasi (KLT) algorithm [19] at a frequency of 15 Hz. Then, a 2-pt RANSAC [20] is used for initial outlier rejection. The SR-ISWF estimator maintains a sliding window of 15 poses, which are selected at a frequency of about 7 Hz (depending on the motion).

A. Assessment of the Motion's Impact

We compare the localization results of the VINS, within the same environment, between two motion profiles: (i) Generic motion, where we hand-hold the tablet and walk regularly, and (ii) Constant (local) acceleration motion, where the tablet is mounted on the Pioneer robot that follows a circular motion. Fig. 4 illustrates the VICON[8] ground truth and the VINS filter's estimated trajectory. Note that as expected in practice, the vehicle's path (ground truth) is not a perfect circle; Instead, it only approximately follows one of the special motions considered here. Regardless, as evident from Fig. 4, significant scale error appears in the VINS estimates, which validates the conclusion of Thm. 1.

[6] Similar results were observed when using the native Google Tango [17] VINS onboard the tablet, and are omitted from here due to lack of space.

[7] http://www.mobilerobots.com/ResearchRobots/PioneerP3DX.aspx

[8] http://www.vicon.com/

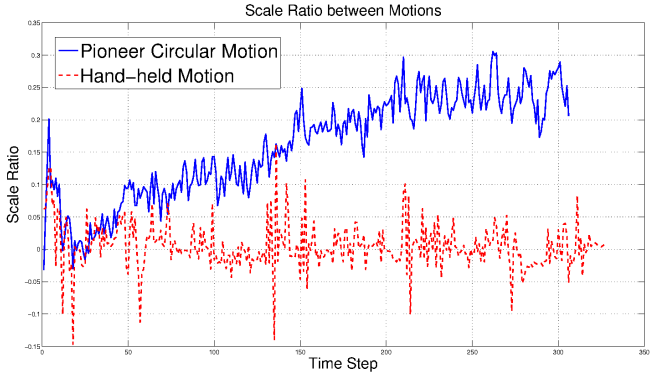

Fig. 5. Scale ratio results for: (i) Pioneer circular motion (blue solid line) with mean 0.16 and std 0.08; (ii) Hand-held motion (red dashed line) with mean 3e-3 and std 0.03.

We further compare the *scale ratio* between the two motions considered, as shown in Fig. 5. The scale ratio is computed as the estimated distance between consecutive poses divided by that of the ground truth, and shifted by one:

$$SR = \frac{d_{est}}{d_{gt}} - 1 \tag{23}$$

which measures the quality of the scale estimates, i.e., the closer this quantity is to zero, the better is the scale estimate. As evident, the scale ratio corresponding to the hand-held motion stays around zero, while that of the circular motion drifts away. Finally, the positioning root mean square error (RMSE) of the hand-held vs. circular motion is 14 cm vs. 81 cm, respectively. These results confirm that when a vehicle undergoes (even approximately) special motions, the reduced information available to the VINS along the unobservable directions significantly degrades the localization accuracy of the corresponding estimator.

B. System Performance Test

We further test the localization accuracy of our system on the Pioneer robot. Five datasets are collected by driving the Pioneer each time for $\sim$ 1 km through a large building. In addition to the IMU-camera data, the Pioneer wheel encoders provide readings at 10 Hz. We compare the localization results among the following setups: (i) VINS only, (ii) VINS plus odometer (VO), (iii) VINS plus odometer plus deterministic planar constraint (VOD), and (iv) VINS plus odometer plus stochastic planar constraint (VOS). The ground truth is computed from the batch least squares (BLS) offline, using all available (visual, inertial, and odometric) measurements.

Fig. 6 illustrates the estimated trajectories, overlayed on the building's floor plan as reference. As evident, the pure VINS suffers from very large errors due to the restricted motion (mostly constant-speed, on straight lines), while as more information becomes available, the positioning accuracy improves significantly. Also, the VOS outperforms the VOD, since the stochastic constraint better models the *approximately* planar motion due to the vibrations of the moving platform and the unevenness of the ground surface. Table I compares quantitatively the positioning error between different methods across all datasets (DS), where each block contains the following RMSE (in meters) results: xy - z - xyz total position - as percentage of the total distance traveled. From these results, we draw the following conclusions: First, between VO and VINS, when the odometer measurements are added, the x-y positioning accuracy is improved dramatically, since more scale information is injected. Second, by comparing VOD and VOS to VINS and VO, it is evident that the planar motion constraints improve mostly the estimates in the z direction, as the error along the perpendicular direction is restricted by the constraint. Lastly, the stochastic constraint of VOS consistently improves the positioning accuracy, while the deterministic one of VOD has a negative impact, due to its modeling error.

In terms of efficiency, our system runs in real time on the tablet. Specifically, the whole VINS pipeline is taking 68 msec per cloned pose, including the 36 msec spent on the SR-ISWF filter update. Note also that our efficient implementation of the proposed methods (for processing odometer data and planar constraints) takes less than 1 msec for each. Overall, $\sim 50\%$ of the total CPU is used by our program when performing updates at $\sim$ 7 Hz.

Finally, it is worth mentioning that our system is able to work robustly in both indoor and outdoor environments. Demonstrating videos are available at: http://mars.cs.umn.edu/research/VINSodometry.php

VI. Conclusions and Future Work

In this paper, we proved that the VINS scale, or 2 additional dof of its global orientation, become unobservable when the robot moves with constant acceleration, or it is not rotating, respectively. For this reason, and as demonstrated in our experiments, directly employing VINS on a wheeled robot results in inaccurate pose estimates. To address this issue, we incorporated wheel-encoder measurements into VINS and showed that the scale becomes observable. Furthermore, we introduced mVINS that properly models the ground robot's almost-planar motion and directly employs this information in the estimator. Experimental results showed that special motions indeed lead to larger positioning errors when using VINS on a wheeled robot. Incorporating, however, odometry measurements, as well as stochastic constraints modeling the vehicle's planar motion, provide additional information and lead to significant improvements in positioning accuracy.

As part of our future work, we plan to extend the proposed mVINS such that it allows to model more complex robot motions (e.g., moving between multiple flat surfaces and climbing stairs), as well as to compensate for wheel slippage.

Appendix I
Proof of Theorem 1

In this section, we prove that the scale in (6) is an unobservable direction of the VINS model, if and only if the platform is moving with constant *local* linear acceleration [see (5)]. We follow the approach presented in [8], that examines the right null space of the observability matrix of the corresponding linearized VINS model. As is the case

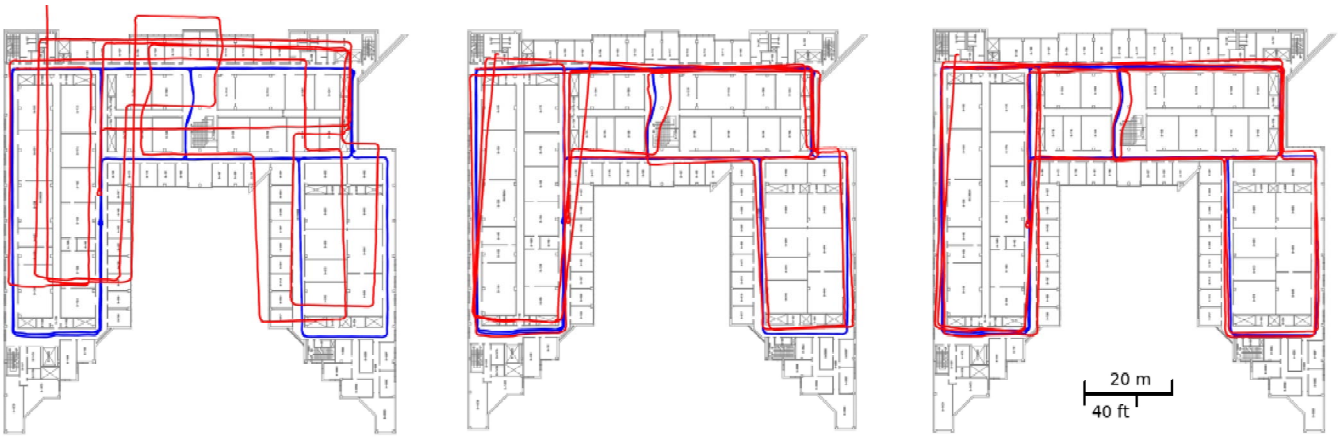

Fig. 6. Illustration of the indoor Pioneer navigation trajectories, shown in red, estimated by the VINS only (left), the VOD (middle), and the VOS (right), overlayed on the floor plan. The ground truth, computed from the BLS method offline, is shown in blue.

TABLE I

COMPARISONS: POSITIONING RMSE (IN METERS) OF DIFFERENT METHODS ACROSS DATASETS (XY - Z - **XYZ** - %)

DS	Path (m)	VINS	VO	VOD	VOS
1	1080	9.8 - 1.5 - **9.9** - 0.91%	2.4 - 1.6 - **2.9** - 0.27%	4.3 - 0.1 - **4.3** - 0.4%	2.7 - 0.08 - **2.7** - 0.25%
2	876	13.8 - 1.1 - **13.8** - 1.6%	1.9 - 1.2 - **2.2** - 0.26%	4.5 - 0.14 - **4.5** - 0.52%	1.9 - 0.09 - **1.9** - 0.22%
3	954	8.3 - 1.2 - **8.4** - 0.88%	3.2 - 1.5 - **3.5** - 0.37%	7.8 - 0.22 - **7.8** - 0.82%	3.1 - 0.11 - **3.1** - 0.32%
4	1048	11.7 - 0.99 - **11.7** - 1.1%	3.7 - 1.0 - **3.8** - 0.37%	7.6 - 0.26 - **7.6** - 0.73%	3.6 - 0.07 - **3.6** - 0.34%
5	1034	9.7 - 0.99 - **9.8** - 0.94%	1.6 - 1.4 - **2.1** - 0.2%	3.2 - 0.12 - **3.2** - 0.31%	1.6 - 0.08 - **1.6** - 0.15%

in [8], and for clarity of presentation, we include only one feature in the state vector (the extension to multiple features is straightforward).

As previously shown (see (51) in [8]), any block row, $\mathbf{M}_k$, of the observability matrix has the following structure:

$$\mathbf{M}_k = \mathbf{H}_k \mathbf{\Phi}_{k,1} = \mathbf{\Gamma}_1 \begin{bmatrix} \mathbf{\Gamma}_2 & \mathbf{\Gamma}_3 & -\delta t_k \mathbf{I}_3 & \mathbf{\Gamma}_4 & -\mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} \quad (24)$$

for any time $t_k \geq t_0$, with the matrices $\mathbf{\Gamma}_i, i = 1, \ldots, 4$, defined by (52)-(55) in [8]. From the property of the observability matrix, the scale direction, $\mathbf{N}_s$, is unobservable, if and only if, $\mathbf{M}_k \mathbf{N}_s = \mathbf{0}$ [21]. From (24) and (6), together with the definition of the matrices $\mathbf{\Gamma}_i$, we obtain:

$$\mathbf{M}_k \mathbf{N}_s = \mathbf{\Gamma}_1 (-{}^G\mathbf{v}_{I_0} \delta t_k - \mathbf{\Gamma}_4 {}^I\mathbf{a} - {}^G\mathbf{p}_{I_0} + {}^G\mathbf{f}) \quad (25)$$

$$\text{with} \quad \mathbf{\Gamma}_4 {}^I\mathbf{a} = \int_{t_0}^{t_k} \int_{t_0}^{s} {}^G_{I_\tau}\mathbf{C} \, d\tau ds \cdot {}^I\mathbf{a} \quad (26)$$

$$= \int_{t_0}^{t_k} \int_{t_0}^{s} {}^G_{I_\tau}\mathbf{C} {}^I\mathbf{a} \, d\tau ds \quad (27)$$

$$= \int_{t_0}^{t_k} \int_{t_0}^{s} {}^G_{I_\tau}\mathbf{C} {}^{I_\tau}\mathbf{a}(\tau) \, d\tau ds \quad (28)$$

$$= \int_{t_0}^{t_k} \int_{t_0}^{s} {}^G\mathbf{a}(\tau) \, d\tau ds \quad (29)$$

$$= \int_{t_0}^{t_k} ({}^G\mathbf{v}_{I_s} - {}^G\mathbf{v}_{I_0}) \, ds \quad (30)$$

$$= {}^G\mathbf{p}_{I_k} - {}^G\mathbf{p}_{I_0} - {}^G\mathbf{v}_{I_0} \delta t_k \quad (31)$$

where the equality from (27) to (28) holds if and only if the constant acceleration assumption in (5) is satisfied. Substituting (31) into (25) yields:

$$\mathbf{M}_k \mathbf{N}_s = \mathbf{\Gamma}_1 ({}^G\mathbf{f} - {}^G\mathbf{p}_{I_k}) = \mathbf{H}_{c,k} {}^{I_k}_G\mathbf{C} ({}^G\mathbf{f} - {}^G\mathbf{p}_{I_k}) = \mathbf{H}_{c,k} {}^{I_k}\mathbf{f} = \mathbf{0} \quad (32)$$

where the last equality holds since the camera perspective-projection Jacobian matrix, $\mathbf{H}_{c,k}$, has as its right null space the feature position in the IMU frame (see (30) in [8]).

Lastly, this new unobservable direction is in addition to the four directions corresponding to global translation and yaw, i.e., $\mathbf{N}_s$ and $\mathbf{N}_1$ in (57) of [8] are independent, since the 4th block element of $\mathbf{N}_1$ is zero while that of $\mathbf{N}_s$ is not.

APPENDIX II
PROOF OF THEOREM 2

In what follows, we prove that the 3-dof global orientation in (8) is an unobservable direction of the VINS model, if and only if the platform does not rotate [see (7)]. Similarly to the proof presented in Appendix I, in this case, we need to show that $\mathbf{M}_k \mathbf{N}_o = \mathbf{0}$. From (24) and (8), together with the definition of the matrices $\mathbf{\Gamma}_i, i = 1, \ldots, 4$, we obtain:

$$\mathbf{M}_k \mathbf{N}_o = \mathbf{\Gamma}_1 (\mathbf{\Gamma}_4 {}^{I_0}_G\mathbf{C} - \frac{1}{2}\delta t_k^2 \mathbf{I}_3) \lfloor {}^G\mathbf{g} \rfloor \quad (33)$$

$$= \mathbf{\Gamma}_1 (\int_{t_0}^{t_k} \int_{t_0}^{s} {}^G_{I_\tau}\mathbf{C} \, d\tau ds \cdot {}^{I_0}_G\mathbf{C} - \frac{1}{2}\delta t_k^2 \mathbf{I}_3) \lfloor {}^G\mathbf{g} \rfloor \quad (34)$$

$$= \mathbf{\Gamma}_1 (\int_{t_0}^{t_k} \int_{t_0}^{s} {}^G_{I_0}\mathbf{C} \, d\tau ds \cdot {}^{I_0}_G\mathbf{C} - \frac{1}{2}\delta t_k^2 \mathbf{I}_3) \lfloor {}^G\mathbf{g} \rfloor \quad (35)$$

$$= \mathbf{\Gamma}_1 (\int_{t_0}^{t_k} \int_{t_0}^{s} 1 \, d\tau ds \cdot {}^G_{I_0}\mathbf{C} {}^{I_0}_G\mathbf{C} - \frac{1}{2}\delta t_k^2 \mathbf{I}_3) \lfloor {}^G\mathbf{g} \rfloor$$

$$= \mathbf{\Gamma}_1 (\frac{1}{2}\delta t_k^2 \mathbf{I}_3 - \frac{1}{2}\delta t_k^2 \mathbf{I}_3) \lfloor {}^G\mathbf{g} \rfloor = \mathbf{0} \quad (36)$$

where the equality from (34) to (35) holds if and only if the no rotation (i.e., constant orientation) assumption in (7) is satisfied.

Lastly, these new unobservable directions are in addition to the three directions corresponding to global translation, i.e., $\mathbf{N}_o$ and $\mathbf{N}_{t,1}$ in (57) of [8] are independent, since the first block element of $\mathbf{N}_{t,1}$ is zero while that of $\mathbf{N}_o$ is a (full-rank) rotational matrix.

APPENDIX III
PROOF OF THEOREM 3

We hereafter prove that the scale in (6) is observable for the VINS model when an odometer is present. Specifically, the odometer provides measurements of the 2-dof planar component of the robot's linear velocity:

$$\mathbf{v}_k = \Lambda^{O_k}\mathbf{v}_{O_k} = \Lambda_I^O\mathbf{C}({}^{I_k}_{G}\mathbf{C}\,{}^{G}\mathbf{v}_{I_k} + \lfloor\boldsymbol{\omega}_m(t_k) - \mathbf{b}_g(t_k)\rfloor\,{}^{I}\mathbf{p}_O)$$

from which we obtain the following measurement Jacobians with respect to the states involved:

$$\begin{aligned}\mathbf{H}^O_{\delta\theta} &= \Lambda_I^O\mathbf{C}\lfloor{}^{I_k}_{G}\mathbf{C}\,{}^{G}\mathbf{v}_{I_k}\rfloor\ ,\quad \mathbf{H}^O_{b_g} = \Lambda_I^O\mathbf{C}\lfloor{}^{I}\mathbf{p}_O\rfloor\\ \mathbf{H}^O_{v} &= \Lambda_I^O\mathbf{C}\,{}^{I_k}_{G}\mathbf{C}\end{aligned} \tag{37}$$

The odometry measurements provide extra block rows in the observability matrix, in addition to the ones corresponding to the camera observations [see (24)]. From (37) and the analytical form of the state transition matrix, $\mathbf{\Phi}_{k,1}$ (see (44) in [8]), it can be verified that these extra block rows have the following structure:

$$\begin{aligned}\mathbf{M}^O_k &= \mathbf{H}^O_k\mathbf{\Phi}_{k,1}\\ &= \mathbf{\Gamma}^O_1\begin{bmatrix}\mathbf{\Gamma}^O_2 & \mathbf{\Gamma}^O_3 & {}^{I_k}_{G}\mathbf{C} & {}^{I_k}_{G}\mathbf{C}\mathbf{\Phi}^{(3,4)}_{k,1} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3}\end{bmatrix}\\ \text{with}\ \ \mathbf{\Gamma}^O_1 &= \Lambda^O_I\mathbf{C}\\ \mathbf{\Gamma}^O_2 &= \lfloor{}^{I_k}_{G}\mathbf{C}\,{}^{G}\mathbf{v}_{I_k}\rfloor\mathbf{\Phi}^{(1,1)}_{k,1} + {}^{I_k}_{G}\mathbf{C}\mathbf{\Phi}^{(3,1)}_{k,1}\\ \mathbf{\Gamma}^O_3 &= \lfloor{}^{I_k}_{G}\mathbf{C}\,{}^{G}\mathbf{v}_{I_k}\rfloor\mathbf{\Phi}^{(1,2)}_{k,1} + \lfloor{}^{I}\mathbf{p}_O\rfloor + {}^{I_k}_{G}\mathbf{C}\mathbf{\Phi}^{(3,2)}_{k,1}\end{aligned} \tag{38}$$

for any time $t_k \geq t_0$, with $\mathbf{\Phi}^{(i,j)}_{k,1}$ denoting the (i,j)-th block element of the state transition matrix $\mathbf{\Phi}_{k,1}$. From Thm. 1, the scale becomes unobservable if and only if the acceleration is constant. Therefore, it suffices to show that $\mathbf{M}^O_k\mathbf{N}_s \neq \mathbf{0}$ when (5) is satisfied. Specifically:

$$\mathbf{M}^O_k\mathbf{N}_s = \mathbf{\Gamma}^O_1({}^{I_k}_{G}\mathbf{C}\,{}^{G}\mathbf{v}_{I_0} - {}^{I_k}_{G}\mathbf{C}\mathbf{\Phi}^{(3,4)}_{k,1}\,{}^{I}\mathbf{a}) \tag{39}$$

$$= \Lambda^O_I\mathbf{C}^{I_k}_{G}\mathbf{C}({}^{G}\mathbf{v}_{I_0} + \int_{t_0}^{t_k}{}^{G}_{I_\tau}\mathbf{C}\ d\tau\cdot{}^{I}\mathbf{a}) \tag{40}$$

$$= \Lambda^O_I\mathbf{C}^{I_k}_{G}\mathbf{C}^{G}\mathbf{v}_{I_k} = \Lambda^{O_k}\mathbf{v}_{I_k} \tag{41}$$

where we have followed the same reasoning as in (26)-(30). The quantity in (41) is non-zero, if the velocity of the IMU frame, expressed in the odometer frame, does not vanish along the $x-y$ directions, i.e., if the platform has translational motion along the horizontal plane. Under this condition, which is satisfied in practice as long as the vehicle does not stay static forever, the odometer measurements make the scale observable.

REFERENCES

[1] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (slam): Part i the essential algorithms," *IEEE Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–110, june 2006.

[2] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (slam): part ii," *IEEE Robotics Automation Magazine*, vol. 13, no. 3, pp. 108–117, Sept 2006.

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

[4] 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, Mar. 2015.

[5] M. Li, B. H. Kim, and A. I. Mourikis, "Real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera," in *Proc. of the IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, May 6-10 2013, pp. 4697–4704.

[6] K. J. Wu, A. Ahmed, G. Georgiou, and S. I. Roumeliotis, "A square root inverse filter for efficient vision-aided inertial navigation on mobile devices," in *Proc. of Robotics: Science and Systems*, Rome, Italy, Jul. 13-17 2015.

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

[8] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Consistency analysis and improvement of vision-aided inertial navigation," *IEEE Trans. on Robotics*, vol. 30, no. 1, pp. 158–176, Feb. 2014.

[9] A. Martinelli, "Closed-form solution of visual-inertial structure from motion," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 138–152, 2013.

[10] 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.

[11] 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, no. 2, pp. 216–235, February 2012.

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

[13] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Hamburg, Germany, Sep. 28 – Oct. 2 2015, pp. 298–304.

[14] K. J. Wu and S. I. Roumeliotis, "Unobservable directions of VINS under special motions," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep., September 2016. [Online]. Available: http://mars.cs.umn.edu/research/VINSodometry.php

[15] C. X. Guo, F. M. Mirzaei, and S. I. Roumeliotis, "An analytical least-squares solution to the odometer-camera extrinsic calibration problem," in *Proc. of the IEEE International Conference on Robotics and Automation*, Saint Paul, Minnesota, May 14–18 2012, pp. 3962–3968.

[16] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Springer, 2006.

[17] "Project tango https://www.google.com/atap/projecttango," *Online*.

[18] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. of the 9th European Conference on Computer Vision*, vol. 3951, Graz, Austria, May 7–13 2006, pp. 430–443.

[19] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of the International Joint Conference on Artificaial Intelligence*, Vancouver, British Columbia, Aug. 24–28 1981, pp. 674–679.

[20] L. Kneip, M. Chli, and R. Siegwart, "Robust real-time visual odometry with a single camera and an IMU," in *Proc. of The British Machine Vision Conference*, Dundee, Scotland, Aug. 2011, pp. 1–11.

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