Exploiting Urban Scenes for Vision-aided Inertial Navigation

Dimitrios G. Kottas and Stergios I. Roumeliotis
Department of Computer Science and Engineering
University of Minnesota, Minneapolis, MN
{dkottas|stergios} at cs.umn.edu.

***Abstract*—This paper addresses the problem of visual-inertial navigation when processing camera observations of both point and line features detected within a Manhattan world. First, we prove that the observations of: (i) a single point, and (ii) a single line of *known* direction perpendicular to gravity (e.g., a non-vertical structural line of a building), provide sufficient information for rendering all degrees of freedom of a vision-aided inertial navigation system (VINS) observable, up to global translations. Next, we examine the observability properties of the linearized system employed by an extended Kalman filter (EKF) for processing line observations of known direction, and show that the rank of the corresponding observability matrix erroneously increases. To address this problem, we introduce an elegant modification which enforces that the linearized EKF system has the correct number of unobservable directions, thus improving its consistency. Finally, we validate our findings experimentally in urban scenes and demonstrate the superior performance of the proposed VINS over alternative approaches.**

I. Introduction and Related Work

Current approaches for estimating the position and attitude of a vehicle navigating in 3D rely on inertial measurement units (IMUs) that provide rotational velocity and linear acceleration measurements. However, when using commercial-grade IMUs, the integration of the noise and bias in the inertial measurements renders the resulting estimates unreliable even after a short period of time. For this reason, most inertial navigation systems (INSs) often rely on GPS to provide periodic corrections in what is known as GPS-aided INS. In many cases (e.g., indoors or within urban canyons), however, the GPS signals are either unavailable or unreliable and thus alternative means for aiding INS are necessary. Vision-aided INS (VINS) employs cameras to extract motion (and often structure) information from images of the surroundings of the IMU-camera pair, necessary for providing periodic corrections to the inertial estimates. Recent advances in VINS have led to successful applications to ground [14, 7], aerial [20, 19], and space exploration [15] vehicles.

Most work to date on VINS has focused on *point* features that are tracked through sequences of images, or re-detected when returning to the same location. Point features can be found in most natural and man-made environments, and as shown in [7, 10, 8], they provide sufficient information for estimating all degrees of freedom (d.o.f.) of a VINS, except the global position and yaw. In the absence of alternative sources of information, however, the yaw errors typically grow linearly with time and can cause even faster increase of the position errors. One way to bound the uncertainty in the estimated heading is to use (straight) *line* features of *known direction*. Lines,[1] which are abundant in structured environments, are typically aligned with gravity (e.g., vertical edges of a building), or are perpendicular to it (e.g., edges along the length of a corridor). An environment where most lines are parallel to one of the three cardinal directions is often referred to as a "Manhattan world", and there exist numerous algorithms for estimating the attitude of a camera based on their observations [21, 3]. In particular, it is well known that measurements of two sets of parallel lines (or equivalently of two vanishing points) suffice for determining all three d.o.f. of the camera's attitude [3].

Employing line features for improving the accuracy of VINS has received limited attention to date. In one of the earlier works [16], it was shown that all 6 d.o.f. of a bias-free VINS become observable when detecting lines of known direction *and* position, and a Luenberger observer was proposed for fusing them. The assumption of an *a priori* known map of lines was removed in [9] where a single vanishing point, corresponding to lines of known direction, was used along with measurements of the gravity to determine the 3 d.o.f. attitude of a *static*, bias-free, IMU-camera pair. More recently, line observations have been employed for reducing attitude errors by either requiring two or more vanishing points to completely recover the camera's global attitude (e.g., the loosely-coupled filter of [17]), or by directly processing line measurements along at least two directions [6]. In both cases, however, the impact of additional information that may become available through the observation of point features is *not* considered when examining the conditions under which the IMU-camera's attitude becomes observable.

To the best of our knowledge, the only work that considers both point *and* line observations for improving the accuracy of VINS is that of Williams et al. [20]. It focuses, however, only on *vertical* lines whose observations allow improving the

This work was supported by the University of Minnesota through the Digital Technology Center (DTC) and AFOSR (FA9550-10-1-0567).

[1] Note that contrary to point features, edges and in particular straight lines, can be extracted in a stable manner over different viewing angles. Moreover, lines commonly appear at the occluding boundaries of a scene, and can be tracked reliably as the scene changes, a case usually fatal for point-feature tracking methods.

accuracy of the roll and pitch estimates (already observable when detecting points) but provide no information about the global heading. In contrast, we are interested in examining the information available to a VINS when detecting points, as well as lines aligned with *any* of the cardinal directions. In particular, the main contributions of this paper are:

- We study the observability of a VINS that uses measurements of both points and lines and prove that the observation of a *single* point and a *single line of known direction different than gravity*, provide sufficient information for rendering *all d.o.f. of a VINS observable, up to global translations.*
- We improve the consistency of an extended Kalman filter (EKF) that processes visual measurements of lines of known direction, by ensuring that no information is acquired along the unobservable directions of its linearized system model.
- We provide a simple framework for incorporating measurements of lines aligned with the cardinal directions into existing VINS and experimentally demonstrate the superior accuracy of the proposed system.

The remainder of this paper is structured as follows: In Sect. II, we present the VINS system model as well as the measurement model corresponding to lines of known direction. Sect. III studies the observability of a VINS employing measurements of point features and lines of known direction. Sect. IV introduces a method for improving the consistency of the EKF when processing measurements to lines of known direction, while Sect. V discusses the major components of the proposed framework for exploiting urban regularities in VINS. Sect. VI, presents the results of the experimental comparison of the proposed approach against alternative vision-inertial methods. Finally, Sect. VII summarizes the key findings of this work and provides an outline of future research directions.

II. VINS State and Measurement Models

In this section, we present the system model used for state and covariance propagation based on inertial measurements (Sect. II-A), and then describe the measurement model for processing straight-line observations of known direction (Sect. II-B).

A. IMU State and Covariance Propagation Model

The 19×1 IMU state vector is:

$$\mathbf{x}_R = \begin{bmatrix} {}^I\bar{q}_G^T & \mathbf{b}_g^T & {}^G\mathbf{v}_I^T & \mathbf{b}_a^T & {}^G\mathbf{p}_I^T & {}^G\mathbf{f}^T \end{bmatrix}^T \tag{1}$$

where ${}^I\bar{q}_G(t)$ is the quaternion which represents the orientation of the global frame $\{G\}$ in the IMU frame $\{I\}$. The position and velocity of the IMU frame in the global frame are denoted by ${}^G\mathbf{p}_I(t)$, and ${}^G\mathbf{v}_I(t)$, and $\mathbf{b}_g(t)$, $\mathbf{b}_a(t)$ are the gyroscope and accelerometer biases, while ${}^G\mathbf{f}$ is a mapped feature.

The system model is (see [18]):

$${}^I\dot{\bar{q}}_G(t) = \tfrac{1}{2}\mathbf{\Omega}({}^I\omega(t))\,{}^I\bar{q}_G(t), {}^G\dot{\mathbf{p}}_I(t) = {}^G\mathbf{v}_I(t), \tag{2}$$

$${}^G\dot{\mathbf{v}}_I(t) = {}^G\mathbf{a}(t), \dot{\mathbf{b}}_g(t) = \mathbf{n}_{wg}(t), \dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa}(t), {}^G\dot{\mathbf{f}}(t) = 0 \tag{3}$$

where ${}^I\omega$ and ${}^G\mathbf{a}$ are the rotational velocity and linear acceleration, respectively, while $\mathbf{n}_{wg}$ and $\mathbf{n}_{wa}(t)$ are the white-noise processes driving the IMU biases, and

$$\mathbf{\Omega}(\boldsymbol{\omega}) \triangleq \begin{bmatrix} -\lfloor \boldsymbol{\omega} \times \rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix}, \quad \lfloor \boldsymbol{\omega} \times \rfloor \triangleq \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}.$$

The gyroscope and accelerometer measurements are:

$$\boldsymbol{\omega}_m(t) = {}^I\omega(t) + \mathbf{b}_g(t) + \mathbf{n}_g(t) \tag{4}$$

$$\mathbf{a}_m(t) = \mathbf{C}({}^I\bar{q}_G(t))\left({}^G\mathbf{a}(t) - {}^G\mathbf{g}\right) + \mathbf{b}_a(t) + \mathbf{n}_a(t). \tag{5}$$

where $\mathbf{C}(\bar{q})$ is the rotation matrix corresponding to the quaternion $\bar{q}$, ${}^G\mathbf{g}$ is the gravitational acceleration expressed in $\{G\}$, and $\mathbf{n}_g(t)$ and $\mathbf{n}_a(t)$ are white-noise processes contaminating the corresponding measurements.

Following existing literature on inertial navigation [15], we obtain the continuous-time linearized system model. By defining the 18×1 error-state vector as:[2]

$$\widetilde{\mathbf{x}} = \begin{bmatrix} {}^I\boldsymbol{\delta\theta}_G^T & \widetilde{\mathbf{b}}_g^T & {}^G\widetilde{\mathbf{v}}_I^T & \widetilde{\mathbf{b}}_a^T & {}^G\widetilde{\mathbf{p}}_I^T & {}^G\widetilde{\mathbf{f}}^T \end{bmatrix}^T, \tag{6}$$

the continuous-time IMU error-state equation becomes:

$$\dot{\widetilde{\mathbf{x}}}(t) = \begin{bmatrix} \mathbf{F}_c(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \widetilde{\mathbf{x}}(t) + \mathbf{G}(t)\mathbf{n}(t) \tag{7}$$

$$= \mathbf{F}(t)\,\widetilde{\mathbf{x}}(t) + \mathbf{G}(t)\mathbf{n}(t) \tag{8}$$

where $\mathbf{G}(t)$ is the input noise matrix [5], and $\mathbf{F}_c$ is the error-state transition matrix:

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

with $\hat{\mathbf{a}}(t) = \mathbf{a}_m(t) - \hat{\mathbf{b}}_a(t)$, and $\hat{\boldsymbol{\omega}}(t) = \boldsymbol{\omega}_m(t) - \hat{\mathbf{b}}_g(t)$.

The discrete-time state transition matrix from time t_1 to t_k, $\mathbf{\Phi}_{k,1}$, is the solution to the matrix differential equation $\dot{\mathbf{\Phi}}_{k,1} = \mathbf{F}(t_k)\,\mathbf{\Phi}_{k,1}, \ \mathbf{\Phi}_{1,1} = \mathbf{I}_{18}$ and has the following structure [4]:

$$\mathbf{\Phi}_{k,1} = \begin{bmatrix} \mathbf{\Phi}_{k,1}^{(1,1)} & \mathbf{\Phi}_{k,1}^{(1,2)} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{\Phi}_{k,1}^{(3,1)} & \mathbf{\Phi}_{k,1}^{(3,2)} & \mathbf{I}_3 & \mathbf{\Phi}_{k,1}^{(3,4)} & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{\Phi}_{k,1}^{(5,1)} & \mathbf{\Phi}_{k,1}^{(5,2)} & \delta t_k \mathbf{I}_3 & \mathbf{\Phi}_{k,1}^{(5,4)} & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \tag{9}$$

where $\delta t_k = (t_k - t_1)$. The specific elements of $\mathbf{\Phi}_{k,1}$, which we will employ later on, for the observability analysis are given by [4]:

$$\mathbf{\Phi}_{k,1}^{(1,1)} = \mathbf{C}\left({}^{I_k}\bar{q}_{I_1}\right) \tag{10}$$

$$\mathbf{\Phi}_{k,1}^{(1,2)} = -\int_{t_1}^{t_k} \mathbf{C}\left({}^{I_k}\bar{q}_{I_\tau}\right) \mathrm{d}\tau. \tag{11}$$

[2] For the IMU position, velocity, and biases, we use a standard additive error model (i.e., $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ is the error in the estimate $\hat{\mathbf{x}}$ of a random variable $\mathbf{x}$). To ensure minimal representation for the covariance, we employ a multiplicative attitude error model where the error between the quaternion $\bar{q}$ and its estimate $\hat{\bar{q}}$ is the 3×1 angle-error vector, $\boldsymbol{\delta\theta}$, implicitly defined by the *error quaternion* $\delta\bar{q} = \bar{q} \otimes \hat{\bar{q}}^{-1} \simeq \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}^T & 1 \end{bmatrix}^T$, where $\delta\bar{q}$ describes the small rotation that causes the true and estimated attitude to coincide.

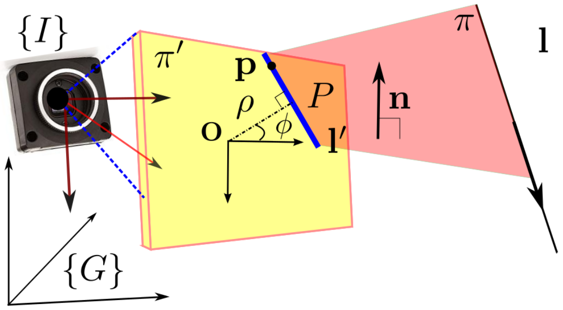


Fig. 1: The geometric constraint employed for deriving an inferred measurement model in (17).

B. Measurement Model

Due to space limitations, we omit the description of the measurement model and the corresponding observability matrix for point features, since this has been investigated thoroughly in existing literature (see [15, 8] and references therein). Instead we hereafter focus on the measurement model, corresponding to observations of lines of known direction.

For simplicity, we assume that the IMU frame of reference $\{I\}$ coincides with the camera frame of reference[3]. The optical center of the camera $\{I\}$, together with the 3D line $\mathbf{l}$, define a plane π in space. Let O denote the principal point of the image plane π'. The image sensor captures the intersection line $\mathbf{l}'$ of the plane π and π', and parameterizes it by an angle ϕ and distance ρ (see Fig. 1).

A point $\mathbf{p}$ with homogeneous image coordinates $\mathbf{p}^T = \begin{bmatrix} u & v & 1 \end{bmatrix}$, lies on the line $\mathbf{l}'$ if it satisfies the equality:

$$\begin{bmatrix} \cos\phi & \sin\phi & -\rho \end{bmatrix} \mathbf{p} = 0. \tag{12}$$

Let ${}^{I}\mathbf{u} = \begin{bmatrix} \sin\phi & -\cos\phi & 0 \end{bmatrix}^T$ be a (free) unit vector along the line $\mathbf{l}'$ on the image plane and let P denote the point on $\mathbf{l}'$ that is closest to O. From Fig. 1, the vectors ${}^{I}\mathbf{u}$ and $IP = IO + OP = \begin{bmatrix} \rho\cos\phi & \rho\sin\phi & 1 \end{bmatrix}^T$ define the plane π. The unit vector ${}^{I}\mathbf{n}$, perpendicular to the plane π, is:

$${}^{I}\mathbf{n} = \frac{IP \times {}^{I}\mathbf{u}}{||IP \times {}^{I}\mathbf{u}||_2} = \frac{1}{\sqrt{1+\rho^2}} \begin{bmatrix} \cos\phi & \sin\phi & -\rho \end{bmatrix}^T. \tag{13}$$

So as to couple the VINS state with the *known* line direction ${}^{G}\mathbf{l}$, we formulate the following geometric constraint:

$$h(\mathbf{x}, {}^{I}\mathbf{n}) = {}^{I}\mathbf{n}^T \mathbf{C}({}^{I}\bar{q}_G){}^{G}\mathbf{l} = 0 \tag{14}$$

which captures the fact that the line direction ${}^{G}\mathbf{l}$, expressed in the IMU frame $\{I\}$, lies in the plane π, and hence is perpendicular to the normal ${}^{I}\mathbf{n}$. In practice, the camera measures[4]

$$\mathbf{z} = \begin{bmatrix} \phi & \rho \end{bmatrix}^T + \xi \tag{15}$$

where ξ follows a zero-mean Gaussian distribution $\mathcal{N}(\mathbf{0}_{2\times 1}, \mathbf{R}_{\phi\rho})$ and represents the noise induced by the camera sensor and the line extraction algorithm. The effect of ξ on $h(\mathbf{x}, {}^{I}\mathbf{n})$, denoted by w can be approximated through linearization as:

$$w \approx \mathbf{A}_{\mathbf{1}\times\mathbf{3}}\mathbf{B}_{\mathbf{3}\times\mathbf{2}}\xi \tag{16}$$

where $\mathbf{A} = \nabla_{{}^{I}\mathbf{n}} h$ and $\mathbf{B} = \begin{bmatrix} \nabla_{\phi}{}^{I}\mathbf{n} & \nabla_{\rho}{}^{I}\mathbf{n} \end{bmatrix}$. Hence, w can be approximated by a zero-mean Gaussian distribution $\mathcal{N}(0, \sigma^2),\ \sigma^2 = \mathbf{A}\mathbf{B}\mathbf{R}_{\phi\rho}\mathbf{B}^T\mathbf{A}^T$.

Using the above noise parameterization, we arrive at the following inferred measurement model that couples the measurement of line $\mathbf{l}$ at time-step t_k, with the VINS state vector, $\mathbf{x}_k$:

$$z_{l,t_k} = h(\mathbf{x}_k, {}^{I_k}\mathbf{n}) + w_k = {}^{I_k}\mathbf{n}^T \mathbf{C}({}^{I_k}\bar{q}_G){}^{G}\mathbf{l} + w_k. \tag{17}$$

The measurement Jacobian of line $\mathbf{l}$, at time-step t_k, takes the form:

$$\mathbf{H}_k^l = \begin{bmatrix} {}^{I_k}\mathbf{n}^T \lfloor \mathbf{C}({}^{I_k}\bar{q}_G){}^{G}\mathbf{l} \times \rfloor & \mathbf{0}_{1\times 15} \end{bmatrix}. \tag{18}$$

III. Observability Analysis

In this section, we study the observability properties of a VINS observing a single point feature and a line of known direction, over k time steps. The observability matrix $\mathbf{M}$, of the linearized VINS, can be partitioned in two sub-matrices (see (19)), corresponding to measurements of a point feature and a single line of known direction, respectively:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}^{\mathbf{f}} \\ \hline \mathbf{M}^{\mathbf{l}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1^{\mathbf{f}} \\ \mathbf{H}_2^{\mathbf{f}}\mathbf{\Phi}_{2,1} \\ \vdots \\ \mathbf{H}_k^{\mathbf{f}}\mathbf{\Phi}_{k,1} \\ \hline \mathbf{H}_1^{\mathbf{l}} \\ \mathbf{H}_2^{\mathbf{l}}\mathbf{\Phi}_{2,1} \\ \vdots \\ \mathbf{H}_k^{\mathbf{l}}\mathbf{\Phi}_{k,1} \end{bmatrix}. \tag{19}$$

As it has been recently shown [5], the right nullspace of $\mathbf{M}^{\mathbf{f}}$, is spanned by:

$$\mathbf{N}^{\mathbf{f}} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{C}\left({}^{I_1}\bar{q}_G\right){}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_3 & -\lfloor {}^{G}\mathbf{v}_{I_1} \times \rfloor {}^{G}\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor {}^{G}\mathbf{p}_{I_1} \times \rfloor {}^{G}\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^{G}\mathbf{f} \times \rfloor {}^{G}\mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{N}_{t,1} & | & \mathbf{N}_{r,1} \end{bmatrix}. \tag{20}$$

Notice that the three directions $\mathbf{N}_{t,1}$ correspond to global translations of the platform and the observed feature, while $\mathbf{N}_{r,1}$, describes rotations around the gravitational acceleration vector, ${}^{G}\mathbf{g}$. As we show in Appendix A, the right nullspace of $\mathbf{M}^{\mathbf{l}}$, is given by:

$$\mathbf{N}^{\mathbf{l}} = \begin{bmatrix} \mathbf{C}\left({}^{I_1}\bar{q}_G\right){}^{G}\mathbf{l} & \mathbf{0}_{3\times 12} \\ \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 12} \\ \mathbf{0}_{12\times 1} & \mathbf{I}_{12\times 12} \end{bmatrix} = \begin{bmatrix} \mathbf{N}^{\mathbf{l},(:,1)} & \mathbf{N}^{\mathbf{l},(:,2:13)} \end{bmatrix}. \tag{21}$$

[3] In practice, we perform IMU-camera extrinsic calibration following the approach of [12].

[4] We estimate ϕ and ρ, and their associated covariances by fitting a straight line to points that belong to an image edge.

Hence, the right nullspace $\mathbf{M}$ is:

$$\text{null}(\mathbf{M}) = \text{range}(\mathbf{N}^{\mathbf{f}}) \cap \text{range}(\mathbf{N}^{\mathbf{l}}). \tag{22}$$

The subspace $\text{range}(\mathbf{N}^{\mathbf{f}}) \cap \text{range}(\mathbf{N}^{\mathbf{l}})$ can be parameterized as:

$$\mathbf{N}^{\mathbf{f}}\mathbf{N}_{\alpha} = \mathbf{N}^{\mathbf{l}}\mathbf{N}_{\beta} \tag{23}$$

for full column rank matrices $\mathbf{N}_{\alpha},\ \mathbf{N}_{\beta}$. From (23), we get:

$$\begin{bmatrix}\mathbf{N}^{\mathbf{f}} & \mid & -\mathbf{N}^{\mathbf{l}}\end{bmatrix} \underbrace{\begin{bmatrix}\mathbf{N}_{\alpha} \\ \mathbf{N}_{\beta}\end{bmatrix}}_{\boldsymbol{\Delta}} = \mathbf{0}. \tag{24}$$

Hence $\boldsymbol{\Delta}$ is the right nullspace of $\begin{bmatrix}\mathbf{N}^{\mathbf{f}} & \mid & -\mathbf{N}^{\mathbf{l}}\end{bmatrix}$, which is:

$$\boldsymbol{\Delta} = \begin{cases} \begin{bmatrix} \mathbf{I}_3 \\ \mathbf{0}_{1\times 3} \\ \mathbf{0}_{1\times 3} \\ \mathbf{0}_3 \\ \mathbf{0}_3 \\ \mathbf{I}_3 \\ \mathbf{I}_3 \end{bmatrix} & \text{, if } {}^G\mathbf{l} \not\parallel {}^G\mathbf{g} \\ \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 1 \\ \mathbf{0}_{1\times 3} & \gamma \\ \mathbf{0}_3 & -\lfloor {}^G\mathbf{v}_{I_1} \times \rfloor {}^G\mathbf{g} \\ \mathbf{0}_3 & \mathbf{0}_{3\times 1} \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{p}_{I_1} \times \rfloor {}^G\mathbf{g} \\ \mathbf{I}_3 & -\lfloor {}^G\mathbf{f} \times \rfloor {}^G\mathbf{g} \end{bmatrix} & \text{, otherwise.} \end{cases}$$

where γ is a scalar such that ${}^G\mathbf{g} = \gamma {}^G\mathbf{l}$. After recovering $\mathbf{N}_{\alpha}$, we get the span of $\text{range}(\mathbf{N}^{\mathbf{f}}) \cap \text{range}(\mathbf{N}^{\mathbf{l}})$ as:

$$\text{null}(\mathbf{M}) = \mathbf{N}^{\mathbf{f}}\mathbf{N}_{\alpha} = \begin{cases} \begin{bmatrix}\mathbf{N}_{t,1}\end{bmatrix} & \text{if } {}^G\mathbf{l} \not\parallel {}^G\mathbf{g} \\ \begin{bmatrix}\mathbf{N}_{t,1} & \mid & \mathbf{N}_{r,1}\end{bmatrix} & \text{otherwise.} \end{cases}$$

In summary, we have shown that observations of a *single* point feature and a *single 3D line of known direction, different from gravity*, provide enough information, for rendering *all degrees of freedom of a VINS observable, up to the global translations* of the observed feature and the IMU-camera platform.

IV. Consistent EKF Updates for Lines of Known Direction

As defined in [1], a state estimator is consistent if the estimation errors are zero-mean and have covariance smaller than or equal to the one calculated by the filter. Unfortunately, processing observations of lines of known direction using an EKF can lead to injection of spurious information and hence inconsistencies.

As we show in Appendix A, for the true linearized system, the direction corresponding to rotations around $\mathbf{l}$ is unobservable and the observability matrix $\mathbf{M}^{\mathbf{l}}$ is of rank 5. However, when the system and measurement Jacobians are evaluated at the current state estimates, the corresponding observability matrix $\hat{\mathbf{M}}^{\mathbf{l}}$ becomes of rank 6. This causes the EKF to erroneously perceive the direction corresponding to rotations around the known line $\mathbf{l}$ as observable and decrease its uncertainty, causing estimator inconsistency. We hereafter examine this problem in more detail and propose a simple, yet powerful, method for addressing it.

Consider a system observing a *single* line of known direction $\mathbf{l}$ over k time-steps, $t_1, \ldots, t_k$. For the linearized system, employed by the EKF the first row of the observability matrix $\hat{\mathbf{M}}^{\mathbf{l}}$, takes the form:[5]

$$\hat{\mathbf{M}}^{\mathbf{l}}_1 = \begin{bmatrix} {}^{I_1}\mathbf{n}^T \lfloor \mathbf{C}\left({}^{I_{1|0}}\bar{q}_G\right) {}^G\mathbf{l} \times \rfloor & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 12} \end{bmatrix} \tag{25}$$

while its $k-th$ row has the form:

$$\hat{\mathbf{M}}^{\mathbf{l}}_k = \begin{bmatrix} {}^{I_k}\mathbf{n}^T \lfloor \mathbf{C}\left({}^{I_{k|k-1}}\bar{q}_G\right) {}^G\mathbf{l} \times \rfloor \hat{\boldsymbol{\Gamma}}_k & \mathbf{E}_k & \mathbf{0}_{1\times 12} \end{bmatrix}. \tag{26}$$

where:

$$\begin{aligned} \hat{\boldsymbol{\Gamma}}_k &= \boldsymbol{\Phi}^{(1,1)}_{k|k-1,k-1|k-1} \cdots \boldsymbol{\Phi}^{(1,1)}_{2|1,1|1} \\ &= \mathbf{C}\left({}^{I_{k|k-1}}\bar{q}_{I_{k-1|k-1}}\right) \ldots \mathbf{C}\left({}^{I_{2|1}}\bar{q}_{I_{1|1}}\right) \end{aligned} \tag{27}$$

and $\mathbf{E}_k$ is a time-varying full-rank matrix that does not affect the present analysis.

From the structure of the observability matrix $\hat{\mathbf{M}}^{\mathbf{l}}$, the directions $\mathbf{N}^{\mathbf{l},(:,2:13)}$ (see (21)) are independent of the current state estimate and satisfy $\hat{\mathbf{M}}^{\mathbf{l}}_k \mathbf{N}^{\mathbf{l},(:,2:13)} = 0$, for any k. However, the same is not true for the direction $\mathbf{N}^{\mathbf{l},(:,1)}$ which corresponds to rotations around the line $\mathbf{l}$.

Specifically, at time-step t_1, the direction $\mathbf{N}^{\mathbf{l},(:,1)}$ is given by:

$$\hat{\mathbf{N}}^{\mathbf{l},(:,1)}_{1|0} = \begin{bmatrix} \mathbf{C}\left({}^{I_{1|0}}\bar{q}_G\right) {}^G\mathbf{l} \\ \mathbf{0}_{3\times 1} \\ \mathbf{0}_{12\times 1} \end{bmatrix}. \tag{28}$$

For $\hat{\mathbf{M}}^{\mathbf{l}}_k \hat{\mathbf{N}}^{\mathbf{l},(:,1)}_{1|0} = 0$ to hold, we would need:

$${}^{I_k}\mathbf{n}^T \lfloor \mathbf{C}\left({}^{I_{k|k-1}}\bar{q}_G\right) {}^G\mathbf{l} \times \rfloor \hat{\boldsymbol{\Gamma}}_k \mathbf{C}\left({}^{I_{1|0}}\bar{q}_G\right) {}^G\mathbf{l} = 0, \ \forall\ k \tag{29}$$

which in general does not hold, for the $\hat{\boldsymbol{\Gamma}}_k$ in (27) and leads to injection of spurious information. To address this problem, we seek a modified $\boldsymbol{\Gamma}^{\star}_k$ such that:

$$\boldsymbol{\Gamma}^{\star}_k \mathbf{C}\left({}^{I_{1|0}}\bar{q}_G\right) {}^G\mathbf{l} = \mathbf{C}\left({}^{I_{k|k-1}}\bar{q}_G\right) {}^G\mathbf{l}. \tag{30}$$

One possible solution, for (30), is to evaluate $\boldsymbol{\Phi}^{(1,1)}$, using the propagated orientation estimates, i.e.,

$$\boldsymbol{\Phi}^{\star(1,1)}_{k|k-1,k-1|k-1} = \mathbf{C}\left({}^{I_{k|k-1}}\bar{q}_{I_{k-1|k-2}}\right). \tag{31}$$

In that case, the $k-th$ row of the observability matrix would include a matrix $\boldsymbol{\Gamma}^{\star}_k$ such that:

$$\boldsymbol{\Gamma}^{\star}_k = \mathbf{C}\left({}^{I_{k|k-1}}\bar{q}_{I_{k-1|k-2}}\right) \ldots \mathbf{C}\left({}^{I_{2|1}}\bar{q}_{I_{1|0}}\right) \tag{32}$$

which satisfies (30).

V. VINS in an Urban Scene aligned with Gravity

We make the assumption that we navigate in an environment whose lines are predominantly mutually orthogonal and aligned with the gravitational field (i.e., parallel or perpendicular to the gravity vector). After initializing the roll and pitch angles w.r.t. a global frame $\{G\}$, whose z-axis is parallel to gravity, finding the initial yaw is achieved through the process described hereafter.

[5] $I_{k|k-1}$ denotes the estimate for frame $\{I\}$ at time-step k, utilizing all available inertial measurements and line observations only up to time-step $k-1$.

A. Yaw Initialization

Firstly, we employ a RANSAC-based vanishing point estimator that uses triplets of lines for generating hypotheses of all three orthogonal vanishing points at once [13]. We prune these hypotheses, by keeping the one that corresponds to a rotational matrix with roll and pitch angles closest to those estimated by the filter. Following this method, we define a frame $\{B\}$ (corresponding to the building) whose cardinal axes are aligned with the structural lines of the scene. Next, we estimate the yaw angle, θ, between frames $\{G\}$ and $\{B\}$. As an example, consider N observations, $\mathbf{n}_i$, $i = 1 \dots N$, of lines parallel to the direction $\mathbf{e}_1 = \begin{bmatrix}1 & 0 & 0\end{bmatrix}^T$ of the building. From these measurements, we arrive at the following least squares problem:

$$\min_{\theta} \; J(\theta) \tag{33}$$

$$J(\theta) = \sum_{i=1}^{N} \left(\mathbf{n}_i^T \mathbf{C}({}^I\bar{q}_G) \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{e}_1 \right)^2$$

After defining,

$$\bar{\mathbf{u}} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}, \text{ and } \Pi = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \tag{34}$$

we can re-arrange $J(\theta)$ as,

$$\begin{aligned} J(\bar{\mathbf{u}}) &= \sum_{i=1}^{N} \left(\mathbf{n}_i^T \mathbf{C}({}^I\bar{q}_G) \Pi \bar{\mathbf{u}} \right)^2 \\ &= \bar{\mathbf{u}}^T \underbrace{\Pi^T \mathbf{C}({}^I\bar{q}_G)^T \sum_{i=1}^{N} \left(\mathbf{n}_i \mathbf{n}_i^T \right) \mathbf{C}({}^I\bar{q}_G) \Pi}_{\mathbf{S}} \bar{\mathbf{u}} \end{aligned} \tag{35}$$

When $\bar{\mathbf{u}}^T \bar{\mathbf{u}} = 1$, (35) corresponds to the Rayleigh quotient of the matrix $\mathbf{S}$ and is minimized for a $\bar{\mathbf{u}}^*$ equal to the eigenvector corresponding to the minimum eigenvalue of $\mathbf{S}$. Hence, we recover θ, as:

$$\begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} = \pm \bar{\mathbf{u}}^* \tag{36}$$

Note, that this procedure takes place once, at the initialization of the filter, so as to disambiguate the cardinal directions of the building. Hereafter, we assume that the cardinal axes of the frame $\{G\}$ are parallel with the cardinal axes of the building.

B. Line Classification

After we perform gradient edge detection using the Canny Edge detector [2] and extract straight lines using OpenCV's probabilistic Hough transform [11], we need to classify whether a line is parallel to one of the three cardinal directions, $\mathbf{e}_j$, $j = 1 \dots 3$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \; \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{37}$$

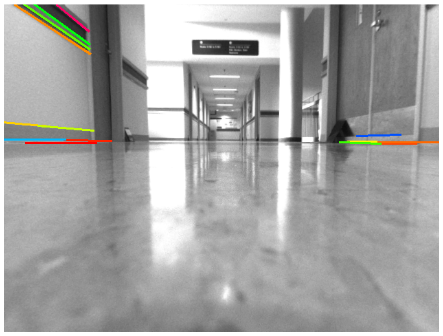


Fig. 2: Lines corresponding to the building's x cardinal axis, classified using a RANSAC-based vanishing point estimator, for performing heading (yaw) initialization.

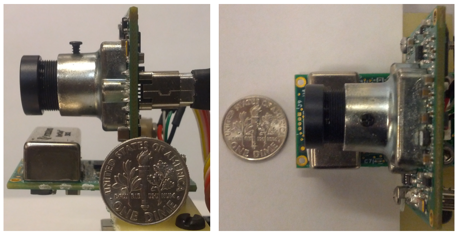


Fig. 5: The hardware setup comprises of a miniature monochrome Point Grey Chameleon camera recording images at 7.5 Hz, and a rigidly attached InterSense NavChip IMU operating at 100 Hz. A coin (US dime, radius 1.8 cm) is included as a size reference.

For the proposed VINS, this classification problem can be solved *reliably*, since *absolute* roll, pitch, and yaw angles are *observable* (see Sect. III). We achieve this, by taking advantage of the current filter's state estimate $\mathbf{x}_{k|k-1}$ and covariance $\mathbf{P}_{k|k-1}$, as described in Algorithm 1.

Algorithm 1: Line Classification using state and covariance estimates at time-step t_k, over N measurements.

```
for Line measurement $i = 1 \dots N$ do
    for Cardinal Direction $j = 1 \dots 3$ do
        Compute residual $r_i^j = -{}^{I,k}\mathbf{n}_i^T \mathbf{C}({}^{I_k}\bar{q}_G)^G\mathbf{e}_j$;
        Compute measurement Jacobian $\mathbf{H}_k^{\mathbf{e}_j}$ using (18);
        Perform 1 d.o.f. $\chi^2$ Mahalanobis distance test:;
        $\epsilon_{i,j} = \frac{(r_i^j)^2}{\left(\mathbf{H}_k^{\mathbf{e}_j}\mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathbf{e}_j\,T} + \sigma_i^2\right)}$;
    end
    Keep measurement $i$ if it succeeded only a single
    hypothesis test;
end
```

VI. Experimental Results

Our experimental setup comprises a PointGrey Chameleon camera[6] and a Navchip IMU[7], which are rigidly attached on a light-weight (100g) platform (see Fig. 5). IMU signals

[6] http://www.ptgrey.com

[7] http://www.intersense.com

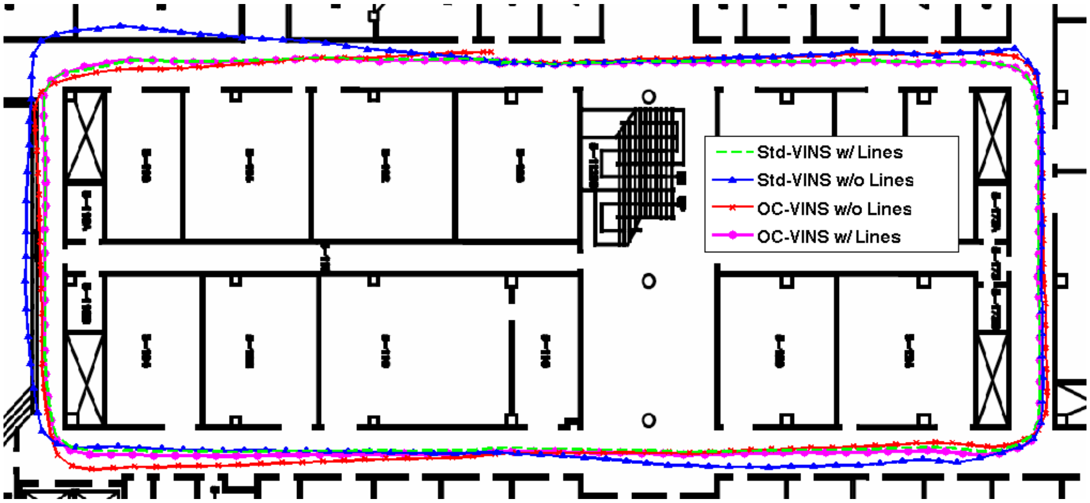
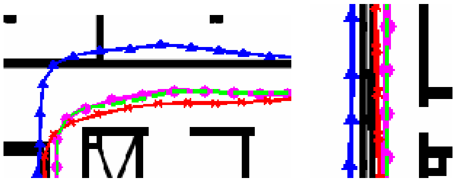


Fig. 3: Overhead x-y view of the 3-D trajectory, projected on the building's floor plans. The Std-VINS (–△–) violates the correct observability properties and hence results in an infeasible trajectory, (i.e., passing through walls), the OC-VINS (–×–) remains consistent with the ground truth floor drawings, however it *cannot recover* from its accumulated yaw errors. In contrast, the OC-VINS and Std-VINS augmented with observations of *lines of known directions* (–○– and ——), are able to both correct their yaw (heading) error and remain consistent with the floor plans.



Visual Inertial Odometry	Final Error (m)	Pct. (%)
Std-VINS	0.57	0.40
OC-VINS	0.80	0.56
Std-VINS w/ Lines	0.47	0.33
OC-VINS w/ Lines	0.31	0.22

Fig. 4 & TABLE I: Zoom-in to narrow paths of the trajectory (Left and Middle). The width of the narrow corridor (at the left) is approximately 1.5 m, emphasizing the fact that the filters which do not employ line measurements result in infeasible trajectories. (Right) Final position error reduction, when visual observations of lines are employed. Note that the proposed method (OC-VINS w/ Lines) achieves approximately 50% final position error reduction compared to Std-VINS.

were sampled at a frequency of 100 Hz while camera images were acquired at 7.5 Hz. The experiment was conducted in a corridor scene dominated by orthogonal lines, aligned with or perpendicular to gravity, and for which the floor plans are available but used only for assessing accuracy.

In our experiments we compared the performance of the following VINS algorithms:

- The Multi-state constraint Kalman filter (MSC-KF) of [14] using only observations of points (referred to as Std-VINS w/o lines in Fig. 3). The main advantage of the MSC-KF is that it processes all geometric constraints induced by camera measurements of points features, over a finite window of image frames, with computational complexity only *linear* in the number of observed features. This is accomplished by not including a map of the environment in the state vector, but rather using all available information for motion estimation.
- The MSC-KF modified to also process visual observations of lines of known direction (referred to as Std-VINS w/ lines in Fig. 3).
- The Observability-Constraint OC-MSC-KF of [8] using only observations of points (referred to as OC-VINS w/o lines in Fig. 3). As explained in [8], due to estimation errors, the linearized error-state model of the MSC-KF erroneously perceives rotations about the gravity as observable, thus leading to inconsistency. The OC-MSC-KF addresses this problem by appropriately modifying the corresponding Jacobians and ensuring that no information from the measurements is injected along the direction of rotations about gravity.
- The proposed OC-MSC-KF where observations of points as well as of lines of known directions are used (referred to as OC-VINS w/ lines in Fig. 3). The main difference with the OC-MSC-KF of [8] is that the system Jacobians are appropriately modified (as explained earlier) to ensure that no information about rotations around the detected line of known direction is acquired by the filter.

The results described hereafter are from an experiment where a person holding the IMU-camera pair traversed a path of 144 m and returned back to his initial position for providing an estimate of the final position error.

As shown in Fig. 3, the Std-VINS w/o lines erroneously perceives the global attitude as observable and results in infeasible trajectories that violate the building's floor plan (see

Fig. 3). In contrast, the OC-VINS w/ lines (see Sect. V) is able to recover from heading errors and also remain *consistent* with the ground-truth floor plans. As depicted in Table I, the proposed method achieves approximately 45% position accuracy improvement when compared to the Std-VINS w/o lines. Furthermore, the superior performance of OC-VINS w/ lines when compared to the Std-VINS w/ lines (approximately 34%), underlines the negative impact of a mismatch between the observability properties of the true system and the one employed for estimation purposes, as discussed in Sect. IV.

VII. Conclusion

In this work, we presented an extended Kalman filter-based algorithm for processing observations of structural lines, so as to improve the accuracy of vision-aided inertial navigation systems (VINS), when operating in man-made environments. Our observability analysis showed that measurements, across time, of a *single* point feature and a *single 3D line of known direction different than gravity*, provide sufficient information, for rendering *all degrees of freedom of a VINS observable*, up to global translations of the sensor platform. Leveraging the results of our observability analysis, we introduced a method for improving the consistency of an EKF estimator that processes measurements of such lines, by ensuring that no spurious information (i.e., rotations around the observed line) is acquired by the estimator. Finally, we developed a simple framework for incorporating camera observations of structural lines into existing VINS algorithms and demonstrated, through experimental validation, the superior accuracy of the resulting VINS. In our future work, we aim at the improvement of VINS' accuracy, by further taking advantage of structural constraints present in urban environments, such as the existence of mutually orthogonal or parallel planes.

Appendix A

In this section, we study the observability properties of a VINS, observing a *single* line ${}^{G}\mathbf{l}$ of known direction. The observability matrix $\mathbf{M}^{\mathbf{l}}$, corresponding to the linearized, around the *true* state estimates, system is given by:

$$\mathbf{M}^{\mathbf{l}} = \begin{bmatrix} \mathbf{H}^{\mathbf{l}}_1 \\ \mathbf{H}^{\mathbf{l}}_2 \boldsymbol{\Phi}_{2,1} \\ \vdots \\ \mathbf{H}^{\mathbf{l}}_k \boldsymbol{\Phi}_{k,1} \end{bmatrix} = \begin{bmatrix} \mathbf{M}^{\mathbf{l},(\mathbf{1:6})} & \mathbf{0}_{k\times 12} \end{bmatrix}. \tag{38}$$

In what follows, we will prove that $\mathbf{M}^{\mathbf{l}}$ has (for generic rotations of the sensor platform) rank 5. Substituting the analytic expressions for $\boldsymbol{\Phi}_{k,1}$ (See (9) and (10) in Sect. II) and $\mathbf{H}^{\mathbf{l}}_k$ (See (18) in Sect. II-B), we arrive at:

$$\mathbf{M}^{\mathbf{l},(\mathbf{1:6})} = \begin{bmatrix} \mathbf{M}^{\mathbf{l},(\mathbf{1:3})} & \mathbf{M}^{\mathbf{l},(\mathbf{4:6})} \end{bmatrix} \tag{39}$$

$$\mathbf{M}^{\mathbf{l},(\mathbf{1:3})} = \begin{bmatrix} {}^{I_1}\mathbf{n}^T \lfloor {}^{I_1}\mathbf{l} \times \rfloor \\ {}^{I_2}\mathbf{n}^T \lfloor {}^{I_2}\mathbf{l} \times \rfloor \mathbf{C}\left({}^{I_2}\bar{q}_{I_1}\right) \\ \vdots \\ {}^{I_k}\mathbf{n}^T \lfloor {}^{I_k}\mathbf{l} \times \rfloor \mathbf{C}\left({}^{I_k}\bar{q}_{I_1}\right) \end{bmatrix} \tag{40}$$

$$\mathbf{M}^{\mathbf{l},(\mathbf{4:6})} = \begin{bmatrix} \mathbf{0}_{1\times 3} \\ -{}^{I_2}\mathbf{n}^T \lfloor {}^{I_2}\mathbf{l} \times \rfloor \int_{t_1}^{t_2} \mathbf{C}\left({}^{I_2}\bar{q}_{I_\tau}\right) d\tau \\ \vdots \\ -{}^{I_k}\mathbf{n}^T \lfloor {}^{I_k}\mathbf{l} \times \rfloor \int_{t_1}^{t_k} \mathbf{C}\left({}^{I_k}\bar{q}_{I_\tau}\right) d\tau \end{bmatrix}. \tag{41}$$

Since the last 12 columns of $\mathbf{M}^{\mathbf{l}}$ are always zero, we can drop them and study the right nullspace corresponding to its first 6 columns. Let us denote:

$$\breve{\mathbf{n}}_k = \mathbf{C}\left({}^{G}\bar{q}_{I_k}\right) {}^{I_k}\mathbf{n} \tag{42}$$

$$\mathbf{B}_k = \int_{t_1}^{t_k} \mathbf{C}\left({}^{G}\bar{q}_{I_\tau}\right) d\tau. \tag{43}$$

We can write:

$$\mathbf{M}^{\mathbf{l},(\mathbf{1:6})} = \begin{bmatrix} \breve{\mathbf{n}}_1{}^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{C}\left({}^{G}\bar{q}_{I_1}\right) & \mathbf{0}_{1\times 3} \\ \breve{\mathbf{n}}_2{}^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{C}\left({}^{G}\bar{q}_{I_1}\right) & -\breve{\mathbf{n}}_2^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{B}_2 \\ \vdots & \vdots \\ \breve{\mathbf{n}}_k{}^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{C}\left({}^{G}\bar{q}_{I_1}\right) & -\breve{\mathbf{n}}_k^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{B}_k \end{bmatrix}. \tag{44}$$

For finding the nullspace of $\mathbf{M}^{\mathbf{l},(\mathbf{1:6})}$, we seek to determine vectors $\alpha, \beta \in \mathbb{R}^3$, such that:

$$\mathbf{M}^{\mathbf{l},(\mathbf{1:6})} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \mathbf{0} \tag{45}$$

which is a system of k equations in α, β, for any time steps $t_1 \dots t_k$. For the first equation to hold, we can enumerate the possible solutions for α, by $\alpha_{1\dots4}$.

$$\breve{\mathbf{n}}_1{}^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{C}\left({}^{G}\bar{q}_{I_1}\right) \alpha = 0 \implies \tag{46}$$

$$\alpha_1 = \mathbf{C}\left({}^{I_1}\bar{q}_{G}\right) {}^{G}\mathbf{l}, \text{ or} \tag{47}$$

$$\alpha_2 = \mathbf{C}\left({}^{I_1}\bar{q}_{G}\right) \breve{\mathbf{n}}_1, \text{ or} \tag{48}$$

$$\alpha_3 = \mathbf{0}, \text{ or} \tag{49}$$

$$\alpha_4 = \epsilon_1 \alpha_1 + \alpha_2 \tag{50}$$

where ϵ_1 can be any scalar different than 0. For cases 1 and 3, the rest of the $k-1$ equations take the form:

$$\begin{bmatrix} \breve{\mathbf{n}}_2^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{B}_2 \\ \vdots \\ \breve{\mathbf{n}}_k^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{B}_k \end{bmatrix} \beta = \mathbf{0}. \tag{51}$$

Since $\mathbf{B}_k$ is always full-rank for any k, this would require:

$$\exists \beta \in \mathbb{R}^3, \text{ s.t. } \mathbf{B}_k \beta = {}^{G}\mathbf{l}, \ \forall\ k \tag{52}$$

which for generic rotations of the platform, has a solution only for $\beta = \mathbf{0}$. While for cases 2 and 4:

$$\begin{bmatrix} \breve{\mathbf{n}}_2^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{B}_2 \\ \vdots \\ \breve{\mathbf{n}}_k^T \lfloor {}^{G}\mathbf{l} \times \rfloor \mathbf{B}_k \end{bmatrix} \beta = \begin{bmatrix} \breve{\mathbf{n}}_2{}^T \lfloor {}^{G}\mathbf{l} \times \rfloor \breve{\mathbf{n}}_1 \\ \vdots \\ \breve{\mathbf{n}}_k{}^T \lfloor {}^{G}\mathbf{l} \times \rfloor \breve{\mathbf{n}}_1 \end{bmatrix}. \tag{53}$$

Similarly to the previous case, since $\mathbf{B}_k$ is always full-rank for any k, this would require:

$$\exists \beta \in \mathbb{R}^3, \text{s.t. } \mathbf{B}_k \beta = \breve{\mathbf{n}}_1, \forall k \tag{54}$$

which for generic rotations of the platform, has no solution. This concludes our proof that the right nullspace of $\mathbf{M}^{\mathbf{l}}$, is:

$$\mathbf{N}^{\mathbf{l}} = \begin{bmatrix} \mathbf{C}\left({}^{I_1}\bar{q}_G\right) {}^{G}\mathbf{l} & \mathbf{0}_{3\times 12} \\ \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 12} \\ \mathbf{0}_{12\times 1} & \mathbf{I}_{12\times 12} \end{bmatrix}. \tag{55}$$

References

[1] Y. Bar-Shalom, X. Rong Li, and Kirubarajan T. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, New York, NY, 2001.

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[3] H. H. Chen. Pose determination from line-to-plane correspondences: existence condition and closed-form solutions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(6):530–541, June 1991.

[4] J.A. Hesch, D.G. Kottas, S.L. Bowman, and S.I. Roumeliotis. Observability-constrained Vision-aided Inertial Navigation. Technical Report 2012-001, University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, February 2012. URL http://www-users.cs.umn.edu/~dkottas/pdfs/vins_tr_winter_2012.pdf.

[5] J.A. Hesch, D.G. Kottas, S.L. Bowman, and S.I. Roumeliotis. Towards consistent vision-aided inertial navigation. In *Int. Workshop on the Algorithmic Foundations of Robotics*, Cambridge, MA, June 13–15, 2012.

[6] M. Hwangbo and T. Kanade. Visual-Inertial UAV attitude estimation using urban scene regularities. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 2451–2458, Shanghai, China, May 9–13, 2011.

[7] E.S. Jones and S. Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *Int. Journal of Robotics Research*, 30(4):407–430, April 2011.

[8] D.G. Kottas, J.A. Hesch, S.L. Bowman, and S.I. Roumeliotis. On the consistency of vision-aided inertial navigation. In *Int. Symposium on Experimental Robotics*, Quebec City, Canada, June 17–21, 2012.

[9] J. Lobo and J. Dias. Vision and inertial sensor cooperation using gravity as a vertical reference. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(12): 1597–1608, Dec 2003.

[10] A. Martinelli. Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination. *IEEE Trans. on Robotics*, 28(1):44–60, February 2012.

[11] J. Matas, C. Galambos, and J. Kittler. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding*, 78(1):119–137, April 2000.

[12] F.M. Mirzaei and S.I. Roumeliotis. A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. *IEEE Trans. on Robotics*, 24(5):1143–1156, October 2008.

[13] F.M. Mirzaei and S.I. Roumeliotis. Optimal estimation of vanishing points in a manhattan world. In *Proc. of the Int. Conf. on Computer Vision*, pages 2454–2461, Barcelona, Spain, November 6-12 2011. IEEE.

[14] A.I. Mourikis and S.I Roumeliotis. A multi-state constraint Kalman filter for vision-aided inertial navigation. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 3565–3572, Rome, Italy, April 10–14, 2007.

[15] A.I Mourikis, N. Trawny, S.I Roumeliotis, A.E. Johnson, A. Ansar, and L. Matthies. Vision-aided inertial navigation for spacecraft entry, descent, and landing. *IEEE Trans. on Robotics*, 25(2):264–280, April 2009.

[16] H. Rehbinder and B.K. Ghosh. Pose estimation using line-based dynamic vision and inertial sensors. *IEEE Transactions on Automatic Control*, 48(2):186–199, February 2003.

[17] L. Ruotsalainen, J. Bancroft, and G. Lachapelle. Mitigation of attitude and gyro errors through vision aiding. In *Proc. of the IEEE Int. Conf. on Indoor Positioning and Indoor Navigation*, pages 1–9, Sydney, Australia, November 13–15, 2012.

[18] N. Trawny and S. I. Roumeliotis. Indirect Kalman filter for 3D attitude estimation. Technical Report 2005-002, University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, March 2005. URL http://www-users.cs.umn.edu/~trawny/Publications/Quaternions_3D.pdf.

[19] S. Weiss, M.W. Achtelik, S. Lynen, M. Chli, and R. Siegwart. Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environment. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, St. Paul, MN, May 14–18, 2012.

[20] B. Williams, N. Hudson, N. Tweddle, R. Brockers, and L. Matthies. Feature and pose constrained visual aided inertial navigation for computationally constrained aerial vehicles. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pages 431–438, Shanghai, China, May 9–13, 2011.

[21] T.S. Huang Y. Liu and O.D. Faugeras. Determination of camera location from 2D to 3D line and point correspondences. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 82–88, Ann Arbor, MI, June 5–9, 1988.