Preintegrated Velocity Bias Estimation to Overcome Contact Nonlinearities in Legged Robot Odometry

David Wisth, Marco Camurri, Maurice Fallon

Abstract— In this paper, we present a novel factor graph formulation to estimate the pose and velocity of a quadruped robot on slippery and deformable terrain. The factor graph introduces a preintegrated velocity factor that incorporates velocity inputs from leg odometry and also estimates related biases. From our experimentation we have seen that it is difficult to model uncertainties at the contact point such as slip or deforming terrain, as well as leg flexibility. To accommodate for these effects and to minimize leg odometry drift, we extend the robot's state vector with a bias term for this preintegrated velocity factor. The bias term can be accurately estimated thanks to the tight fusion of the preintegrated velocity factor with stereo vision and IMU factors, without which it would be unobservable. The system has been validated on several scenarios that involve dynamic motions of the ANYmal robot on loose rocks, slopes and muddy ground. We demonstrate a 26% improvement of relative pose error compared to our previous work and 52% compared to a state-of-the-art proprioceptive state estimator.

I. INTRODUCTION

The increased maturity of legged robotics has been demonstrated by the initial industrial deployments of quadruped robots, as well as impressive results achieved by academic research. State estimation plays a key role in field deployment of legged machines: without an accurate estimate of its location and velocity, the robot cannot build a useful representation of its environment or plan/execute trajectories to reach goal positions.

Most legged robots are equipped with a high frequency ($>250\,\text{Hz}$) proprioceptive state estimator for control and local mapping purposes. These are typically implemented as nonlinear filters fusing high frequency signals such as kinematics and IMU [1]. In ideal conditions (i.e. high friction, rigid terrain), these estimators have a limited (yet unavoidable) drift that is acceptable for local mapping and control.

However, deformable terrains, leg flexibility and slippage can degrade the estimation performance up to a point where local terrain reconstruction is unusable and multi-step trajectories cannot be executed, even over short ranges. This problem is more evident when a robot is moving dynamically.

Recent works have attempted to improve kinematic-inertial estimation accuracy by detecting unstable contacts and reducing their influence on the overall estimation [2], [3]. Alternatively, some works have focused on incorporating additional exteroceptive sensing modalities into the estimator to help reduce the pose error [4].

The authors are with the Oxford Robotics Institute, Department of Engineering Science, University of Oxford, UK. {davidw, mcamurri, mfallon}@robots.ox.ac.uk

Fig. 1. ANYmal trotting over a field of small rocks at the Swiss Military Rescue Centre in Wangen an der Aare (Switzerland). The ground truth was collected using a Leica TS16 laser tracker (visible in the background). Video: http://youtu.be/w1Sx6dIqgQo

These approaches model the contact locations as being fixed and affected only by Gaussian noise. Both assumptions fail in conditions such as non-rigid terrain, kinematic chain flexibility, and foot slippage.

A. Motivation

Our work is motivated by the observation that there is an approximately constant velocity bias from the kinematic-inertial state estimator on the ANYmal robot during dynamic locomotion. An example is shown in Fig. 2, where the robot's estimated altitude grows linearly as the robot moves. We attribute this behavior to the compression of legs and the ground during the contact events.

One approach would be to further model the dynamic properties of the robot [5] or the terrain directly within the estimator. However, this is likely to be robot specific and terrain dependent: improving performance in one situation but degrading it elsewhere. Instead, we propose to extend the state of the estimator with a velocity bias term which is estimated using vision and then to reject all such effects.

Inspired by the IMU bias estimation and preintegration methods from [6], we propose a novel leg odometry factor that performs online velocity preintegration and bias estimation to compensate for characteristic drift in leg odometry.

This factor was implemented as a concurrent thread within our VILENS framework [7], a visual-inertial-legged estimator which uses GTSAM for optimization [8]. Thanks to forward propagation, the thread can output the best pose and velocity estimates at $400\,\text{Hz}$ directly, or just update the bias terms of the estimator running inside the robot's control loop. The optimized estimate is available from the optimizer thread at at $30\,\text{Hz}$ for effective local mapping.
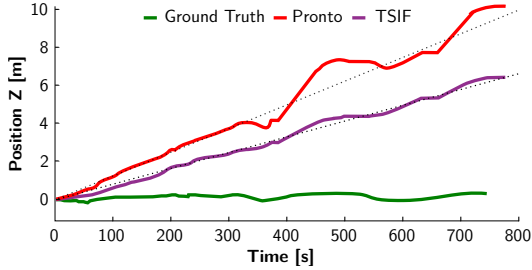
Fig. 2. Comparison between estimated robot altitude by Pronto [2] (blue) and TSIF [9] (magenta) kinematic-inertial state estimators, against ground truth (green) on the SMR1 dataset. Despite local fluctuations, the drift has a characteristic linear growth.

Sensor	Model	Hz	Specs
IMU	Xsens MTi-100	400	*Init Bias:* $0.2\,°/\text{s}$ | $5\,\text{mg}$ *Bias Stab:* $10\,°/\text{h}$ | $15\,\text{mg}$
Stereo Camera	RealSense D435i	30	*Resolution*: 848×480 px *FoV:* $91.2° \times 65.5°$ *Imager:* IR global shutter
Encoder	ANYdrive	400	*Resolution:* $<0.025°$
Torque	ANYdrive	400	*Resolution:* $<0.1\,\text{N m}$

TABLE I

B. Contribution

This paper builds upon state-of-the-art methods for online IMU bias estimation [6] and the authors' previous work [7] to improve state estimation performance of legged robots in a variety of difficult scenarios where kinematic-inertial estimates would drift significantly. Compared to previous research, we present the following contributions:

- We present a novel factor graph approach that tightly fuses leg odometry as velocity constraints (as opposed to position constraints), with stereo vision and IMU measurements;
- We present the first visual-inertial-legged odometry solution that explicitly accounts for error in leg odometry (that can be caused by terrain/leg deformation and slippage) by extending the state with a velocity bias term;
- We show that estimating leg odometry error can reduce RPE by 26% in extensive outdoor experiments on muddy ground, slopes, and rockbeds with the ANYmal robot (Fig. 1).

The remainder of this article is presented as follows: in Section II we review the literature on mobile state estimation with a focus on challenging, outdoor conditions; Section III formally defines the problem addressed by the paper and provides the required mathematical background; Section IV describes the factors used in our proposed formulation; Section V presents the implementation details of our physical system; Section VI presents the experimental results and their interpretation; Section VII concludes with final remarks.

II. RELATED WORK

In legged robotics, slippage and/or deformation have been typically addressed by assuming the contact location of a stance foot is always static throughout the stance period (yet affected by Gaussian noise). Thus, the main focus has been on detecting and ignoring the feet that are not in fixed contact with the ground. These methods would typically perform filtering using only proprioceptive sensing, with a few exceptions.

Bloesch et al. [1] proposed an Unscented Kalman Filter design that fuses IMU and differential kinematics. The approach used a threshold on the Mahalanobis distance of the filter innovation to infer outliers which were then ignored.

Ma et al. [10] proposed an Extended Kalman Filter (EKF) design that was Visual Odometry (VO) driven. They incorporated kinematics only when VO failed or a simple heuristic criteria was met (e.g., when roll or pitch are greater than $45°$ slippage was assumed and leg odometry ignored). Using high-grade sensors, they were able to achieve 1% error over several kilometers of experiments.

Camurri et al. [2] proposed an EKF fusing IMU and differential kinematics similar to [1]. Instead of the Mahalanobis distance on the filter innovation, they developed probabilistic contact and impact detectors. The contact detector learns the optimal force threshold to detect a foot in contact for a specific gait, while the impact detector adapts the measurement covariance to reject unreliable measurements. We used this approach to fuse each leg's kinematic measurements into a single velocity measurement for our proposed factor graph method.

Recently, Jenelten et al. [3] presented a probabilistic contact and slip detector which used a Hidden Markov Model for the ANYmal quadruped robot. Using differential kinematics, the authors were able to successfully detect slippage events and robustify locomotion on slippery surfaces. However, they did not address pose estimate drift.

III. PROBLEM STATEMENT

Our quadruped robot has 12 active Degrees-of-Freedom (DoF) and is equipped with a stereo camera, an IMU, joint encoders and torque sensors (see Table I for the specifications). We aim to estimate the history of the robot's base link pose and its velocity (linear and angular) over time. In contrast to previous works, we propose to estimate velocity biases (in addition to IMU biases) to compensate for leg odometry drift, as detailed in the following section.

The relevant reference frames are specified in Fig. 3 and include: the left camera frame C, the IMU frame I, the fixed-world frame W, and the base frame B. When a foot is in contact with the ground, a contact frame K is also defined.

Unless otherwise specified, position ${}_\text{W}\mathbf{p}_\text{WB}$ and orientation $\mathbf{R}_\text{WB}$ of the base are expressed in world coordinates, velocities of the base ${}_\text{B}\mathbf{v}_\text{WB}, {}_\text{B}\boldsymbol{\omega}_\text{WB}$ are in base coordinates (see [11]), IMU biases ${}_\text{I}\mathbf{b}^g$, ${}_\text{I}\mathbf{b}^a$ are expressed in the IMU frame, and the velocity biases ${}_\text{B}\mathbf{b}^\omega$, ${}_\text{B}\mathbf{b}^v$ are expressed in the base frame.

A. State Definition

The robot state at time t_i is defined as follows:

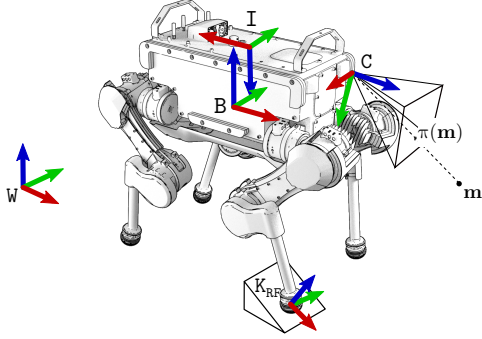$$\boldsymbol{x}_i \triangleq [\mathbf{R}_i, \mathbf{p}_i, \mathbf{v}_i, \mathbf{b}_i] \tag{1}$$

Fig. 3. Reference frames conventions. The world frame W is fixed to earth, while the base frame B, the camera's optical frame C, and the IMU frame, I are rigidly attached to the robot's chassis. When a foot touches the ground (e.g., the Right Front, RF), a contact frame K (perpendicular to the ground and parallel to W's y-axis) is defined. The projection of a landmark point $\mathbf{m}$ onto the image plane is $\pi(\mathbf{m})$.

where: $\mathbf{R}_i \in \mathrm{SO}(3)$ is the orientation, $\mathbf{p}_i \in \mathbb{R}^3$ is the position, $\mathbf{v}_i \in \mathbb{R}^3$ is the linear velocity. The bias vector $\mathbf{b}_i$ is composed as follows:

$$\mathbf{b}_i = [\mathbf{b}_i^g \ \mathbf{b}_i^a, \ \mathbf{b}_i^\omega, \ \mathbf{b}_i^v] \in \mathbb{R}^{12} \tag{2}$$

where the first two elements are the usual IMU gyro and accelerometer biases, and the last two $[\mathbf{b}_i^\omega, \ \mathbf{b}_i^v] = \mathbf{b}_i^Q$ are our proposed angular/linear velocity biases from leg odometry.

This builds upon the formulation from our previous work [7], by incorporating a preintegrated velocity factor (as opposed to a relative pose factor where the integration was operated by an external filter).

In addition to the robot state, we estimate the position of all observed visual landmarks $\mathbf{m}_\ell$. The objective of our estimation problem is then the union of all the robot states and landmarks visible up to the current time t_k:

$$\mathcal{X}_k \triangleq \bigcup_{\forall i \in \mathsf{K}_k} \left[\{\boldsymbol{x}_i\}, \bigcup_{\forall \ell \in \mathsf{M}_i} \{\mathbf{m}_\ell\} \right] \tag{3}$$

where $\mathsf{K}_k, \mathsf{M}_i$ are the lists of all the keyframe indices up to time t_k and all landmark indices visible at time t_i, respectively.

B. Measurements Definition

For each new stereo camera frame $\mathcal{C}_i$, collected at time t_i, we receive a number of IMU measurements $\mathcal{I}_{ij}$ collected between t_i and t_j. We also define $\mathcal{V}_{ij}$ as the angular and linear velocity measurements from an external source of leg odometry (Pronto or TSIF). This source would account for the fusion of multiple legs in contact into one velocity measurement per joint state measurement. The set of all measurements up to time t_k is therefore defined as:

$$\mathcal{Z}_k \triangleq \bigcup_{\forall i \in \mathsf{K}_k} \{\mathcal{C}_i, \mathcal{I}_{ij}, \mathcal{V}_{ij}\} \tag{4}$$

C. Maximum-a-Posteriori Estimation

We wish to maximize the likelihood of the measurements $\mathcal{Z}_k$ given the history of states $\mathcal{X}_k$:

$$\mathcal{X}_k^* = \arg\max_{\mathcal{X}_k} p(\mathcal{X}_k|\mathcal{Z}_k) \propto p(\mathcal{X}_0)p(\mathcal{Z}_k|\mathcal{X}_k) \tag{5}$$
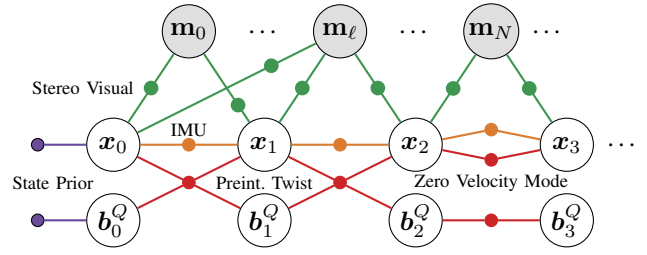

Fig. 4. VILENS factor graph structure, showing initial prior, visual, IMU, and preintegrated velocity factors. When a zero velocity state is detected (e.g. between x_2 and x_3) then the velocity bias is not used and is kept constant (we assume the bias is only present when the robot is moving).

As the measurements are formulated as conditionally independent and corrupted by white Gaussian noise, Eq. (5) can be formulated as a least squares minimization problem:

$$\mathcal{X}_k^* = \arg\min_{\mathcal{X}_k} \|\mathbf{r}_0\|_{\Sigma_0}^2 + \sum_{i \in \mathsf{K}_k} \|\mathbf{r}_{\mathcal{I}_{ij}}\|_{\Sigma_{\mathcal{I}_{ij}}}^2 + \sum_{i \in \mathsf{K}_k} \|\mathbf{r}_{\mathcal{V}_{ij}}\|_{\Sigma_{\mathcal{V}_{ij}}}^2$$
$$+ \sum_{i \in \mathsf{K}_k} \|\mathbf{r}_{\mathbf{b}_{ij}}\|_{\Sigma_\mathbf{b}}^2 + \sum_{i \in \mathsf{K}_k} \sum_{\ell \in \mathsf{M}_i} \|\mathbf{r}_{i,\mathbf{m}_\ell}\|_{\Sigma_\mathbf{m}}^2 \tag{6}$$

where each term is the residual associated to a factor type, weighted by its covariance matrix; specifically the residuals are: state prior, IMU, velocity, biases and landmarks.

IV. FACTOR GRAPH FORMULATION

In the following sections we describe the measurements, residuals and covariances of the factors which compose the factor graph shown in Fig. 4. For convenience, we summarize the IMU factors from [6] in Section IV-A; our novel velocity factor is detailed in Section IV-B; Sections IV-C and IV-D describe the bias and stereo visual residuals, which are adapted from [6], [7] to include the velocity bias term and stereo cameras, respectively.

A. Preintegrated IMU Factors

In the standard manner, the IMU measurements are preintegrated to constrain the pose and velocity between two consecutive nodes of the graph, and provide high frequency state updates between them. This uses a residual of the form:

$$\mathbf{r}_{\mathcal{I}_{ij}} = \left[\mathbf{r}_{\Delta\mathbf{R}_{ij}}^\mathsf{T}, \mathbf{r}_{\Delta\mathbf{v}_{ij}}^\mathsf{T}, \mathbf{r}_{\Delta\mathbf{p}_{ij}}^\mathsf{T} \right] \tag{7}$$

where $\mathcal{I}_{ij}$ are the IMU measurements between t_i and t_j. The individual elements of the residual are defined as:

$$\mathbf{r}_{\Delta\mathbf{R}_{ij}} = \mathrm{Log}\left(\Delta\tilde{\mathbf{R}}_{ij}(\mathbf{b}_i^g) \right) \mathbf{R}_i^\mathsf{T} \mathbf{R}_j \tag{8}$$

$$\mathbf{r}_{\Delta\mathbf{v}_{ij}} = \mathbf{R}_i^\mathsf{T} \left(\mathbf{v}_j - \mathbf{v}_i - \mathbf{g}\Delta t_{ij} \right) - \Delta\tilde{\mathbf{v}}_{ij}(\mathbf{b}_i^g, \mathbf{b}_i^a) \tag{9}$$

$$\mathbf{r}_{\Delta\mathbf{p}_{ij}} = \mathbf{R}_i^\mathsf{T} \left(\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i \Delta t_{ij} - \frac{1}{2}\mathbf{g}\Delta t_{ij}^2 \right)$$
$$- \Delta\tilde{\mathbf{p}}_{ij}(\mathbf{b}_i^g, \mathbf{b}_i^a) \tag{10}$$

for the definition of the preintegrated IMU measurements $\Delta\tilde{\mathbf{R}}_{ij}, \Delta\tilde{\mathbf{p}}_{ij}, \Delta\tilde{\mathbf{v}}_{ij}$, the noise terms $\delta\boldsymbol{\phi}, \delta\mathbf{v}, \delta\mathbf{p}$, and the covariance matrix $\Sigma_{\mathcal{I}_{ij}}$, the reader is invited to consult [6].

B. Preintegrated Velocity Factors

1) Leg Odometry: When a leg is in rigid, non-slipping contact with the ground, the robot's linear velocity can be computed from the foot velocity and position in base frame:

$$_B\mathbf{v}_{WB} = -_B\mathbf{v}_{BK} - _B\boldsymbol{\omega}_{WB} \times _B\mathbf{p}_{BK} \qquad (11)$$

From the sensed joint positions and velocities $\tilde{\mathbf{q}}, \dot{\tilde{\mathbf{q}}}$ and noise $\boldsymbol{\eta}^q, \boldsymbol{\eta}^{\dot{q}}$ we can rewrite Eq. (11) as a linear velocity measurement [1]:

$$\tilde{\mathbf{v}} = -J(\tilde{\mathbf{q}} - \boldsymbol{\eta}^q) \cdot (\dot{\tilde{\mathbf{q}}} - \boldsymbol{\eta}^{\dot{q}}) - \boldsymbol{\omega} \times f(\tilde{\mathbf{q}} - \boldsymbol{\eta}^q) \qquad (12)$$

where $f(\cdot)$ and $J(\cdot)$ are the forward kinematics function and its Jacobian, respectively.

Eq. (12) is valid only when the corresponding leg is in contact with the ground. However, this happens intermittently while the robot moves. Since multiple legs can be in contact simultaneously, measurement fusion is necessary. To do this we take advantage of the contact detection and sensor fusion features of the EKF filter in [2] and use it as an independent source of unified velocity measurements $\tilde{\mathbf{v}}, \tilde{\boldsymbol{\omega}}$.

2) Velocity Bias: On slippery/deformable terrains, the constraint from Eq. (11) might not be respected, leading to leg odometry drift and inconsistency with visual odometry. In our experience this drift is constant and gait dependent.

For these reasons, we relax Eq. (11) by adding a slowly varying bias term $\mathbf{b}^v$ to Eq. (12). As in [1], we also collect all the effects of encoder noise into a single term, leading to:

$$\tilde{\mathbf{v}} = -J(\tilde{\mathbf{q}})\dot{\tilde{\mathbf{q}}} - \boldsymbol{\omega} \times f(\tilde{\mathbf{q}}) + \mathbf{b}^v + \boldsymbol{\eta}^v \qquad (13)$$

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega} + \mathbf{b}^{\omega} + \boldsymbol{\eta}^{\omega} \qquad (14)$$

where the parameters for $\boldsymbol{\eta}^v, \boldsymbol{\eta}^{\omega}$ are provided by the source of velocity measurements.

3) Preintegrated Measurements: In the following, we derive the the preintegrated position and noise only. For the respective rotational quantities $\Delta\tilde{\mathbf{R}}$ and $\delta\boldsymbol{\phi}$, we refer to [6], as they have the same form as for IMU measurements.

Assuming constant velocity between t_i and t_j, we can iteratively calculate the position at time t_j as:

$$\mathbf{p}_j = \mathbf{p}_i + \sum_{k=i}^{j-1} [\mathbf{R}_k(\tilde{\mathbf{v}}_k - \mathbf{b}_i^v - \boldsymbol{\eta}_k^v)\Delta t] \qquad (15)$$

From Eq. (15) a relative measurement can be obtained:

$$\Delta\mathbf{p}_{ij} = \mathbf{R}_i^\mathsf{T}(\mathbf{p}_j - \mathbf{p}_i) = \sum_{k=i}^{j-1} [\Delta\mathbf{R}_{ik}(\tilde{\mathbf{v}}_k - \mathbf{b}_k^v - \boldsymbol{\eta}_k^v)\Delta t)] \qquad (16)$$

With the substitution $\Delta\mathbf{R}_{ik} = \Delta\tilde{\mathbf{R}}_{ik}\mathrm{Exp}\left(-\delta\boldsymbol{\phi}_{ik}\right)$ to include the preintegrated rotation measurement, and the approximation $\mathrm{Exp}\left(\boldsymbol{\phi}\right) \simeq \mathbf{I} + \boldsymbol{\phi}^\wedge$, Eq. (16) becomes:

$$\Delta\mathbf{p}_{ij} \simeq \sum_{k=i}^{j-1} \left[\Delta\tilde{\mathbf{R}}_{ik}(I - \delta\boldsymbol{\phi}_{ik}^\wedge)(\tilde{\mathbf{v}}_k - \mathbf{b}_i^v - \boldsymbol{\eta}_k^v)\Delta t\right] \quad (17)$$

By separating the measurement and noise components of Eq. (17) and ignoring the higher order terms, we can define

the preintegrated position *measurement* $\Delta\tilde{\mathbf{p}}$ and *noise* $\delta\mathbf{p}$ as:

$$\Delta\tilde{\mathbf{p}}_{ij} \triangleq \sum_{k=i}^{j-1} \left[\Delta\tilde{\mathbf{R}}_{ik}(\tilde{\mathbf{v}}_k - \mathbf{b}_i^v)\Delta t\right] \qquad (18)$$

$$\delta\mathbf{p}_{ij} \triangleq \sum_{k=i}^{j-1} \left[\Delta\tilde{\mathbf{R}}_{ik}\boldsymbol{\eta}_k^v\Delta t - \Delta\tilde{\mathbf{R}}_{ik}(\tilde{\mathbf{v}}_k - \mathbf{b}_i^v)^\wedge\delta\boldsymbol{\phi}_{ij}\Delta t\right]$$

$$(19)$$

Note that both quantities still depend on the bias states $\mathbf{b}^Q = [\mathbf{b}^\omega, \mathbf{b}^v]$; when these change, we would like to avoid the recomputation of Eq. (18). Given a small change $\delta\mathbf{b}^Q$ such that $\mathbf{b}^Q = \bar{\mathbf{b}}^Q + \delta\mathbf{b}^Q$, we approximate the measurement as:

$$\Delta\tilde{\mathbf{p}}_{ij}(\mathbf{b}^Q) \simeq \Delta\tilde{\mathbf{p}}_{ij}(\bar{\mathbf{b}}^Q) + \frac{\partial\Delta\mathbf{p}_{ij}}{\partial\mathbf{b}^\omega}\delta\mathbf{b}^\omega + \frac{\partial\Delta\mathbf{p}_{ij}}{\partial\mathbf{b}^v}\delta\mathbf{b}^v \quad (20)$$

4) Residuals: the velocity residual can be expressed as:

$$\mathbf{r}_{\mathcal{V}_{ij}} = \left[\mathbf{r}_{\Delta\mathbf{R}_{ij}}^\mathsf{T}, \mathbf{r}_{\Delta\mathbf{p}_{ij}}^\mathsf{T}\right] \qquad (21)$$

where $\mathbf{r}_{\Delta\mathbf{R}_{ij}}$ has the same form as Eq. (8) and $\mathbf{r}_{\Delta\mathbf{p}_{ij}}$ is:

$$\mathbf{r}_{\Delta\mathbf{p}_{ij}} = \mathbf{R}_i^\mathsf{T}(\mathbf{p}_j - \mathbf{p}_i) - \Delta\tilde{\mathbf{p}}_{ij}(\mathbf{b}_i^\omega, \mathbf{b}_i^v) \qquad (22)$$

5) Covariance: After simple manipulation of Eq. 19, the covariance of the residual $\mathbf{r}_{\Delta\mathbf{p}_{ij}}$ can be expressed as a linear combination of the preintegrated and current sensor noise:

$$\Sigma_{\mathcal{V}_{ik+1}} = A\,\Sigma_{\mathcal{V}_{ik}}\,A^\mathsf{T} + B\,\Sigma_\mathcal{V}^\eta\,B^\mathsf{T} \qquad (23)$$

$\Sigma_{\mathcal{V}_{ik}}$ evolves over time while $\Sigma_\mathcal{V}^\eta$ is fixed and depends sensor specifications. The multiplicative terms are:

$$A = \begin{bmatrix} \Delta\tilde{\mathbf{R}}_{ik}^\mathsf{T} & 0 \\ -\Delta\tilde{\mathbf{R}}_{ik}(\tilde{\mathbf{v}}_k - \mathbf{b}^v)^\wedge\Delta t & \mathbf{I} \end{bmatrix}, B = \begin{bmatrix} J_r^k\Delta t & 0 \\ 0 & \Delta\tilde{\mathbf{R}}_{ik}\Delta t \end{bmatrix}$$

$$(24)$$

where J_r is the right Jacobian of $SO(3)$ and the other terms are manipulations of $\delta\boldsymbol{\phi}$ and $\delta\mathbf{p}$ from [6] and Eq. (19).

C. Bias Residuals

The biases from Eq. (2) are intended to change slowly and are therefore modeled as a Gaussian random walk. The residual term for the cost function is therefore:

$$\|\mathbf{r}_{\mathbf{b}_{ij}}\|_{\Sigma_\mathbf{b}}^2 \triangleq \|\mathbf{b}_j^g - \mathbf{b}_i^g\|_{\Sigma_{\mathbf{b}^g}}^2 + \|\mathbf{b}_j^a - \mathbf{b}_i^a\|_{\Sigma_{\mathbf{b}^a}}^2 +$$
$$+ \|\mathbf{b}_j^\omega - \mathbf{b}_i^\omega\|_{\Sigma_{\mathbf{b}^\omega}}^2 + \|\mathbf{b}_j^v - \mathbf{b}_i^v\|_{\Sigma_{\mathbf{b}^v}}^2 \quad (25)$$

where the covariance matrices are determined by the expected rate of change of these quantities, depending on the IMU specifications or the drift rate of the leg odometry.

D. Stereo Visual Factors

Given a stereo pair of rectified images, the stereo visual odometry residual is the difference between the measured landmark pixel locations $(u^L, v), (u^R, v)$, and the reprojection of the estimated landmark location into image coordinates, $(\pi_u^L, \pi_v), (\pi_u^R, \pi_v)$ using the standard radial-tangential distortion model. The residual at pose i for landmark $\mathbf{m}_\ell$ is:

$$\mathbf{r}_{i,\mathbf{m}_\ell} = \begin{pmatrix} \pi_u^L(\mathbf{R}_i, \mathbf{p}_i, \mathbf{m}_\ell) - u_{i,\ell}^L \\ \pi_u^R(\mathbf{R}_i, \mathbf{p}_i, \mathbf{m}_\ell) - u_{i,\ell}^R \\ \pi_v(\mathbf{R}_i, \mathbf{p}_i, \mathbf{m}_\ell) - v_{i,\ell} \end{pmatrix} \qquad (26)$$

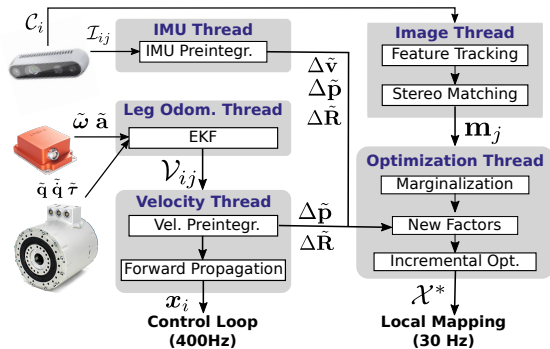where $\Sigma_\mathbf{m}$ is computed using an uncertainty of 0.5 pixels.

Fig. 5. The VILENS architecture with preintegrated velocity bias estimation.

V. IMPLEMENTATION

The state estimation architecture is shown in Fig. 5. Four parallel threads execute the following operations: preintegration of the IMU factor, preintegration of the velocity factor, stereo feature tracking, and optimization. This approach outputs $400\,\mathrm{Hz}$ velocity and pose estimates from the preintegration thread for use by the robot's control system, and a $30\,\mathrm{Hz}$ output from the factor-graph optimization thread for use by local mapping. When a new keyframe is processed, the preintegrated measurements and tracked landmarks are collected by the optimization thread, while the other threads process the next set of measurements. The factor graph optimization is implemented using the efficient incremental optimization solver iSAM2 [12], which is part of the GTSAM library [13]. We limit the number of states in the graph to 500 to keep the optimization time approximately constant.

A. Visual Feature Tracking

We detect features using the FAST corner detector, and track them between successive frames using the KLT feature tracker. Outliers are rejected using a RANSAC-based rejection method. Thanks to the parallel architecture and incremental optimization, all frames are used as keyframes, achieving $30\,\mathrm{Hz}$ nominal output. In contrast to [7], we use the Dynamic Covariance Scaling (DCS) [14] robust cost function to reduce the effect of landmark correspondence outliers on the optimization.

B. Zero Velocity Update Factors

To limit drift and factor graph growth when the robot is stationary, we enforce zero velocity updates on the different sensor modalities (camera, IMU, and leg odometry). If two out of three modalities report no motion, a zero velocity constraint factor is added to the graph. The IMU and leg odometry threads report zero velocity when position (rotation) is less than $0.1\,\mathrm{mm}$ $(0.5°)$ between two keyframes. The image thread reports zero velocity when less than 0.5 pixels displacement of all the features is detected over the same period.

VI. EXPERIMENTAL RESULTS

We have tested our proposed algorithm on a variety of terrain types for a total time of $53\,\mathrm{min}$ and $403\,\mathrm{m}$ traveled distance. The datasets consist of four scenarios (Fig. 6):

Fig. 6. Onboard camera feed from the four scenarios evaluated. *Top left*: wet concrete (FSC); *Top right:* gravel and grass (SMR1); *Bottom left:* rockbeds (SMR2); *Bottom right:* muddy grass uphill (SMR3).

10 m Relative Pose Error (RPE) $\mu(\sigma)$ [m]				
Data	**TSIF**[9]	**V-VI**	**V-RP**[7]	**V-VB**
FSC	0.49 (0.36)	0.42 (0.32)	0.47 (0.40)	**0.36** (0.30)
SMR1	0.96 (0.44)	**0.36** (0.29)	**0.36** (0.32)	**0.36** (0.33)
SMR2	0.69 (0.23)	**0.24** (0.16)	0.45 (0.10)	**0.24** (0.16)
SMR3	0.87 (0.42)	0.43 (0.48)	0.53 (0.46)	**0.39** (0.48)

TABLE II

- **FSC** a $240\,\mathrm{m}$ long trajectory consisting of three loops on wet concrete, standing water/oil, gravel and mud;
- **SMR1** a $106\,\mathrm{m}$ straight trot over concrete, gravel and high grass, followed by two loops on short grass alternating between dynamic and static gaits;
- **SMR2** a $22\,\mathrm{m}$ straight trot on rockbeds;
- **SMR3** a $35\,\mathrm{m}$ trot in a loop uphill with grass, mud and external disturbances applied to the robot to cause slippage events.

The first dataset was collected at Fire Service College (FSC), Moreton-on-Marsh, UK; the other three at the Swiss Military Rescue Center (SMR), Wangen an der Aare, Switzerland. Different copies of the ANYmal robot were used in the experiments. The attached video gives a sense of the conditions.

To generate ground truth, we tracked the robot using the Leica TS16 laser tracker (shown in Fig. 1), which provides millimeter accurate position measurements at $5\,\mathrm{Hz}$. The orientation was reconstructed with an optimization method similar to that used by the EuRoC dataset [15].

We have evaluated the Relative Pose Error (RPE) over a distance of $10\,\mathrm{m}$ for the following algorithms:

- **V-VB:** VILENS with our proposed velocity bias factors;
- **V-RP:** VILENS with leg odometry integrated as relative pose factors, as used in our previous work [7];
- **V-VI:** a pure visual inertial navigation system (i.e., VILENS without leg odometry factors);
- **TSIF:** default ANYmal state estimator [9], our baseline.

Note that the same IMU and camera settings have been used for all configurations and datasets. Also, in comparison to our previous work [7], the robustness of visual feature tracking has been improved due to the introduction of stereo factors, higher framerates and robust cost functions.

The results are summarized in Table II. In three of the four datasets (FSC, SMR2 and SMR3), V-VI is able to

Fig. 7. Aerial view of the trajectories for method V-VB and the ground truth on the FSC dataset (240 m traveled). The common start point is indicated with a circle and the two endpoints with a cross. The lower part of the trajectory has no ground truth. Note that V-VB has no loop-closure system.
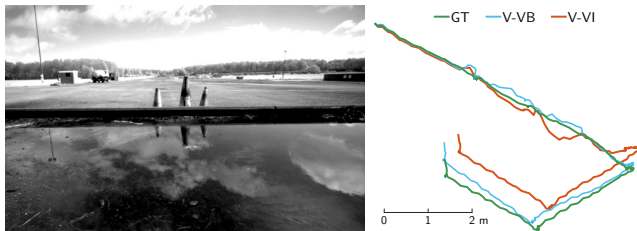

Fig. 8. *Left:* Onboard image of a visually challenging scene containing reflections from a large, oily puddle. *Right:* Top-down comparison of V-VB and V-VI trajectories aligned with ground truth while crossing the puddle.
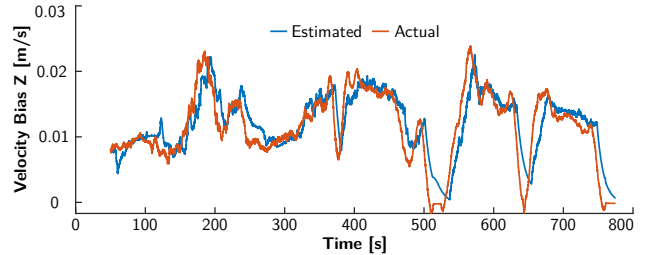

Fig. 9. Using visual information, VILENS is able to accurately and stably estimate the bias of the kinematic-inertial estimate. Experiment: SMR1.
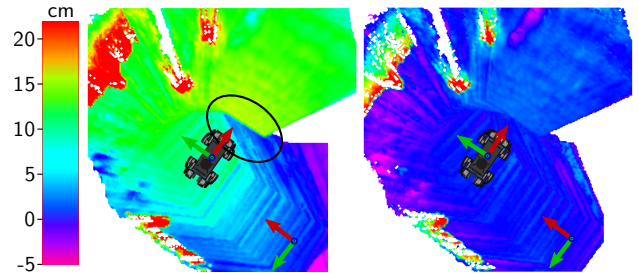

Fig. 10. Terrain reconstruction comparison between TSIF [9] (left) and VILENS (right). The robot walks from the bottom right corner to the center of the image and turns 90° right. VILENS eliminates the 15 cm drift and drastically reduces the number of artifacts on the elevation map.

outperform V-RP. This is because the datasets are designed to be particularly challenging for leg odometry. Therefore, the inclusion of relative pose factors without compensating for leg odometry drift actually degrades the performance compared to a visual-inertial only system. With the preintegrated velocity bias estimation, leg odometry improves the estimate up to 14 % compared to V-VI and 26 % compared to V-RP.

The global performance is shown in Fig. 7, which depicts the estimated and ground truth trajectories on the 240 m FSC dataset. By incorporating visual information to reject drift, the final z position of V-VB is 8.6 cm above ground truth, compared to a drift of 4.02 m from the TSIF kinematic-inertial estimator. Note that since VILENS is an odometry system, no loop closures have been performed.

A. Analysis of Visually Challenging Episodes

Most of the datasets presented favorable conditions for VO (well lit static scenes with texture). However, there were also certain locations where feature tracking struggled.

Fig. 8 shows a situation from FSC dataset where the robot traverses a large puddle. V-VI tracks the features on the water, causing drift in the lateral direction. Instead, V-VB maintains a better pose estimate by relying on leg odometry, whose drift is suppressed using the estimated velocity bias.

B. Velocity Bias Evolution

We have compared the estimated online bias in the z-axis to a lowpass filtered version of the same signal from the Pronto EKF [2] (Fig. 9). The sequence analyzed is the same as the one shown in Fig. 2. Since the z-axis position and average velocity of the robot are zero, the high correlation

between the two signals demonstrates the effectiveness of leg odometry drift rejection.

C. Terrain Reconstruction Assessment

We have evaluated the quality of local terrain mapping during a sequence of walking and turning on flat ground from the FSC dataset (Fig. 10). Due to drift in the ANYmal's internal filter [9], the elevation map contains a phantom discontinuity in front of the robot (encircled in black). With VILENS, the drift is reduced for effective footstep planning.

VII. CONCLUSION

We have presented a novel factor graph formulation for state estimation that estimates preintegrated velocity factors for leg odometry and velocity bias estimation to accommodate for leg odometry drift. These bias effects are difficult to directly model, we instead infer them from vision. The redundancy of our approach is also demonstrated in visual impoverished situations which vision along would struggle. In these situations, our system gracefully relies on leg odometry and the velocity bias estimation compensates for its drift. We have demonstrated the robustness of our method with outdoor experiments which include conditions such as slippery and deformable terrain, reflections, and external disturbances applied to the robot.

VIII. ACKNOWLEDGEMENTS

References

[1] M. Bloesch, C. Gehring, P. Fankhauser, M. Hutter, M. A. Hoepflinger, and R. Siegwart, "State estimation for legged robots on unstable and slippery terrain," in *IEEE International Conference on Intelligent Robots and Systems*, 2013, pp. 6058–6064.

[2] M. Camurri, M. Fallon, S. Bazeille, A. Radulescu, V. Barasuol, D. G. Caldwell, and C. Semini, "Probabilistic contact estimation and impact detection for state estimation of quadruped robots," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1023–1030, 2017.

[3] F. Jenelten, J. Hwangbo, F. Tresoldi, C. D. Bellicoso, and M. Hutter, "Dynamic locomotion on slippery ground," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4170–4176, Oct 2019.

[4] R. Hartley, M. G. Jadidi, L. Gan, J.-K. Huang, J. W. Grizzle, and R. M. Eustice, "Hybrid contact preintegration for visual-inertial-contact state estimation within factor graphs," in *IEEE International Conference on Intelligent Robots and Systems*, 2018.

[5] T. Koolen, S. Bertrand, G. Thomas, T. De Boer, T. Wu, J. Smith, J. Englsberger, and J. Pratt, "Design of a momentum-based control framework and application to the humanoid robot atlas," *International Journal of Humanoid Robotics*, vol. 13, no. 01, p. 1650007, 2016.

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

[7] D. Wisth, M. Camurri, and M. Fallon, "Robust legged robot state estimation using factor graph optimization," *IEEE Robotics and Automation Letters*, 2019.

[8] F. Dellaert, "Factor Graphs and GTSAM: A Hands-on Introduction, Tech. Rep. September, 2012. [Online]. Available: https://research.cc. gatech.edu/borg/sites/edu.borg/files/downloads/gtsam.pdf

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

[10] J. Ma, M. Bajracharya, S. Susca, L. Matthies, and M. Malchano, "Real-time pose estimation of a dynamic quadruped in GPS-denied environments for 24-hour operation," *International Journal of Robotics Research*, vol. 35, no. 6, pp. 631–653, 2016.

[11] P. Furgale, "Representing Robot Pose: The good, the bad, and the ugly," 2014. [Online]. Available: http://paulfurgale.info/news/2014/6/9/ representing-robot-pose-the-good-the-bad-and-the-ugly

[12] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. 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, 2012.

[13] F. Dellaert and M. Kaess, *Factor Graphs for Robot Perception*, 2017, vol. 6, no. 1-2. [Online]. Available: http://www.nowpublishers.com/ article/Details/ROB-043

[14] K. MacTavish and T. D. Barfoot, "At All Costs: A comparison of robust cost functions for camera correspondence outliers," in *Conference on Computer and Robot Vision*, 2015, pp. 62–69.

[15] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.