Hybrid Contact Preintegration for Visual-Inertial-Contact State Estimation Using Factor Graphs

Ross Hartley, Maani Ghaffari Jadidi, Lu Gan, Jiunn-Kai Huang, Jessy W. Grizzle, and Ryan M. Eustice

Abstract— The factor graph framework is a convenient modeling technique for robotic state estimation where states are represented as nodes, and measurements are modeled as factors. When designing a sensor fusion framework for legged robots, one often has access to visual, inertial, joint encoder, and contact sensors. While visual-inertial odometry has been studied extensively in this framework, the addition of a preintegrated contact factor for legged robots has been only recently proposed. This allowed for integration of encoder and contact measurements into existing factor graphs, however, new nodes had to be added to the graph every time contact was made or broken. In this work, to cope with the problem of switching contact frames, we propose a hybrid contact preintegration theory that allows contact information to be integrated through an arbitrary number of contact switches. The proposed hybrid modeling approach reduces the number of required variables in the nonlinear optimization problem by only requiring new states to be added alongside camera or selected keyframes. This method is evaluated using real experimental data collected from a Cassie-series robot where the trajectory of the robot produced by a motion capture system is used as a proxy for ground truth. The evaluation shows that inclusion of the proposed preintegrated hybrid contact factor alongside visual-inertial navigation systems improves estimation accuracy as well as robustness to vision failure, while its generalization makes it more accessible for legged platforms.

I. INTRODUCTION AND RELATED WORK

Long-term state estimation and mapping for legged robots requires a flexible sensor fusion framework that allows for reducing drift and correcting past estimates as the robot perceives new information. During long-term missions, odometry systems can drift substantially since the absolute position and yaw (rotation about gravity) are unobservable [1], [2], leading to an unbounded growth in the covariance of the estimate and an undesirable expansion of the search space for data association tasks. The factor graph smoothing framework [3]–[6] offers suitable machineries for building such systems in which real-time performance is achieved by exploiting the sparse structure of the Simultaneous Localization and Mapping (SLAM) problem [7], [8]. In addition, the incorporation of loop-closures [8] into the graph, upon availability, is convenient.

Legged robot perception often involves fusing kinematic odometry, Inertial Measurement Unit (IMU) data, and visual/depth measurements to infer the robot's trajectory, controller inputs such as velocity, and calibration parameters [9]–[12]. The challenge in such perception problems is the rigorous real-time performance requirements in legged

The authors are with the College of Engineering, University of Michigan, Ann Arbor, MI 48109 USA {rosshart, maanigj, ganlu, bjhuang, grizzle, eustice}@umich.edu.
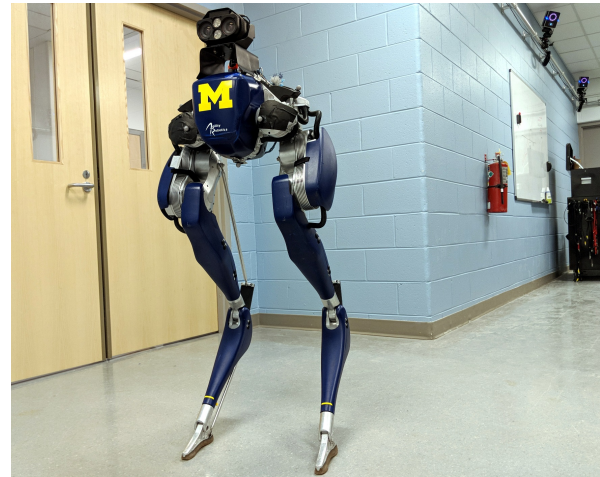
Fig. 1: Experiments were conducted on a Cassie-series robot designed by Agility Robotics in an indoor laboratory environment. The motion capture system is used to record the robot trajectory as a proxy for ground truth data. The Cassie-series robot has 20 degrees of freedom, 10 actuators, joint encoders, an IMU, and mounted with a Multisense S7 stereo camera.

robots arising from their direct and switching contact with the environment [1], [13]–[16]. Furthermore, kinematic leg odometry involves estimating relative transformations and velocity using kinematic and contact information, which can be noisy due to the encoder noise and foot slip [2], [17].

In our previous work [18], we made progress towards a long-term, legged robot state estimator by developing two novel factors that integrate the use of forward kinematics and the notion of contact between the robot and the environment into the factor graph framework. While these novel factors improved the estimator's performance, new challenges emerged due to frequent switching contacts as the robot navigates through an environment. In particular, a new node needs to be added to the graph every time contact was made or broken. This leads to an larger number of optimization variables, which increases the complexity of the graph and ultimately slows down the estimator. This problem is only exacerbated as the number of contact points increases (quadraped or hexapod robots).

In this paper, we generalize the idea of the preintegrated contact factor in [18] to a *hybrid preintegrated contact factor* that alleviates the effect of frequently changing contacts with the environment. We develop a novel method for preintegrating contact information though an arbitrary number of contact switches. The proposed hybrid modeling approach reduces the number of required variables in the nonlinear optimization problem by only requiring new states to be added alongside camera or selected keyframes. The present

work has the following contributions:

i. A generic forward kinematic factor modeled in SE(3) using the manipulator Jacobian that supports both prismatic and revolute joints; the factor incorporates noisy encoder measurements to estimate an end-effector pose at any time-step.

ii. A hybrid preintegrated contact factor modeled in SE(3) that allows for an arbitrary number of contact switches between camera or selected keyframes.

iii. Real-time implementation and experimental evaluation of the derived factors on a Cassie-series biped robot in an indoor laboratory environment where ground truth data was collected using a motion capture system.

The remainder of this paper is organized as follows. Section II provides the mathematical background and preliminaries. We formulate the problem using the factor graph approach in Section III. Section IV explains the forward kinematic modeling and derives the forward kinematic factor. The proposed hybrid rigid contact model and hybrid contact preintegration are derived in Section V. Experimental evaluations of the proposed methods on a 3D biped robot (shown in Fig. 1) are presented in Section VI. Finally, Section VII concludes the paper and provides future work suggestions.

II. MATHEMATICAL BACKGROUND AND PRELIMINARIES

We first review the Lie group theory corresponding to the rotation and motion groups [19], [20]. Afterwards, we discuss the optimization technique on matrix Lie groups [21] and the choice of retraction for deriving forward kinematic and hybrid preintegrated contact factors.

Matrices are capitalized in bold, such as in $\mathbf{X}$, and vectors are column-wise in lower case bold type, such as in $\mathbf{x}$. We denote $\|\mathbf{e}\|_{\boldsymbol{\Sigma}}^2 \triangleq \mathbf{e}^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \mathbf{e}$. The n-by-n identity matrix and the n-by-m matrix of zeros are denoted by $\mathbf{I}_n$ and $\mathbf{0}_{n\times m}$ respectively. The vector constructed by stacking x_i, $\forall\, i \in \{1, \ldots, n\}$ is denoted by $\mathrm{vec}(x_1, \ldots, x_n)$. The covariance of a random vector is denoted by $\mathbf{Cov}[\cdot]$. Finally, we denote the base (IMU) frame of the robot by B, the world frame by W, and contact frame by C.

A. Matrix Lie Group of Rotation and Motion in $\mathbb{R}^3$

The general linear group of degree n, denoted by $\mathrm{GL}_n(\mathbb{R})$, is the set of all $n \times n$ nonsingular real matrices, where the group binary operation is the ordinary matrix multiplication. The three-dimensional (3D) special orthogonal group, denoted by $\mathrm{SO}(3) = \{\mathbf{R} \in \mathrm{GL}_3(\mathbb{R}) \mid \mathbf{R}\mathbf{R}^{\mathsf{T}} = \mathbf{I}_3, \det \mathbf{R} = +1\}$ is the rotation group on $\mathbb{R}^3$. The 3D special Euclidean group, denoted by

$$\mathrm{SE}(3) = \{\mathbf{H} = \begin{bmatrix} \mathbf{R} & \mathbf{p} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \in \mathrm{GL}_4(\mathbb{R}) \mid \mathbf{R} \in \mathrm{SO}(3), \mathbf{p} \in \mathbb{R}^3 \}$$

is the group of rigid transformations on $\mathbb{R}^3$. The Lie algebra (tangent space at the identity together with Lie bracket) of $\mathrm{SO}(3)$, denoted by $\mathfrak{so}(3)$, is the set of 3×3 skew-symmetric

matrices such that for any $\boldsymbol{\omega} \triangleq \mathrm{vec}(\omega_1, \omega_2, \omega_3) \in \mathbb{R}^3$:

$$\boldsymbol{\omega}^{\wedge} \triangleq \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix},$$

and $(\boldsymbol{\omega}^{\wedge})^{\vee} = \boldsymbol{\omega}$. The Lie algebra of $\mathrm{SE}(3)$, denoted by $\mathfrak{se}(3)$, can be identified by 4×4 matrices such that for any $\boldsymbol{\omega}, \mathbf{v} \in \mathbb{R}^3$ and the twist is defined as $\boldsymbol{\xi} \triangleq \mathrm{vec}(\boldsymbol{\omega}, \mathbf{v}) \in \mathbb{R}^6$:

$$\boldsymbol{\xi}^{\wedge} \triangleq \begin{bmatrix} \boldsymbol{\omega}^{\wedge} & \mathbf{v} \\ \mathbf{0}_{1\times 3} & 0 \end{bmatrix} \quad (1)$$

where the wedge operator ($^{\wedge}$) for twist is overloaded.

The exponential map $\exp : \mathfrak{se}(3) \to \mathrm{SE}(3)$ can be used to map a member of $\mathfrak{se}(3)$ around a neighborhood of zero to a member of $\mathrm{SE}(3)$ around a neighborhood of the identity. The logarithm map is the inverse, i.e., $\log : \mathrm{SE}(3) \to \mathfrak{se}(3)$, and $\exp(\log(\mathbf{H})) = \mathbf{H}$, $\mathbf{H} \in \mathrm{SE}(3)$. Now we can define the difference between a transformation $\mathbf{H} \in \mathrm{SE}(3)$ and its estimate with a small perturbation $\tilde{\mathbf{H}} \in \mathrm{SE}(3)$ as [20], [22]:

$$\boldsymbol{\epsilon}^{\wedge} = \log(\mathbf{H}^{-1}\tilde{\mathbf{H}}),$$

where $\boldsymbol{\epsilon}^{\wedge} \in \mathfrak{se}(3)$. To define the norm and covariance of the error term, we exploit the fact that $\mathfrak{se}(3)$ is isomorphic to $\mathbb{R}^6$, i.e., $\boldsymbol{\epsilon}^{\wedge} \mapsto \boldsymbol{\epsilon} \in \mathbb{R}^6$ using the $\vee$ operator. Therefore, we can define the 6×6 covariance matrix conveniently as $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \mathbf{Cov}[\boldsymbol{\epsilon}]$. We use the following adopted simplified notations from [6]:

$$\begin{array}{llll} \mathrm{Exp} : \mathbb{R}^6 & \ni \boldsymbol{\xi} & \to \exp(\boldsymbol{\xi}^{\wedge}) & \in \mathrm{SE}(3) \\ \mathrm{Log} : \mathrm{SE}(3) & \ni \mathbf{H} & \to \log(\mathbf{H})^{\vee} & \in \mathbb{R}^6. \end{array}$$

The adjoint representation of a Lie group is a linear map that captures the non-commutative structure of the group. For $\mathrm{SE}(3)$, the matrix representation of the adjoint map [20] is

$$\mathrm{Ad}_{\mathbf{H}} = \begin{bmatrix} \mathbf{R} & \mathbf{0}_{3\times 3} \\ \mathbf{p}^{\wedge}\mathbf{R} & \mathbf{R} \end{bmatrix}. \quad (2)$$

For any $\mathbf{H} \in \mathrm{SE}(3)$ and $\boldsymbol{\xi} \in \mathfrak{se}(3)$, the adjoint map asserts

$$\mathrm{Ad}_{\mathbf{H}}\boldsymbol{\xi}^{\wedge} = \mathbf{H}\boldsymbol{\xi}^{\wedge}\mathbf{H}^{-1} \Rightarrow \mathrm{Exp}(\mathrm{Ad}_{\mathbf{H}}\boldsymbol{\xi}) = \mathbf{H}\mathrm{Exp}(\boldsymbol{\xi})\mathbf{H}^{-1}.$$

B. Retraction Map and Optimization on Manifold

Given a retraction mapping and the associated manifold, we can optimize over the manifold by iteratively *lifting* the cost function of our optimization problem to the tangent space, solving the re-parameterized problem, and then mapping the updated solution back to the manifold using the retraction [21]. For $\mathrm{SE}(3)$ we use its exponential map as the natural retraction, $\mathcal{R}_{\mathbf{H}}(\delta\boldsymbol{\xi}) = \mathbf{H}\mathrm{Exp}(\delta\boldsymbol{\xi})$, where $\delta\boldsymbol{\xi} \in \mathbb{R}^6$ is the twist defined earlier in (1). Therefore, the retractions on the base and the contact poses are defined as

$$\mathbf{X}_i \leftarrow \mathbf{X}_i\mathrm{Exp}(\delta\mathbf{x}_i) \quad \text{and} \quad \mathbf{C}_i \leftarrow \mathbf{C}_i\mathrm{Exp}(\delta\mathbf{c}_i). \quad (3)$$

These retractions are used to both update the state during optimization and to compute the Jacobians of the residuals.

III. PROBLEM STATEMENT AND FORMULATION

In this section, we formulate the state estimation problem using the factor graph framework where independent sensor

measurements can be incorporated by introducing additional factors based on the associated measurement models. The biped robot, shown in Figure 1, is equipped with a stereo camera, an IMU mounted on the torso, joint encoders, and a series of springs used for binary contact detection. Without loss of generality, we assume the IMU and camera are collocated with the base frame of the robot.

A. State Representation

The state variables (at time t_i) include the 3D body pose, $\mathbf{H}_{\mathrm{WB}}(t) \in \mathrm{SE}(3)$, and velocity, $_{\mathrm{W}}\mathbf{v}_{\mathrm{WB}}(t) \in \mathbb{R}^3$, in the world frame, the 3D contact pose in the world frame, $\mathbf{H}_{\mathrm{WC}}(t) \in \mathrm{SE}(3)$, and the IMU bias, $\mathbf{b}(t) \triangleq \mathrm{vec}\,(\mathbf{b}^g(t), \mathbf{b}^a(t)) \in \mathbb{R}^6$, where $\mathbf{b}^g(t) \in \mathbb{R}^3$ and $\mathbf{b}^a(t) \in \mathbb{R}^3$ are the gyroscope and accelerometer biases, respectively. All together, the state at any time-step i is a tuple as follows:

$$\mathcal{T}_i \triangleq \left(\mathbf{H}_{\mathrm{WB}}(t), _{\mathrm{W}}\mathbf{v}_{\mathrm{WB}}(t), \mathbf{H}_{\mathrm{WC}}(t), \mathbf{b}(t)\right) \triangleq (\mathbf{X}_i, \mathbf{v}_i, \mathbf{C}_i, \mathbf{b}_i) \quad (4)$$

Further, it is convenient to denote the trajectory of the state variables up to time-step k by $\mathcal{X}_k \triangleq \bigcup_{i=1}^{k} \mathcal{T}_i$. Foot slip is the major source of drift in kinematic leg odometry. The inclusion of the contact pose in the state tuple allows for isolating the noise at the contact point.

B. Factor Graph Formulation

Let $\mathcal{L}_{ij} \in \mathrm{SE}(3)$ be a perceptual loop-closure measurement relating poses at time-steps i and j ($j > i$) computed from an independent sensor, e.g. using a point cloud matching algorithm. The forward kinematic measurements at time-step i are denoted by $\mathcal{F}_i$. The IMU and contact sensors provide measurements at higher frequencies. Between any two time-steps i and j, we denote the set of all IMU and contact measurements by $\mathcal{I}_{ij}$ and $\mathcal{C}_{ij}$, respectively. Let $\mathcal{K}_k$ be the index set of time-steps (or keyframes) up to time-step k. We denote the set of all measurements up to time-step k by $\mathcal{Z}_k \triangleq \{\mathcal{L}_{ij}, \mathcal{I}_{ij}, \mathcal{F}_i, \mathcal{C}_{ij}\}_{i,j \in \mathcal{K}_k}$.

By assuming the measurements are conditionally independent and are corrupted by additive zero mean white Gaussian noise, the posterior probability of the *full SLAM* problem can be written as $p(\mathcal{X}_k | \mathcal{Z}_k) \propto p(\mathcal{X}_0) p(\mathcal{Z}_k | \mathcal{X}_k)$, where

$$p(\mathcal{Z}_k | \mathcal{X}_k) = \prod_{i,j \in \mathcal{K}_k} p(\mathcal{L}_{ij} | \mathcal{X}_j) p(\mathcal{I}_{ij} | \mathcal{X}_j) p(\mathcal{F}_i | \mathcal{X}_i) p(\mathcal{C}_{ij} | \mathcal{X}_j).$$

The *Maximum-A-Posteriori* (MAP) estimate of $\mathcal{X}_k$ can be computed by solving the following optimization problem:

$$\underset{\mathcal{X}_k}{\mathrm{minimize}} \;\; -\log p(\mathcal{X}_k | \mathcal{Z}_k)$$

in which due to the noise assumption mentioned earlier is equivalent to the following nonlinear least-squares problem:

$$\underset{\mathcal{X}_k}{\mathrm{minimize}} \; \|\mathbf{r}_0\|_{\hat{\boldsymbol{\Sigma}}_0}^2 + \sum_{i,j \in \mathcal{K}_k} \|\mathbf{r}_{\mathcal{L}_{ij}}\|_{\boldsymbol{\Sigma}_{\mathcal{L}_{ij}}}^2 + \sum_{i,j \in \mathcal{K}_k} \|\mathbf{r}_{\mathcal{I}_{ij}}\|_{\boldsymbol{\Sigma}_{\mathcal{I}_{ij}}}^2$$
$$+ \sum_{i \in \mathcal{K}_k} \|\mathbf{r}_{\mathcal{F}_i}\|_{\boldsymbol{\Sigma}_{\mathcal{F}_i}}^2 + \sum_{i,j \in \mathcal{K}_k} \|\mathbf{r}_{\mathcal{C}_{ij}}\|_{\boldsymbol{\Sigma}_{\mathcal{C}_{ij}}}^2$$

where $\mathbf{r}_0$ and $\boldsymbol{\Sigma}_0$ represent the prior over the initial state and serves to anchor the graph, $\mathbf{r}_{\mathcal{L}_{ij}}$, $\mathbf{r}_{\mathcal{I}_{ij}}$, $\mathbf{r}_{\mathcal{F}_i}$, $\mathbf{r}_{\mathcal{C}_{ij}}$ are the residual terms associated with the loop closure, IMU, forward kinematic, and contact measurements respectively, i.e.
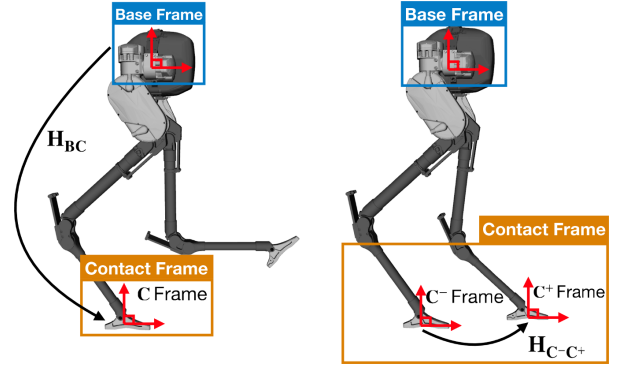

Fig. 2: In this paper, we refer to two separate forward kinematics functions. The pose of the current contact frame relative to the base frame is denoted by $\mathbf{H}_{\mathrm{BC}}$. When the robot has multiple points of contact with the environment, it is possible to transfer this contact from from one frame to another. This transfer of contact is captured by the homogeneous transform $\mathbf{H}_{\mathrm{C^-C^+}}$.

the error between the measured and predicted values given the state, and $\boldsymbol{\Sigma}_{\mathcal{L}_{ij}}$, $\boldsymbol{\Sigma}_{\mathcal{I}_{ij}}$, $\boldsymbol{\Sigma}_{\mathcal{F}_i}$, $\boldsymbol{\Sigma}_{\mathcal{C}_{ij}}$ are the corresponding covariance matrices.

IV. FORWARD KINEMATICS

Forward kinematics refers to the process of computing the relative pose transformation between two frames of a multi-link system. Each individual joint displacement describes how the child link moves with respect to the parent one. This joint displacement can either be an angle (revolute joints) or a distance (prismatic joints).

Let $\boldsymbol{\alpha} \in \mathbb{R}^N$ denote the vector of joint displacements for a general robot. Without loss of generality, we define a *base frame* on the robot, denoted B, that is assumed to be collocated with both the IMU and the camera frames. When the robot is in contact with the static environment, we can also define a *contact frame*, denoted C, on the robot at the point of contact. The homogeneous transformation between the base frame and the contact frame is defined by:

$$\mathbf{H}_{\mathrm{BC}}(\boldsymbol{\alpha}) \triangleq \begin{bmatrix} \mathbf{R}_{\mathrm{BC}}(\boldsymbol{\alpha}) & _{\mathrm{B}}\mathbf{p}_{\mathrm{BC}}(\boldsymbol{\alpha}) \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix}, \quad (5)$$

where $\mathbf{R}_{\mathrm{BC}}(\boldsymbol{\alpha})$ and $_{\mathrm{B}}\mathbf{p}_{\mathrm{BC}}(\boldsymbol{\alpha})$ denote the relative orientation and position of the contact frame with respect to the base frame.

When there are two points in contact with the static environment, it is possible to "transfer" the contact frame from one point to the other (shown in Figure 2). Let C^- denote the old contact frame and C^+ denote the new contact frame. Then, the homogeneous transformation between the old frame and the new frame is defined by:

$$\mathbf{H}_{\mathrm{C^-C^+}}(\boldsymbol{\alpha}) \triangleq \begin{bmatrix} \mathbf{R}_{\mathrm{C^-C^+}}(\boldsymbol{\alpha}) & _{\mathrm{C^-}}\mathbf{p}_{\mathrm{C^-C^+}}(\boldsymbol{\alpha}) \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix}, \quad (6)$$

where $\mathbf{R}_{\mathrm{C^-C^+}}(\boldsymbol{\alpha})$ and $_{\mathrm{C^-}}\mathbf{p}_{\mathrm{C^-C^+}}(\boldsymbol{\alpha})$ denote the relative orientation and position of the new contact frame with respect to the old contact frame.

A. Measurements

We assume the robot's joints are equipped with a set of joint encoders that can measure the joint displacement. These

encoder measurements, $\tilde{\boldsymbol{\alpha}}$, are assumed to be corrupted with Gaussian white noise. This is an explicit measurement coming from physical sensors located on the robot.

$$\tilde{\boldsymbol{\alpha}}(t) = \boldsymbol{\alpha}(t) + \mathbf{w}^\alpha(t), \quad \mathbf{w}^\alpha(t) \sim \mathcal{N}(\mathbf{0}_{N \times 1}, \boldsymbol{\Sigma}^\alpha(t)) \quad (7)$$

The geometric (or manipulator) Jacobian, denoted $\mathbf{J}(\boldsymbol{\alpha})$, provides a method for computing the angular and linear velocity of an end-effector given the vector of joint velocities [19]. In a similar manner, we can use the Jacobian to map incremental angle displacements to changes in the end-effector pose. Let $\mathbf{J}_{\mathrm{BC}}(\boldsymbol{\alpha})$ denote the body manipulator Jacobian of the forward kinematics function (5). Then, the following relationship holds:

$$_\mathrm{C}\boldsymbol{\xi}_{\mathrm{BC}}\, \delta t = \mathbf{J}_{\mathrm{BC}}(\boldsymbol{\alpha})\, \delta\boldsymbol{\alpha} \quad (8)$$

where $\delta\boldsymbol{\alpha}$ and δt are incremental encoder and time quantities and $_\mathrm{C}\boldsymbol{\xi}_{\mathrm{BC}}$ denotes the vector of angular and linear velocities of the contact frame due to $\delta\boldsymbol{\alpha}$ (measured in the contact frame). We can use this relationship (8) to factor out the noise from the forward kinematics equations. Up to a first order approximation, (5) can be factored as:

$$\mathbf{H}_{\mathrm{BC}}(\boldsymbol{\alpha}(t)) \approx \mathbf{H}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}(t))\, \mathrm{Exp}\left(-\mathbf{J}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}(t))\mathbf{w}^\alpha(t)\right). \quad (9)$$

A similar approximation can be found for (6), albeit with a different Jacobian.

$$\mathbf{H}_{\mathrm{C'C^+}}(\boldsymbol{\alpha}(t)) \approx \mathbf{H}_{\mathrm{C'C^+}}(\tilde{\boldsymbol{\alpha}}(t))\, \mathrm{Exp}\left(-\mathbf{J}_{\mathrm{C'C^+}}(\tilde{\boldsymbol{\alpha}}(t))\mathbf{w}^\alpha(t)\right) \quad (10)$$

Remark 1. *In general, the manipulator Jacobian can be derived as spatial or body Jacobian [19], and based on this choice, the noise can appear on the left or right side of the rotation/rigid body transformation, respectively.*

B. Forward Kinematic Factor

The goal of this section is to derive a general *forward kinematic factor* that can be used in the factor graph framework. This will be a unary factor that relates two poses through the forward kinematics equations while accounting for encoder noise. We derive it here for relating the base and contact frames.

The orientation and position of the contact frame with respect to the world frame are given by:

$$\mathbf{H}_{\mathrm{WC}}(t) = \mathbf{H}_{\mathrm{WB}}(t)\, \mathbf{H}_{\mathrm{BC}}(\boldsymbol{\alpha}(t)). \quad (11)$$

Substituting in the state variables at time t_i (4), the forward kinematic equation (5) yields:

$$\mathbf{C}_i = \mathbf{X}_i\, \mathbf{H}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i - \mathbf{w}_i^\alpha). \quad (12)$$

We can now use the first order approximation (9) to factor out the encoder noise to give the following expression:

$$\mathbf{C}_i = \mathbf{X}_i\, \mathbf{H}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i)\, \mathrm{Exp}\left(-\mathbf{J}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i)\mathbf{w}_i^\alpha\right). \quad (13)$$

Defining the zero-mean white Gaussian forward kinematic noise term, $\delta\mathbf{c}_i \triangleq \mathbf{J}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i)\mathbf{w}_i^\alpha$, allows us to write out the *forward kinematic measurement model*:

$$\mathbf{H}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i) = \mathbf{X}_i^{-1}\, \mathbf{C}_i\, \mathrm{Exp}(\delta\mathbf{c}_i), \quad (14)$$

where the forward kinematic noise is characterized by $\delta\mathbf{c}_i \sim \mathcal{N}(\mathbf{0}_{6 \times 1}, \boldsymbol{\Sigma}_{\mathcal{F}_i})$. The residual error is defined in the tangent space and can be written as:

$$\mathbf{r}_{\mathcal{F}i} = \mathrm{Log}\left(\mathbf{C}_i^{-1}\, \mathbf{X}_i\, \mathbf{H}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i)\right). \quad (15)$$

The covariance is computed through the following linear transformation:

$$\boldsymbol{\Sigma}_{\mathcal{F}_i} = \mathbf{J}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i)\, \boldsymbol{\Sigma}_i^\alpha\, \mathbf{J}_{\mathrm{BC}}^\top(\tilde{\boldsymbol{\alpha}}_i), \quad (16)$$

where $\mathbf{J}_{\mathrm{BC}}(\tilde{\boldsymbol{\alpha}}_i)$ is the body manipulator Jacobian evaluated at the current encoder measurements and $\boldsymbol{\Sigma}_i^\alpha$ denotes the encoder covariance matrix at time t_i.

V. HYBRID CONTACT PREINTEGRATION

A continuous hybrid dynamical system, $\mathcal{H}$, can be defined with a continuous dynamics function, $f(\cdot)$, a discrete transition map, $\Delta(\cdot)$, and a switching surface, $\mathcal{S}$ [23]. Trajectories of the hybrid dynamical system evolve according to the continuous dynamics, until the switching surface it hit. At those moments, the state gets mapped through the discrete transition map, after which the trajectory continues according to the continuous dynamics again. The general form of this system can be expressed as follows:

$$\mathcal{H} : \begin{cases} \dot{x}(t) = f(x, t) & (x^-, t^-) \notin \mathcal{S} \\ x^+ = \Delta(x^-) & (x^-, t^-) \in \mathcal{S}. \end{cases} \quad (17)$$

As long as the number of contact points is greater than or equal to one, the switching contact frame dynamics can be modeled as a hybrid system:

$$\mathcal{H} : \begin{cases} \dot{\mathbf{H}}_{\mathrm{WC}}(t) = \mathbf{H}_{\mathrm{WC}}(t)\, (_\mathrm{C}\boldsymbol{\xi}_{\mathrm{WC}}(t))^\wedge & t^- \notin \mathcal{S} \\ \mathbf{H}_{\mathrm{WC^+}} = \mathbf{H}_{\mathrm{WC^-}}\, \mathbf{H}_{\mathrm{C^-C^+}} & t^- \in \mathcal{S}, \end{cases} \quad (18)$$

where the switching surface, $\mathcal{S}$ is simply modeled as the set of all times where contact is switched from one frame to another. Since the sensor measurements are coming in at discrete time-steps, we perform Euler integration from time t to $t+\Delta t$ to discretize the continuous hybrid contact dynamics (18), forming the following discrete hybrid system:

$$\mathcal{H} : \begin{cases} \mathbf{H}_{\mathrm{WC}}(t + \Delta t) = \mathbf{H}_{\mathrm{WC}}(t)\, \mathrm{Exp}\left(_\mathrm{C}\boldsymbol{\xi}_{\mathrm{WC}}(t)\Delta t\right) & t^- \notin \mathcal{S} \\ \mathbf{H}_{\mathrm{WC^+}} = \mathbf{H}_{\mathrm{WC^-}}\, \mathbf{H}_{\mathrm{C^-C^+}} & t^- \in \mathcal{S}. \end{cases} \quad (19)$$

Physically, the continuous dynamics function, f, describes how a single contact frame moves over time while contact is maintained. When a new contact frame is detected, the new contact pose can be computed by applying the transition map, $\Delta(\cdot)$, which describes the homogeneous transformation between the old and new contact frames.

A. Measurements

The angular and linear velocity of the contact point is an implicit measurement that is *inferred* through a binary contact sensor: specifically, when this sensor indicates contact, the pose of the contact frame is *assumed to remain fixed with respect to the world frame*, i.e. the measured velocities are zero. In order to accommodate potential contact slip,
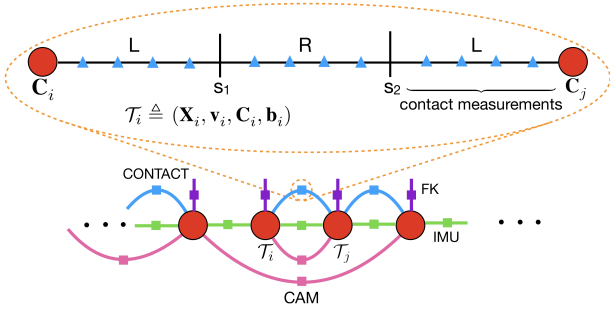
Fig. 3: In the factor graph framework, the robot's state along a discretized trajectory denoted by red circles. Each independent sensor measurement is a factor denoted by lines that constrains the state at separate time-steps. The proposed hybrid contact factor (shown on top) allows preintegration of high-frequency contact data through an arbitrary number of contact switches. In this example, there are two contact switches, where the robot moves from left-stance (L) to right-stance (R), then back to left stance.

this measured twist is modeled to be corrupted with white Gaussian noise, namely

$$_\mathrm{C}\tilde{\boldsymbol{\xi}}_\mathrm{WC}(t) = \mathbf{0}_{6\times1} = {_\mathrm{C}}\boldsymbol{\xi}_\mathrm{WC}(t) + \mathbf{w}^\xi(t), \quad \mathbf{w}^\xi(t) \sim \mathcal{N}(\mathbf{0}_{6\times1}, \boldsymbol{\Sigma}^\xi(t)).$$

Using the state variables (4), forward kinematics definition (6), encoder measurements (7), and contact measurements, the hybrid contact dynamics (19) can be written as:

$$\mathcal{H}: \begin{cases} \mathbf{C}_{k+1} = \mathbf{C}_k \operatorname{Exp}\left(-\mathbf{w}_k^{\xi d}\Delta t\right) & t_k^- \notin \mathcal{S} \\ \mathbf{C}^+ = \mathbf{C}^- \ \mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_k - \mathbf{w}_k^\alpha) & t_k^- \in \mathcal{S} \end{cases} \quad (20)$$

where $\mathbf{w}_k^{vd}$, the discrete time contact noise, is computed using the sample time, $\mathbf{Cov}[\mathbf{w}^{\xi d}(t)] = \dfrac{1}{\Delta t}\mathbf{Cov}[\mathbf{w}^\xi(t)]$.

Remark 2. $\boldsymbol{H}_{\mathrm{C}^-\mathrm{C}^+}(\boldsymbol{\alpha}_k)$ *will depend on the specific contact frames,* C^- *and* C^+, *at time* t_k. *For example, the forward kinematics function to switch from left foot contact to a right foot contact will be different than the one used to switch from right foot to left foot.*

B. Preintegrating Contact Pose

The goal of this section is to formulate a general *hybrid preintegrated contact factor* which relates the contact pose at t_i to the contact pose at t_j. The hybrid nature of this factor comes from the potential switching of contact that occurs naturally in legged locomotion. Following the work on IMU preintegration theory [6], [24], we preintegrate the high-frequency contact measurements to prevent unnecessary computation allowing efficient implementation of the factor.

Let $\mathcal{S}$ denote the sequence of all time indices associated with contact switches, so that each $s_i \in \mathcal{S}$ represents a single time index where a contact switch occurs. To integrate the hybrid contact dynamics model (20) through these contact switches, we integrate the up to the next contact switch time t_{s_i}, apply the transition map $\Delta(\cdot)$, and continue the integration. For the following derivations, a concrete example (depicted in Figure 3) is used to make the derivation easier to follow. We later extend the derivation to an arbitrary number of contact switches. For the example shown in Figure 3, there are two contact switches between times t_i and t_j, i.e.,

$|\mathcal{S}| = 2$. Integrating the discrete hybrid contact model (20) from t_i to t_j yields:

$$\mathbf{C}_j = \mathbf{C}_i \left(\prod_{k=i}^{s_1-1} \operatorname{Exp}(-\mathbf{w}_k^{\xi d}\Delta t)\right) \mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_1} - \mathbf{w}_{s_1}^\alpha)$$

$$\left(\prod_{k=s_1}^{s_2-1} \operatorname{Exp}(-\mathbf{w}_k^{\xi d}\Delta t)\right) \mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_2} - \mathbf{w}_{s_2}^\alpha) \left(\prod_{k=s_2}^{j-1} \operatorname{Exp}(-\mathbf{w}_k^{\xi d}\Delta t)\right).$$

All noise terms can be shifted to the right by using (10) to factor the noise from the forward kinematics term and the adjoint relation (2) to shift the measured kinematics to the left.

$$\mathbf{C}_j = \mathbf{C}_i \ \mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_1}) \ \mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_2})$$
$$\left(\prod_{k=i}^{s_1-1} \operatorname{Exp}(-\operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_2})}\operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_1})}\mathbf{w}_k^{\xi d}\Delta t)\right)$$
$$\operatorname{Exp}(-\operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_2})}\mathbf{J}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_1})\mathbf{w}_{s_1}^\alpha) \quad (21)$$
$$\left(\prod_{k=s_1}^{s_2-1} \operatorname{Exp}(-\operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_2})}\mathbf{w}_k^{\xi d}\Delta t)\right)$$
$$\operatorname{Exp}(-\mathbf{J}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_2})\mathbf{w}_{s_2}^\alpha) \left(\prod_{k=s_2}^{j-1} \operatorname{Exp}(-\mathbf{w}_k^{\xi d}\Delta t)\right)$$

After multiplying both sides by $\mathbf{C}_i^{-1}$, we arrive at a relative pose expression that is independent of states $\mathcal{T}_i$ and $\mathcal{T}_j$.

$$\Delta\mathbf{C}_{ij} \triangleq \mathbf{C}_i^{-1}\mathbf{C}_j = \Delta\tilde{\mathbf{C}}_{ij}\operatorname{Exp}(-\delta\mathbf{c}_{ij}) \quad (22)$$

where $\Delta\tilde{\mathbf{C}}_{ij} \triangleq \prod_{n=1}^{|\mathcal{S}|} \mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_n})$ represents the *hybrid preintegrated contact measurement*, and $\operatorname{Exp}(-\delta\mathbf{c}_{ij})$ groups all of the noise terms together. This noise term is a product of multiple small rigid body transformations. Therefore, it can be approximated as a summation in the tangent space through iterative use of the Baker-Campbell-Hausdorff (BCH) formula [20] (while keeping only the first order terms).

$$\delta\mathbf{c}_{ij} \approx \sum_{k=i}^{s_1-1} \operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_2})}\operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_1})}\mathbf{w}_k^{\xi d}\Delta t$$
$$+ \operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_2})}\mathbf{J}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_1})\mathbf{w}_{s_1}^\alpha + \sum_{k=s_1}^{s_2-1} \operatorname{Ad}_{\mathbf{H}_{\mathrm{C}^-\mathrm{C}^+}^{-1}(\tilde{\boldsymbol{\alpha}}_{s_2})}\mathbf{w}_k^{\xi d}\Delta t$$
$$+ \mathbf{J}_{\mathrm{C}^-\mathrm{C}^+}(\tilde{\boldsymbol{\alpha}}_{s_2})\mathbf{w}_{s_2}^\alpha + \sum_{k=s_2}^{j-1} \mathbf{w}_k^{\xi d}\Delta t$$

The *hybrid preintegrated contact noise*, $\delta\mathbf{c}_{ij}$, is a summation of zero-mean Gaussian terms, and is therefore also zero-mean and Gaussian. It is possible to generalize this noise expression to an arbitrary number of contact switches, however, it becomes much simpler to do so when looking at the iterative propagation form in the following section.

C. Iterative Propagation

It is possible to write both the preintegrated contact measurements, $\Delta\tilde{\mathbf{C}}_{ij}$, and the preintegrated contact noise, $\delta\mathbf{c}_{ij}$, in iterative update forms. This allows the terms to be updated as contact and encoder measurements are coming in. In addition, this form simplifies the expressions and allows for
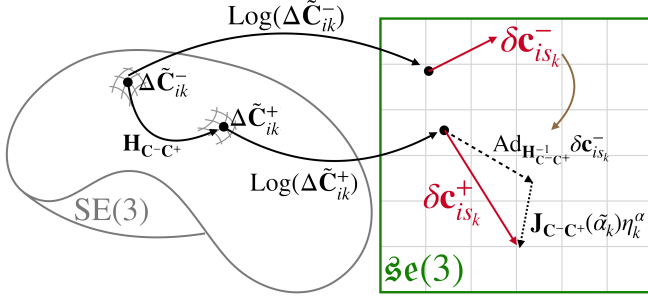
Fig. 4: When a contact switch occurs, the relative contact pose, $\Delta\tilde{\mathbf{C}}_{ik}^-$, gets mapped from one point in SE(3) to another point, $\Delta\tilde{\mathbf{C}}_{ik}^+$, on the manifold. The contact noise, $\delta\mathbf{c}_{ik}^-$, is represented in the tangent space, $\mathfrak{se}(3)$, and is mapped from the tangent space of $\Delta\tilde{\mathbf{C}}_{ik}^-$ to the tangent space of $\Delta\tilde{\mathbf{C}}_{ik}^+$ through the use of the adjoint map of the forward kinematics transformation, $\mathbf{H}_{\mathrm{C^-C^+}}$. However, due to noisy encoders, an addition noise term (computed using the manipulator Jacobian) has to be added to compute $\delta\mathbf{c}_{ik}^+$.

the covariance to be conveniently computed. The following proposition generalizes the hybrid preintegrated contact pose and noise iterative propagation to an arbitrary number of contact switches. The proof is given in the supplementary material.

Proposition 1 (Iterative Propagation of Hybrid Contact Process [25]). *Between any two time-steps t_i and t_j such that $j > i$, starting with $\Delta\tilde{\mathbf{C}}_{ii} = \mathbf{I}_4$ and $\delta\mathbf{c}_{ij} = \mathbf{0}_{3\times 1}$, the hybrid preintegrated contact measurement, $\Delta\tilde{\mathbf{C}}_{ij}$, and noise, $\delta\mathbf{c}_{ij}$, for an arbitrary number of contact switches can be computed iteratively using the following hybrid systems:*

$$\tilde{\mathcal{H}} : \begin{cases} \Delta\tilde{\mathbf{C}}_{ik+1} = \Delta\tilde{\mathbf{C}}_{ik} & t_k^- \notin \mathcal{S} \\ \Delta\tilde{\mathbf{C}}_{ik}^+ = \Delta\tilde{\mathbf{C}}_{ik}^- \, \mathbf{H}_{\mathrm{C^-C^+}}(\tilde{\boldsymbol{\alpha}}_k) & t_k^- \in \mathcal{S} \end{cases} \quad (23)$$

$$\delta\mathcal{H} : \begin{cases} \delta\mathbf{c}_{ik+1} = \delta\mathbf{c}_{ik} + \mathbf{w}_k^{\xi d}\Delta t & t_k^- \notin \mathcal{S} \\ \delta\mathbf{c}_{ik}^+ = \mathrm{Ad}_{\mathbf{H}_{\mathrm{C^-C^+}}^{-1}(\tilde{\boldsymbol{\alpha}}_k)}\delta\mathbf{c}_{ik}^- + \mathbf{J}_{\mathrm{C^-C^+}}(\tilde{\boldsymbol{\alpha}}_k)\mathbf{w}_k^\alpha & t_k^- \in \mathcal{S} \end{cases}$$
$$(24)$$

An abstract representation of these hybrid systems is shown in Figure 4.

D. Rigid Contact Factor

The relative contact pose expression (22) can be used to define the *hybrid preintegrated contact measurement model*:

$$\Delta\tilde{\mathbf{C}}_{ij} = \mathbf{C}_i^{-1}\mathbf{C}_j\mathrm{Exp}(\delta\mathbf{c}_{ij}), \quad (25)$$

where the preintegrated contact noise is characterized by $\delta\mathbf{c}_{ij} \sim \mathcal{N}(\mathbf{0}_{6\times 1}, \boldsymbol{\Sigma}_{\mathcal{C}_{ij}})$. In the factor graph framework, this hybrid preintegrated contact model represents a binary factor that relates the contact frame pose over consecutive time steps. The residual error is defined in the tangent space, and can be written as:

$$\mathbf{r}_{\mathcal{C}_{ij}} = \mathrm{Log}\left(\mathbf{C}_j^{-1}\mathbf{C}_i\Delta\tilde{\mathbf{C}}_{ij}(\tilde{\boldsymbol{\alpha}}_i)\right). \quad (26)$$

The covariance is computed using the hybrid contact noise model (24), starting with $\boldsymbol{\Sigma}_{\mathcal{C}_{ii}} = \mathbf{0}_{6\times 6}$:

$$\begin{cases} \boldsymbol{\Sigma}_{\mathcal{C}_{ik+1}} = \boldsymbol{\Sigma}_{\mathcal{C}_{ik}} + \boldsymbol{\Sigma}_k^\xi \Delta t & t_k^- \notin \mathcal{S} \\ \boldsymbol{\Sigma}_{\mathcal{C}_{ik}}^+ = \mathrm{Ad}_{\mathbf{H}_{\mathrm{C^-C^+}}^{-1}(\tilde{\boldsymbol{\alpha}}_k)}\boldsymbol{\Sigma}_{\mathcal{C}_{ik}}^- \, \mathrm{Ad}_{\mathbf{H}_{\mathrm{C^-C^+}}^{-1}(\tilde{\boldsymbol{\alpha}}_k)}^\mathsf{T} & t_k^- \in \mathcal{S} \\ \qquad + \mathbf{J}_{\mathrm{C^-C^+}}(\tilde{\boldsymbol{\alpha}}_k)\, \boldsymbol{\Sigma}_k^\alpha \, \mathbf{J}_{\mathrm{C^-C^+}}^\mathsf{T}(\tilde{\boldsymbol{\alpha}}_k). \end{cases}$$
$$(27)$$

VI. Experimental Results

We now present experimental evaluations of the proposed factors. In the first experiment, we compare three odometry systems composed by Visual-Inertial-Contact (VIC), Inertial-Contact (IC), and Visual-Inertial (VI) factors. Whenever the contact factor is used, it is assumed that forward kinematic factor is also available; therefore, it is not explicitly mentioned for brevity. In the second experiment, we study the effect of losing visual data for a period of time to see how the contact factor can constrain the graph in the absence of a reliable vision system. In the third experiment, we will evaluate the use of *terrain factors*, where loop-closures can be added to the graph through contact frame poses.

A. Experimental Setup

All experiments are done on a Cassie-series robot designed by Agility Robotics, which has 20 degrees of freedom, 10 actuators, joint encoders, and a VN-100 IMU. A Multisense S7 stereo camera, which contains a separate IMU, is mounted on the top of the Cassie robot as shown in Fig. 1. The robot also has four springs (two on each leg) that can be used as binary contact sensors by thresholding the spring deflection measurements. The Cassie robot has two computers which run MATLAB Simulink Real-Time (for the controller) and Ubuntu (for the estimator) respectively. We use the Robot Operating System (ROS) [26] with the User Datagram Protocol (UDP) to communicate sensor data between the two computers. We also integrate the time synchronization algorithm in [27] into our system to ensure all sensory data is synchronized.

The motion capture system developed by Vicon is used as a proxy for ground truth trajectories. The setup consists of 17 motion capture cameras with four markers attached to the robot to track the IMU pose. The dataset contains the stereo images (20 Hz) and IMU data (750 Hz) from the Multisense S7 camera as well as the joint encoders and IMU data from the Cassie robot (at 400 Hz each).

The proposed kinematic and contact factors were implemented using the GTSAM 4.0 library [4]. We utilized the built-in preintegrated IMU factor [6] with the iSAM2 [3] incremental solver. For visual odometry, we used the semi-direct visual odometry library (SVO 2.0 [28]). The Multisense camera recorded synchronized stereo images at 20 Hz and IMU measurements in about 750 Hz. SVO processes those measurements in real-time and outputs the pose of the left camera in a fixed world frame, $\mathbf{X}_i$, for the current time-step i. The relative transformation of the camera from time-step i to j can be obtained using $\Delta\mathbf{X}_{ij} = \mathbf{X}_i^{-1}\mathbf{X}_j$. We selected keyframes approximately every 0.25 seconds.

B. First Experiment: Odometry Comparison

In this experiment, we had Cassie stand in place for about 15 seconds, then slowly walk forwards and backwards along the length of the lab for approximately 45 seconds. The resulting data was used to compare the odometry performance (processed off-line) of different combinations of factors. The results are shown in Figure 5. The odometry estimate from
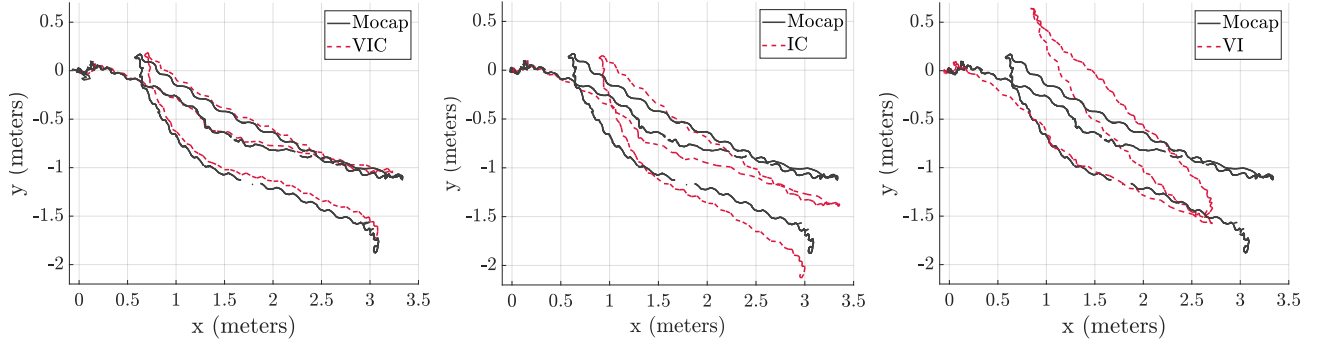
Fig. 5: The odometry results from a 60 second walking experiment using a Cassie-series robot. The Visual-Inertial-Contact (VIC) odometry outperformed both the Inertial-Contact (IC), and the Visual-Inertial (VI) odometry. "Ground-truth" data was collected from a Vicon motion capture system. It is important to note that no loop-closures are being performed, which helps to explain the relatively poor odometry from VI. The video of this experiment is provided at https://youtu.be/WDPhdl5g2MQ.
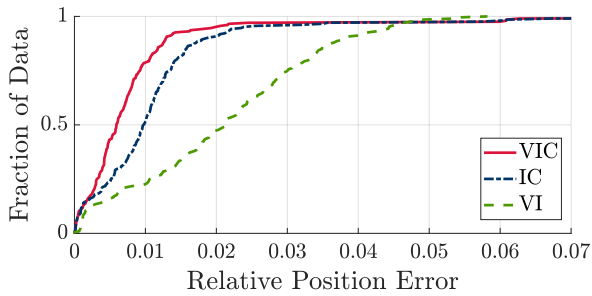

Fig. 6: The Cumulative Distribution Function (CDF) of the relative position error provides a way to analyze the drift in the odometry estimates from various combinations of factors. The fraction of data corresponding to small relative position errors (low-drift odometry) is the larger for Visual-Inertial-Contact (VIC) odometry than for Inertial-Contact (IC) or Visual-Inertial (VI) odometry.
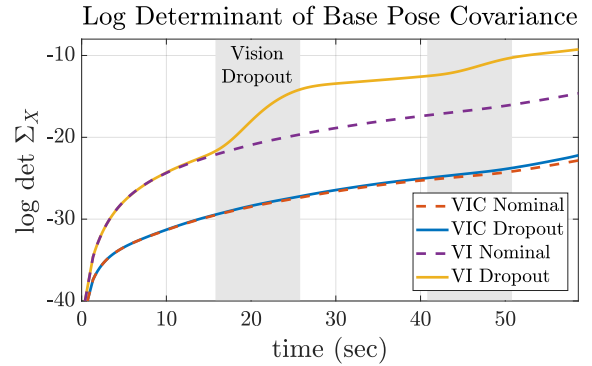

Fig. 7: When vision data is lost, the covariance of the robot's base pose sharply grows for VI odometry due to the lack of additional measurements to constrain the graph. In contrast, during "vision dropout" periods, the additional contact factors allows the covariance estimate from VIC to remains close to the nominal case.

VIC, as expected, outperforms all other combinations of factors. The Cumulative Distribution Function (CDF) of the relative position error is shown in Figure 6. The relative position CDF provides a method for analyzing the drift of an odometry estimate.

From Figure 6, it can be seen that VIC has the highest fraction of data corresponding to smaller relative position errors. This means VIC has lowest drift among all odometry systems. When the robot is walking, the hard impacts cause significant camera shake which leads to motion blur in the images. In addition, the lab environment seemingly lacked numerous quality features. These effects help to explain the relatively poor VI odometry performance.

C. Second Experiment: Vision Dropout

One of the main benefits from including the forward kinematic and contact factors is that the state estimator can be more robust to failure of the vision system. In this experiment, we simulate the effects of "vision dropout" by simply ignoring SVO visual odometry data for two 10-second periods of the experimental data described in the previous section. In other words, during a "vision dropout" period, VIC odometry reduces to IC odometry, and VI odometry

reduces down to inertial (I) odometry. Figure 7 shows the log determinant of the base pose covariance for VIC and VI for these "vision dropout" experiments. The larger the log determinant, the more uncertain the estimator is about the robot's base pose. During the "vision dropout" periods, uncertainty grows for VI odometry. This sharp covariance growth is due to the lack of additional sensor measurements to add into the factor graph. In contrast, the covariance growth for VIC is hardly affected over the same dropout periods.

D. Third Experiment: Terrain Factors for Loop-Closure

Another benefit of adding the proposed contact factors comes from the addition of the contact frame poses into the robot's state. With these new state variables, it becomes simple to place additional constraints that relate the contact pose to a prior map. We test this idea on the collected experiment data by recognizing that the ground was relatively flat in the laboratory. This "zero-height" elevation data serves as our prior map. Figure 8 shows how adding this trivial constraint can reduce position drift in the z-direction. This experiment
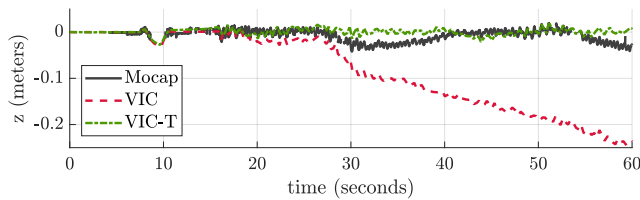
Fig. 8: Since the contact pose is now part of the estimated state, it is possible to add "terrain factors" that relate this contact pose to a prior map. Adding the simple constraint that the contact frame z-translation is zero (VIC-T) improves the drift in the z-direction when compared to the nominal Visual-Inertial-Contact (VIC) case.

simply serves to illustrate the potential for "terrain factors", as the state estimate could be further improved if the robot was actually mapping out the terrain; there was actually a slight slope in the lab (as shown in the motion capture data).

VII. CONCLUSION

We developed a novel method for preintegrating contact information through an arbitrary number of contact switches using a hybrid preintegrated contact factor. The proposed approach was verified through experimental evaluations using a Cassie-series robot where a motion capture system is used as a proxy for ground truth data. Our results indicate that the fusion of contact information with IMU and vision data provide a reliable odometry system for legged robot. Furthermore, we showed that the developed visual-inertial-contact odometry system is robust to occasional vision system failures.

In the future, we plan to incorporate loop closure constraints into our factor graph framework to further improve state estimation, paving the way for long-term mapping on legged robots. We also plan to further investigate the potential utility of "terrain factors" to allow the robot's state to be corrected though detected contact on an estimated map.

ACKNOWLEDGMENT

REFERENCES

[1] M. Bloesch, M. Hutter, M. A. Hoepflinger, S. Leutenegger, C. Gehring, C. D. Remy, and R. Siegwart, "State estimation for legged robots-consistent fusion of leg kinematics and IMU," in *Robotics: Science and Systems*, Berlin, Germany, June 2013.

[2] R. Hartley, M. G. Jadidi, J. Grizzle, and R. M. Eustice, "Contact-aided invariant extended Kalman filtering for legged robot state estimation," in *Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018.

[3] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The Int. J. Robot. Res.*, vol. 31, no. 2, pp. 216–235, 2012.

[4] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012.

[5] L. Carlone, Z. Kira, C. Beall, V. Indelman, and F. Dellaert, "Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors," in *Proc. IEEE Int. Conf. Robot. Automat.* IEEE, 2014, pp. 4290–4297.

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

[7] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *The Int. J. Robot. Res.*, vol. 23, no. 7-8, pp. 693–716, 2004.

[8] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly sparse delayed-state filters for view-based SLAM," *IEEE Trans. Robot.*, vol. 22, no. 6, pp. 1100–1114, 2006.

[9] N. Rotella, M. Bloesch, L. Righetti, and S. Schaal, "State estimation for a humanoid robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2014, pp. 952–958.

[10] M. Benallegue and F. Lamiraux, "Estimation and stabilization of humanoid flexibility deformation using only inertial measurement units and contact information," *Int. J. Humanoid Robot.*, vol. 12, no. 03, p. 1550025, 2015.

[11] J. Eljaik, N. Kuppuswamy, and F. Nori, "Multimodal sensor fusion for foot state estimation in bipedal robots using the extended kalman filter," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2015, pp. 2698–2704.

[12] S. Kuindersma, R. Deits, M. Fallon, A. Valenzuela, H. Dai, F. Permenter, T. Koolen, P. Marion, and R. Tedrake, "Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot," *Auton. Robot*, vol. 40, no. 3, pp. 429–455, 2016.

[13] P. Fankhauser, M. Bloesch, C. Gehring, M. Hutter, and R. Siegwart, "Robot-centric elevation mapping with uncertainty estimates," in *International Conference on Climbing and Walking Robots*, 2014, pp. 433–440.

[14] M. A. Bloesch, "State estimation for legged robots–kinematics, inertial sensing, and computer vision," Ph.D. dissertation, 2017.

[15] M. F. Fallon, M. Antone, N. Roy, and S. Teller, "Drift-free humanoid state estimation fusing kinematic, inertial and lidar sensing," in *IEEE-RAS Int. Conf. on Humanoid Robots.* IEEE, 2014, pp. 112–119.

[16] S. Nobili, M. Camurri, V. Barasuol, M. Focchi, D. G. Caldwell, C. Semini, and M. Fallon, "Heterogeneous sensor fusion for accurate state estimation of dynamic legged robots," in *Robotics: Science and Systems*, 2017.

[17] G. P. Roston and E. P. Krotkov, "Dead reckoning navigation for walking robots," Carnegie-Mellon University, Pittsburgh, PA, Robotics Institute, Tech. Rep., 1991.

[18] R. Hartley, J. Mangelson, L. Gan, M. Ghaffari Jadidi, J. M. Walls, R. M. Eustice, and J. W. Grizzle, "Legged robot state-estimation through combined forward kinematic and preintegrated contact factors," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 4422–4429.

[19] R. M. Murray, Z. Li, S. S. Sastry, and S. S. Sastry, *A mathematical introduction to robotic manipulation.* CRC press, 1994.

[20] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications.* Springer Science & Business Media, 2011.

[21] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds.* Princeton University Press, 2009.

[22] T. D. Barfoot and P. T. Furgale, "Associating uncertainty with three-dimensional poses for use in estimation problems," *IEEE Trans. Robot.*, vol. 30, no. 3, pp. 679–693, 2014.

[23] E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, and B. Morris, *Feedback control of dynamic bipedal robot locomotion.* CRC press, 2007.

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

[25] R. Hartley, M. Ghaffari Jadidi, L. Gan, J.-K. Huang, J. W. Grizzle, and R. M. Eustice, "Supplementary material: Hybrid contact preintegration for visual-inertial-contact state estimation using factor graphs," University of Michigan, Tech. Rep., July 2018. [Online]. Available: https://arxiv.org/abs/1803.07531/anc/supplementary-material.pdf

[26] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source Robot Operating System," in *ICRA workshop on open source software*, vol. 3, no. 3.2, 2009, p. 5.

[27] E. Olson, "A passive solution to the sensor synchronization problem," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, October 2010.

[28] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect visual odometry for monocular and multicamera systems," *IEEE Trans. Robot.*, vol. 33, no. 2, pp. 249–265, 2017.