International Journal of Control, Automation, and Systems VV(X) (YYYY) 1-10
http://dx.doi.org/10.1007/s12555-xxx-xxxx-x
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

VALINOR: a Lightweight Leg Inertial Odometry for Humanoid Robots

Arnaud Demont*, Mehdi Benallegue, Thomas Duvinage, and Abdelaziz Benallegue

Abstract: This article presents VALINOR (Velocity-Aided Leg Inertial Nonlinear Odometry and Registration), a method for Leg-Inertial odometry for humanoid robots addressing the challenge of lightweight yet accurate and certifiable state estimation. VALINOR associates Leg odometry with the Tilt Observer, a computationally efficient complementary filter, which provides accurate estimates of the IMU's tilt and linear velocity with strong mathematical convergence guarantees. We introduce an axis-agnostic method for the fusion of the Leg odometry's yaw with the Tilt Observer's tilt estimate. We argue that this method is less arbitrary and more mathematically sound than those based on other orientation representations, especially on Euler angles. We present an evaluation of the proposed estimator through real-world data on two humanoid robots. We show that, while being 7.5 times faster than the state-of-the-art method used for comparison, VALINOR improves tilt estimation by over 25%, making it a well-suited feedback for balance and walking controllers.

Keywords: Lightweight state estimation, Humanoid robots, Proprioceptive odometry, Tilt estimation, Balancing.

1. INTRODUCTION

Humanoid robot navigation in real-world environments remains a core challenge in robotics. For robots to be deployed at scale in society and industry, whether in healthcare facilities or warehouses, they must operate robustly in dynamic, unstructured, and human-centric environments [1]. This requires strong guarantees on their ability to plan motion, and execute actions in real time. In comparison to other legged robots, the control of humanoids is particularly difficult due to their complex underactuated dynamics controlled through contact with the environment [2]. Their control must therefore be as theoretically grounded as possible, in order to ensure safe, predictable, and certifiable behavior [3].

In the meantime, methods for motion generation continue to gain in complexity and capability. Notably, Model Predictive Control (MPC) [4–6], reinforcement learning-based controllers [7, 8], and foundation models [9, 10] are becoming more prevalent due to their ability to handle long-horizon planning, complex dynamics, and to their versatility. However, these advances come at the cost of increased computational complexity, running them in real-time on resource-limited platforms like embedded systems thus becomes challenging [11–14]. Since the control part is getting heavier, one practical solution is to reduce the computational load of other parts of the pipeline. The state estimation part is one of the cornerstones of the pipeline, since other components depend on its output and often have to wait for it before executing, and thus cannot run in parallel. It is therefore especially critical to try and reduce the computation time of the state estimation part.

State estimation for legged robots has also known notable improvements in the past years, the main trend aiming towards improving its global accuracy by incorporating more and more advanced sensors, especially exteroceptive ones, such as cameras, LiDARs, etc. [15–17]. Such methods show remarkable accuracies. However, they also tend to be more and more computationally expensive (e.g. the order of the millisecond for [17]).

For this reason, there is still a need for lightweight state estimation methods, which rather than adding new sensors, focus on improving the estimation of key variables at high frequency using the proprioceptive sensors readily available on the robot. The most common sensors for humanoids are joint encoders, inertial measurement units (IMUs) and contact force sensors such as FSR or force/torque sensors. IMUs are able to measure angular velocities and linear accelerations, including the gravitational one. It can be used to estimate the vertical direction, which we call *tilt*. This is done often under the assumption that accelerations are negligible compared to the gravitational acceleration (e.g. [18]). Unfortunately, this assumption cannot hold in the case of humanoids, espe-

Arnaud Demont is with the CNRS-AIST JRL (Joint Robotics Laboratory), IRL, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8560 Japan, Université Paris-Saclay, 91190 Gif-sur-Yvette, France, and Laboratoire d'Ingénierie des Systèmes de Versailles, 78140 Vélizy, France (email: arnaud.demont@aist.go.jp). Mehdi Benallegue and Thomas Duvinage are with the CNRS-AIST JRL (Joint Robotics Laboratory), IRL, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8560 Japan (emails: mehdi.benallegue@aist.go.jp, thomas.duvinage@aist.go.jp). Abdelaziz Benallegue is with the CNRS-AIST JRL (Joint Robotics Laboratory), IRL, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8560 Japan, and Laboratoire d'Ingénierie des Systèmes de Versailles, 78140 Vélizy, France (email: abdelaziz.benallegue@uvsq.fr).
* Corresponding author.

cially with impacts and dynamic motions, which significantly degrades the estimation. IMUs signals can also be integrated to provide estimations of the positions and orientation in which is called inertial odometry, but if done for extended periods of time this would produce fast drifts because of noise and initial estimation errors.

Both issues, high accelerations and odometry drifts can be resolved thanks to a single consideration: legged robots are connected to a stable environment through their contacts. Meaning that we consider contacts to be fixed points in the environment, which provides a partial measurement of the robot's kinematics in the world using joint encoders. This consideration only makes the tilt fully observable, however dynamic the motion is [19]. The anchor with the environment and the kinematic data can be used also to perform legged odometry (also called "Leg" odometry) using the transformations between successive contacts to rebuild the world-trajectory [20]. This technique can be merged with IMU data and correct for the drift [21–23]. This is called leg-inertial odometry.

Violations of slippage-free contact assumption can degrade the odometry performance [24, 25]. Work has thus been done to address the problem of contact slippage, for example by discarding unreliable contacts [24–26], or by attempting to correct the contact reference position [21, 27, 28]. Despite that, the position and yaw orientation of the robot in the world frame remain unobservable [21]. Nevertheless, the methods exploiting these contacts have proven to be more than enough in concrete use-cases when the pose needs to be reliable only locally, like for remotely controlled robots [29] or when task planning is also made in the robot's local environment [30], which means that drifts due to slippage remain limited for these use-cases.

In this work we propose VALINOR (Velocity-Aided Leg Inertial Nonlinear Odometry and Registration), a method for Leg-Inertial odometry for legged robots, which relies on a highly accurate tilt estimate provided by a nonlinear complementary filter [19]. This complementary filter allows for a much faster computation than the conventionally used Kalman Filter (e.g. [27]), and offers mathematical convergence guarantees on the tilt and linear velocity estimate, aligning with our objective of lightweight yet accurate and certifiable state estimation. The main contributions introduced in this paper are as follows:

- The proposed state estimator is 7.5 times faster than the start-of-the-art method for proprioceptive odometry, while displaying on par odometry accuracy, and even improving significantly tilt estimation.
- We present the axis-agnostic tilt and yaw fusion, which we show to be more mathematically sound than that based on Euler angle representations.
- We evaluate the proposed method on real experiments with two different humanoid robots.
- The code of the Tilt Observer is open source[1], as well as a ROS wrapper[2].

2. GENERAL NOTATIONS

- The general notation for kinematic variables is ${}^{1}\bigcirc_{2}$, expressing the kinematics of the frame 2 in the frame 1. To simplify the notation, kinematics in the world frame are written without the $\mathcal{W}$ symbol: ${}^{\mathcal{W}}\bigcirc_{2} = \bigcirc_{2}$.
- We note $\hat{\boldsymbol{x}}$ the estimate of the variable $\boldsymbol{x}$.
- The world frame and the robot's IMU frame are denoted $\mathcal{W}$ and $\mathcal{I}$, respectively. The frame associated with the i-th contact is denoted $\mathcal{C}_i$. The anchor point, defined in Section 3, is denoted $\mathcal{A}$.
- $\times$ is the cross-product operator.
- We define n_c the number of current contacts set with the environment.

3. DEFINITION OF THE ANCHOR POINT

We define here the notion of anchor point, since it will be used extensively in the following sections. The position of the anchor point can be merged with the robot's joint encoders and IMU signals to provide an algebraic estimate of the IMU velocity in the world frame. This will be explained in Section 4.2 where we show that it is essential for the estimator design.

The anchor point, denoted $\mathcal{A}$, is defined as a point attached to the robot and which is considered to have zero *instantanous* linear velocity in the world frame. This point does not need to be fixed in the robot frame as well, the only requirement is that its instantaneous position and linear velocity in the frame of the robot's floating base are known.

Contacts with the environment provide naturally good sets of points known to the robot and having zero world velocity. With the no-slippage hypothesis, any point of a contact body would thus be suitable to define an anchor point. However, we need an anchor point that remains always defined and whose position is as continuous as possible, even when contacts are created and broken, as long as there is at least one contact established. Also, the no-slippage hypothesis can hold better for some points than others. For instance, when the contact forces are closer to Coulomb's friction cones the contact is more likely to slip.

To address these two requirements we determine the kinematics of that point through a weighted average of the kinematics of the current contacts, as shown in Figure 1. The weighting coefficients are defined such that weaker contacts, which are prone to violate Coulomb's inequality

[1] https://github.com/ArnaudDmt/state-observation
[2] https://github.com/ArnaudDmt/state_observation_ros.git

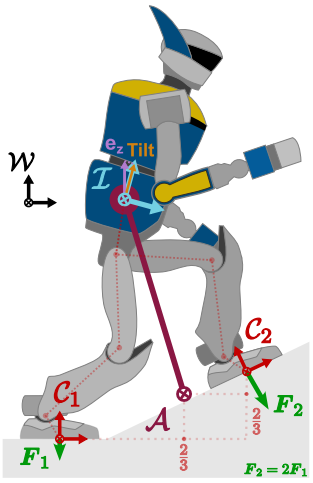

Fig. 1. Illustration of the anchor point position computation and of the reference frames used in VALINOR. $\mathcal{W}$: world frame; $\mathcal{I}$: IMU frame; $\mathcal{C}_i$: frame of the i-th contact. For clarity, only the weighted average of the contact positions is shown. In this example, $\boldsymbol{F}_2 = 2\boldsymbol{F}_1$ and so $\lambda_2 = \frac{2}{3}$.

and thus to slip, contribute less to the anchor point's kinematics computation. Given mg_0 the weight of the robot, we first compute

$$u_i = \frac{\boldsymbol{F}_{i,z}}{\sqrt{\boldsymbol{F}_{i,x}^2 + \boldsymbol{F}_{i,y}^2} + \varepsilon m g_0}, \tag{1}$$

the ratio of the normal force to the tangential force at the contact i. The term ε, arbitrarily small, allows to deal with the case where the tangential force is zero.

The contact's weighting coefficient λ_i is then:

$$\lambda_i = \frac{u_i}{\sum_{j=1}^{n_c} u_i}. \tag{2}$$

Based on this definition, we give the position and linear velocity of the anchor point in the IMU's frame:

$${}^{\mathcal{I}}\boldsymbol{p}_{\mathcal{A}} = \sum_i^{n_c} \lambda_i {}^{\mathcal{I}}\boldsymbol{p}_{\mathcal{C}_i}, \tag{3}$$

$${}^{\mathcal{I}}\dot{\boldsymbol{p}}_{\mathcal{A}'} = \sum_i^{n_c} \lambda_i {}^{\mathcal{I}}\dot{\boldsymbol{p}}_{\mathcal{C}_i}, \tag{4}$$

with ${}^{\mathcal{I}}\boldsymbol{p}_{\mathcal{C}_i}$ the position and ${}^{\mathcal{I}}\dot{\boldsymbol{p}}_{\mathcal{C}_i}$ the linear velocity of the i-th contact in the IMU's frame, which are directly provided by the robot's joint encoders and geometrical model.

It is important to note also that with this definition, although the anchor point is defined to have zero velocity in the world, its instantaneous position may still evolve over time.

4. TILT OBSERVER WITH PROOF OF CONVERGENCE

The proposed estimator relies on a highly accurate estimate of the IMU's tilt provided by a complementary filter, introduced in [19], which we will call *Tilt Observer*.

4.1. Definition of the State Variables

The Tilt Observer is able to provide estimates of the following two variables:

- $\boldsymbol{v}_{\mathcal{I},l} \triangleq \boldsymbol{R}_{\mathcal{I}}^T \boldsymbol{v}_{\mathcal{I}}$ the linear velocity of the IMU's frame in the world frame, expressed in the frame of the IMU.
- $\boldsymbol{R}_{\mathcal{I}}^T \boldsymbol{e}_z$ the tilt of the IMU.

We thus define the state variables:

$$\boldsymbol{x}_1 \triangleq \boldsymbol{v}_{\mathcal{I},l} \quad , \boldsymbol{x}_1 \in \mathbb{R}^3, \tag{5}$$

$$\boldsymbol{x}_2 \triangleq \boldsymbol{R}_{\mathcal{I}}^T \boldsymbol{e}_z \quad , \boldsymbol{x}_2 \in \mathbb{S}^2. \tag{6}$$

The set $\mathbb{S}^2 \subset \mathbb{R}^3$ is the unit sphere centered at the origin, and defined as

$$\mathbb{S}^2 = \left\{ \boldsymbol{x} \in \mathbb{R}^3 \middle| \|\boldsymbol{x}\| = 1 \right\} \tag{7}$$

4.2. Definition of the Measurements

The measurements required by the Tilt Observer are:

- $\boldsymbol{y}_g$ the signal of the IMU's gyrometer.
- $\boldsymbol{y}_a$ the signal of the IMU's accelerometer.
- $\boldsymbol{y}_v$ a measurement of $\boldsymbol{v}_{\mathcal{I},l}$.

Since no sensor provides a direct measurement of $\boldsymbol{v}_{\mathcal{I},l}$, we obtain $\boldsymbol{y}_v$ from an algebraic combination of gyrometer signal $\boldsymbol{y}_a$ and the position and velocity of the anchor point, ${}^{\mathcal{I}}\boldsymbol{p}_{\mathcal{A}}$ from (3) and ${}^{\mathcal{I}}\dot{\boldsymbol{p}}_{\mathcal{A}}$ from (4) respectively, giving:

$$\boldsymbol{y}_v = -\boldsymbol{y}_g \times {}^{\mathcal{I}}\boldsymbol{p}_{\mathcal{A}} - {}^{\mathcal{I}}\dot{\boldsymbol{p}}_{\mathcal{A}} \tag{8}$$

4.3. Definition of the Filter

The state dynamics of our system can be written:

$$\dot{\boldsymbol{x}}_1 = -\boldsymbol{y}_g \times \boldsymbol{x}_1 - g_0 \boldsymbol{x}_2 + \boldsymbol{y}_a, \tag{9}$$

$$\dot{\boldsymbol{x}}_2 = -\boldsymbol{y}_g \times \boldsymbol{x}_2. \tag{10}$$

with g_0 the gravitational acceleration constant. We write a complementary filter which uses the system's dynamics as a feed-forward, and corrects the estimated state using the velocity measurement $\boldsymbol{y}_v$:

$$\begin{cases} \dot{\hat{\boldsymbol{x}}}_1 = -\boldsymbol{y}_g \times \hat{\boldsymbol{x}}_1 - g_0 \hat{\boldsymbol{x}}_2' + \boldsymbol{y}_a + \alpha_1 (\boldsymbol{y}_v - \hat{\boldsymbol{x}}_1), \\ \dot{\hat{\boldsymbol{x}}}_2' = -\boldsymbol{y}_g \times \hat{\boldsymbol{x}}_2 - \dfrac{\alpha_2}{g_0} (\boldsymbol{y}_v - \hat{\boldsymbol{x}}_1), \\ \dot{\hat{\boldsymbol{x}}}_2 = -(\boldsymbol{y}_g - \gamma \hat{\boldsymbol{x}}_2 \times \hat{\boldsymbol{x}}_2') \times \hat{\boldsymbol{x}}_2. \end{cases} \tag{11}$$

α_1, α_2 and γ are positive scalar gains. $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ are estimates of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively.

4.4. Advantages of the Tilt Observer

The use of a complementary filter for the Tilt Observer presents notable strengths in comparison to other methods like the commonly used Kalman Filter. First, it allows us to easily work in the frequency domain when determining the gains. This is particularly suitable for our model since the assumption of fixed contacts used to obtain the velocity measurement $\boldsymbol{y}_v$ is more valid in low frequency than in high frequency. In (9), we thus use the IMU measurements for the high frequency variation of $\hat{\boldsymbol{x}}_1$, and $\boldsymbol{y}_v$ for its low frequency variation. Second, one iteration of the filter only consists in computing three equations, it also doesn't involve any matrix multiplication or inversion and is thus computationally extremely cheap, as will be shown in Section 7.2.1. Finally, the formulation as a complementary filter allows to conduct a convergence analysis of the estimation error, providing strong mathematical guarantees on the estimator's performances. Especially here, it has been shown in [19] that:

- The dynamics of the estimation error is autonomous, and thus does not depend on the motion of the robot.
- The intermediate estimator $\{\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}'_2\}$ is *globally exponentially stable*, with respect to the origin $(0,0)$.
- The full estimator is *almost globally asymptotically stable*, and locally *exponentially stable*. It provides a good filtering of noise and guarantees to respect the normality constraint.

5. LEG ODOMETRY FOR POSITION AND YAW ESTIMATION

While the tilt of the IMU's frame in the world frame is estimated by the Tilt Observer, its position and yaw are obtained using Leg odometry. Once a contact i is created, its initial pose in the world is obtained by forward kinematics from the current IMU's frame pose and the robot's joint encoders. We call it the contact's *reference* pose $\left\{\boldsymbol{p}^\star_{\mathcal{C}_i}, \boldsymbol{R}^\star_{\mathcal{C}_i}\right\}$, which we consider constant to enforce the contact's anchoring role. This pose is then used to recover the pose of the IMU's frame in the world frame. With the proposed pipeline, we thus leverage both the accuracy and mathematical guarantees provided by the Tilt Observer, and the robustness to drift provided by the Leg odometry. Similarly to the computation of the anchor point in Section 3, the contribution of contacts to the Leg odometry is weighted to trust more contacts which are the least prone to slippage, in order to mitigate its effect.

First, we estimate the IMU's orientation from contact information. For the two most reliable contacts[3], we compute the IMU frame's orientation from the contact's reference orientation. For each contact i, we write:

$$\boldsymbol{R}_{\mathcal{I},i} = \boldsymbol{R}^\star_{\mathcal{C}_i} {}^{\mathcal{C}_i}\boldsymbol{R}_{\mathcal{I}}, \tag{12}$$

[3]The two contacts with the highest ratio u defined in (1).

with ${}^{\mathcal{C}_i}\boldsymbol{R}_{\mathcal{I}}$ the orientation of the IMU's frame in the contact frame, provided by the robot's geometrical model and joint encoders.
We then compute the error between the obtained orientations, which is used to compute their weighted average, using the formalism defined by SO(3) the Lie group of rotation matrices:

$$\tilde{\boldsymbol{R}} = \boldsymbol{R}^T_{\mathcal{I},1}\boldsymbol{R}_{\mathcal{I},2} \tag{13}$$

$$\hat{\boldsymbol{R}}_{\mathcal{I}} = \boldsymbol{R}_{\mathcal{I},1}\exp\left(\lambda_2 \log\left(\tilde{\boldsymbol{R}}\right)\right). \tag{14}$$

exp and log are the *exponential* and *logarithm* maps of SO(3). We note that for small angles, we can use the approximation:

$$\log\left(\tilde{\boldsymbol{R}}\right) \simeq \frac{1}{2}\left(\tilde{\boldsymbol{R}} - \tilde{\boldsymbol{R}}^T\right), \tag{15}$$

Once the IMU's orientation $\hat{\boldsymbol{R}}_{\mathcal{I}}$ has been computed, we merge the corresponding yaw with the tilt estimated by the Tilt Observer using a new axis-agnostic fusion of tilt and yaw, that we describe in Section 6, which preserves the estimated tilt with little computations.

The position of the IMU's frame in the world frame is then obtained from the n_c contact reference poses:

$$\hat{\boldsymbol{p}}_{\mathcal{I}} = \sum_i^{n_c} \lambda_i \left(\boldsymbol{p}^\star_{\mathcal{C}_i} - \hat{\boldsymbol{R}}_{\mathcal{I}}{}^{\mathcal{I}}\boldsymbol{p}_{\mathcal{C}_i}\right). \tag{16}$$

We note that this position is obtained using the IMU's orientation estimate that merges the estimated tilt with the yaw coming from the odometry. It takes then full profit from the accurate tilt provided by the Tilt Observer.

6. TILT–YAW FUSION

The Tilt Observer of Section 4 provides $\hat{\boldsymbol{x}}_2$, an accurate estimate of the IMU *tilt* noted $\boldsymbol{\ell} = \boldsymbol{R}_1^\top \boldsymbol{e}_z$, while the contact–based Leg odometry of Section 5 supplies an orientation $\boldsymbol{R}_2 = \hat{\boldsymbol{R}}_{\mathcal{I}}$, whose yaw reflects the robot's heading. We now merge these two pieces into a single rotation $\boldsymbol{R}$.

6.1. Why not Euler angles?

Classically in these situations, we use the Euler angles representation of the rotations, $\boldsymbol{R}_1 = R_z(\psi_1)R_y(\theta_1)R_x(\phi_1)$ and $\boldsymbol{R}_2 = R_z(\psi_2)R_y(\theta_2)R_x(\phi_2)$, the textbook fusion is

$$\boldsymbol{R}_{\text{cl}} = R_z(\psi_2)\,R_y(\theta_1)\,R_x(\phi_1). \tag{17}$$

This solution has two properties

- tilt is conserved, i.e. $\boldsymbol{R}_{\text{cl}}^\top \boldsymbol{e}_z = \boldsymbol{\ell}$.
- among all rotations $\hat{\boldsymbol{R}}$ such that $\hat{\boldsymbol{R}}^\top \boldsymbol{e}_z = \boldsymbol{\ell}$, R_{cl} minimises the horizontal error $\|\hat{\boldsymbol{R}}^\top \boldsymbol{e}_x - \boldsymbol{R}_2^\top \boldsymbol{e}_x\|^2$.

Despite its simplicity, (17) rests on two tacit assumptions that are seldom met in the leg–inertial setting considered here:

- The optimality of (17) is defined with respect to the IMU axis $\boldsymbol{e}_x$ only. This makes sense when $\boldsymbol{e}_x$ *is* the heading carrier, for example with a GPS-based heading for wheeled vehicles. In practice the IMU can be installed with an arbitrary yaw offset, and our yaw information actually comes from leg odometry (average contact frames, Section 5), not from that axis. Treating $\boldsymbol{e}_x$ as "special" therefore introduces a systematic bias.
- Because (17) uses a ZYX decomposition of $\boldsymbol{R}_1$, it inherits the Euler-angle singularity at $\theta_1 = \pm 90°$. Exactly there the pitch axis aligns with $\boldsymbol{e}_z$, roll and yaw collapse, and the horizontal projection of $\boldsymbol{e}_x$ disappears, leaving the yaw numerically indeterminate. Such attitudes are frequent in humanoids that perform multicontact motions.

To avoid these two issues we adopt an *axis-agnostic* fusion that uses solely the measured tilt $\boldsymbol{\ell}$, allows any horizontal vector to encode yaw, and stays regular for all configurations except the physically undefined case $\boldsymbol{\ell} = -\boldsymbol{e}_z$. The construction is detailed next.

6.2. Yaw without Euler angles

The yaw contained in $\boldsymbol{R}_2$ can be extracted without decomposing the matrix into Euler angles. Let's denote the third column of $\boldsymbol{R}_2$ as $\boldsymbol{r}_2$. This column describes the world coordinates of the body $\boldsymbol{z}$-axis, i.e $\boldsymbol{r}_2 = \boldsymbol{R}_2\boldsymbol{e}_z$, which is not the tilt of $\boldsymbol{R}_2$. Using that row we build the horizontal vector

$$\boldsymbol{m}_2 = \begin{cases} \boldsymbol{e}_x, & \text{if } r_{2,x}^2 + r_{2,y}^2 \simeq 0, \\ \dfrac{1}{\sqrt{r_{2,x}^2 + r_{2,y}^2}} [r_{2,y} \;\; -r_{2,x} \; 0]^\top, & \text{otherwise.} \end{cases} \tag{18}$$

where $r_{2,x}$ and $r_{2,y}$ are the first and second components of $\boldsymbol{r}_2$. The condition $r_{2,x}^2 + r_{2,y}^2 \simeq 0$ occurs when the tilt $\boldsymbol{R}_2^T\boldsymbol{e}_z$ is vertical (upward or downward), so we take $\boldsymbol{m} = \boldsymbol{e}_{\boldsymbol{x}}$ to avoid the singularity. Otherwise, the vector $\boldsymbol{m}_2$ always exists and is unique.

Rotating $\boldsymbol{m}_2$ by $\boldsymbol{R}_2^T$ results in another vector $\tilde{\boldsymbol{m}}_2 = \boldsymbol{R}_2^T\boldsymbol{m}_2$, which is horizontal as well. This means that $\boldsymbol{m}_2$ is an invariant horizontal vector.

We define the yaw of $\boldsymbol{R}_2$, as the angle between these two vectors $\boldsymbol{m}_2$ and $\tilde{\boldsymbol{m}}_2$, which is computed as

$$\theta = \operatorname{atan2}\Big(m_{2,x}\tilde{m}_{2,y} - m_{2,y}\tilde{m}_{2,x},\; m_{2,x}\tilde{m}_{2,x} + m_{2,y}\tilde{m}_{2,y}\Big), \tag{19}$$

where the x and y subscripts denote the first and second components of the corresponding vector. The use of horizontal invariant vectors to define the yaw angle guarantees that we only consider rotations around the vertical axis, giving the most sound definition of yaw.

When the tilt $\boldsymbol{R}_2^T\boldsymbol{e}_z$ is vertical, any horizontal vector is invariant. In such a case, taking $\boldsymbol{m} = \boldsymbol{e}_{\boldsymbol{x}}$ amounts at mimicking Euler yaw angle. When this tilt is downward, meaning $\boldsymbol{R}_2$ is a rotation of π around a horizontal axis, the concept of yaw itself is ill-defined. When the tilt is upward, any horizontal vector gives the same yaw, so using Euler's yaw angle by taking $\boldsymbol{m}_2 = \boldsymbol{e}_{\boldsymbol{x}}$ is a valid choice.

Fortunately we don't have to compute this angle explicitly to perform the merge. It can be done directly using the tilt $\boldsymbol{\ell}$ and the second matrix $\boldsymbol{R}_2$.

6.3. Axis-agnostic triad fusion

6.3.1 Select a proper reference vector

Similarly to the yaw computation above, we start with the selection of the reference vector. However, the fusion makes this step slightly more subtile than just identifying the yaw of a single matrix. This is because the two rotations could have a very different tilt. This could lead the tilt vector $\boldsymbol{\ell}$ to be nearly colinear with the rotated horizontal invariant vector $\boldsymbol{R}_2^T\boldsymbol{m}_2$. This would make the fusion singular.

The solution is to find a reference vector that is both horizontal and orthogonal to the real up direction $\boldsymbol{r} = \boldsymbol{R}_2\boldsymbol{\ell} = [v_x \; v_y \; v_z]^\top$. Consequently, a proper reference vector can be defined as

$$\boldsymbol{m} = \begin{cases} \boldsymbol{m}_2, & \text{if } v_x^2 + v_y^2 \simeq 0, \\ \dfrac{1}{\sqrt{v_x^2 + v_y^2}} [v_y \; -v_x \; 0]^\top, & \text{otherwise.} \end{cases}$$

The degenerate case occurs only when $\boldsymbol{r}$ is vertical, corresponding to $\boldsymbol{R}_2^T\boldsymbol{e}_z = \boldsymbol{\ell}$, which means the rotations nearly have the same tilt. In this situation we know that $\boldsymbol{\ell}$ and $\boldsymbol{R}_2^T\boldsymbol{m}_2$ are nearly orthogonal. We thus safely use the yaw angle definition of the previous section.

We define the vector $\boldsymbol{m}$ expressed in the frame of $\boldsymbol{R}_2$ as $\boldsymbol{m}_l = R_2^\top\boldsymbol{m}$.

6.3.2 Build two right-handed bases.

Using cross products, form

$$\boldsymbol{B}_1 = [\boldsymbol{m} \times \boldsymbol{e}_z \quad \boldsymbol{e}_z \times \boldsymbol{m} \times \boldsymbol{e}_z \quad \boldsymbol{e}_z],$$

$$\boldsymbol{B}_2 = \left[\frac{\boldsymbol{m}_l \times \boldsymbol{\ell}}{\|\boldsymbol{m}_l \times \boldsymbol{\ell}\|} \quad \frac{\boldsymbol{\ell} \times \boldsymbol{m}_l \times \boldsymbol{\ell}}{\|\boldsymbol{m}_l \times \boldsymbol{\ell}\|} \quad \boldsymbol{\ell}\right].$$

Each matrix is orthonormal: the first column is the chosen horizontal vector, the third column is the vertical direction, and the second column completes a right-handed frame.

6.3.3 Fuse tilt and yaw.

The desired rotation is obtained by rotating from the basis $\{\boldsymbol{m}_2, \cdot, \boldsymbol{\ell}\}$ to $\{\boldsymbol{m}, \cdot, \boldsymbol{e}_z\}$:

$$\boldsymbol{R} = \boldsymbol{B}_1\,\boldsymbol{B}_2^\top. \tag{20}$$

6.4. Properties

The rotation obtained with (20) shows the following properties:

- *Tilt preservation*: $\boldsymbol{R}^\top \boldsymbol{e}_z = \boldsymbol{\ell}$.
- *Optimal horizontal alignment*: for any rotation $\hat{\boldsymbol{R}}$ satisfying $\hat{\boldsymbol{R}}^\top \boldsymbol{e}_z = \boldsymbol{\ell}$, the choice (20) minimises

$$\left\| \hat{\boldsymbol{R}}^\top \boldsymbol{m} - \boldsymbol{m}_l \right\|^2.$$

  The proof follows exactly the one–parameter maximisation used for the Euler fusion, but with the task-specific direction $\boldsymbol{m}$ in place of the body axis $\boldsymbol{e}_x$.
- *Singularity*: the construction is undefined only when $\boldsymbol{\ell} = -\boldsymbol{e}_z$; at that upside-down pose yaw itself has no meaning.
- *Computation*: implementation requires only vector cross products and normalisations, with no Euler-angle extraction, keeping the cost compatible with real-time embedded execution.

The axis-agnostic fusion is therefore adopted as the default in VALINOR; it avoids the arbitrary preference for the body axis $\boldsymbol{e}_x$, takes into account the difference of tilt and avoids the gimbal-lock issues of (17).

7. EXPERIMENTAL EVALUATION

7.1. Description of the experiments

The proposed estimator has been evaluated accross two experimental scenarios on two different humanoid robots[4]:

- Experiment 1: A walk on a flat ground over about 18 meters with the robot RHP Friends [31], as shown in Figure 2. This experiment was repeated 5 times, for a total distance of about 90 meters.
- Experiment 2: A multi-contact motion over about 2 meters with the robot HRP-5P [32]. This motion involved an additional contact at the robot's left hand, and non-coplanar contacts on tilted obstacles, as shown in Figure 3. This experiment was repeated 4 times for a total distance of about 8 meters.

We compare VALINOR, our proposed estimator, with the Right-Invariant EKF (RI-EKF) from [27], a state-of-the-art method for legged robot proprioceptive odometry. This estimator has demonstrated its effectiveness in practical applications, notably as real-time feedback within a complex control pipeline [29], which is the type of use case intended for VALINOR. It is therefore particularly relevant to compare ourselves to this method.

The comparison is made on both their computation and their estimation performance. For fairness, both estimators received the same contact state information, from

[4]We used the dataset built to evaluate the Kinetics Observer in [28]

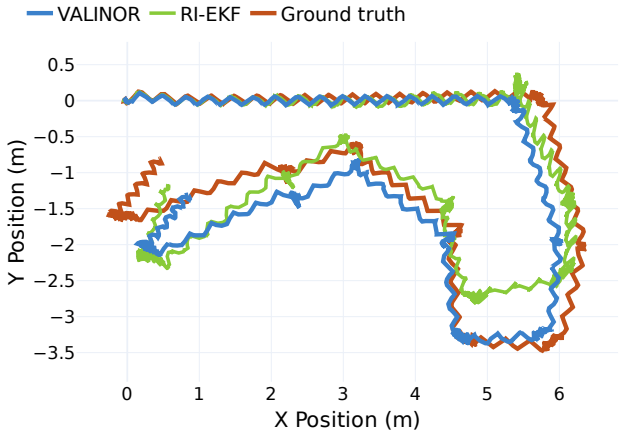

Fig. 2. Ground truth and estimated RHP Friend's pelvis trajectory during the first experiment of walk on flat ground.

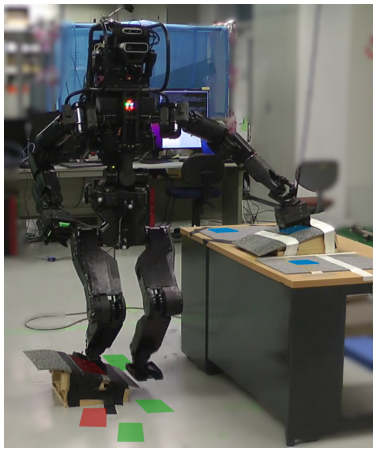

Fig. 3. Picture of HRP-5P during the multi-contact motion. The imprints of the successive contacts are represented by colored trapezoids: Red: Right foot; Green: Left foot; Blue: Left hand. Figure reused from [28].

a Schmitt Trigger on the Ground Reaction Force (with thresholds set to 10% and 15% of the robot's weight). Ground truth pose of the robot's pelvis (to which the IMU is fixed on both robots) is provided by a motion capture system (OptiTrack, 16 Prime^X with 13 cameras), and its ground truth velocity is obtained by finite differences from the ground truth position then filtered with a zero-phase low-pass filter. The pose of the IMU's frame, estimated by both VALINOR and the RI-EKF, is transformed into that of the pelvis using simple rigid body transformation.

The odometry performance of VALINOR and the RI-EKF is evaluated against the ground truth trajectory using the Relative Error (RE), as defined in [33]. RE evaluates the estimation drift over segments of fixed distances. It is particularly relevant for proprioceptive odometry, as it is not affected by global drift, which is unavoidable without exteroceptive sensors. RE therefore provides valuable and

easily interpretable insight into the expected local accuracy of the estimator over a given distance.

We compute the Relative Error for lateral (x, y) and vertical (z) translations separately, since these components are generally influenced by different factors. Accuracy in lateral translation would mainly rely on that of local displacements and of yaw estimation, whereas accuracy in vertical translations would be affected by the accuracy of the estimation of the vertical local displacement, of the tilt, and the reliability of the contact height initialization. Similarily, angular errors in tilt and yaw are also evaluated independently. The error e_ℓ on the tilt estimate is computed with:

$$e_\ell = \arccos\left(\ell_{gt}^T \hat{\ell}\right), \tag{21}$$

with ℓ_{gt} the ground truth tilt, and $\hat{\ell}^T$ the estimated one.
Finally, we also assess the estimation of the IMU's linear velocity in the world. This velocity is expressed in the frame of the IMU such that it is not affected by errors in the orientation estimate. The velocity error is also separated into lateral and vertical components.

7.2. Experimental Results

7.2.1 Computational Performance Evaluation

To evaluate the computation speed performance of VALINOR, we compared its average computation time per iteration over the four multi-contact experiments, to that of the RI-EKF. Both estimators were run on the same laptop, on an Intel Xeon E-2176M CPU (2.70GHz x 12). In average, an iteration of VALINOR was computed in 2.547 µs, against 19.315 µs for the RI-EKF. The proposed method is thus more than 7.5 times faster than the state-of-the-art estimator, while proposing equivalent estimation accuracy, as shown below.

7.2.2 Estimation Accuracy Evaluation

Table 1 regroups the evaluation results of VALINOR's estimation performance compared to those of the RI-EKF during the two experiments on real robots. The presented errors are computed over all the sequences of each scenario: 5 for the walk on flat ground, and 4 for the multi-contact motion.

We first note that the relative strengths of both estimators were consistent accross both experiments, and that overall, they both displayed satisfactory estimation performance. Indeed, without the use of any exteroceptive sensors, their errors in translation estimates remain of the order of the centimer per meter walked, those on orientation estimates remain around 1° per meter walked, and finally their error in velocity estimates remain below 2 cm.s^{-1}. Now we give more insights into the obtained results:

- VALINOR estimated more accurately lateral translations, improving the estimation by over 30% and 40% compared to the RI-EKF. Additionally, as visible in Figure 4 it improved the estimation of the tilt by about 28% in Experiment 1 and by 60% in Experiment 2, reaching an average error of only 0.23°. For both translations and tilt estimates, our method displayed a significantly lower standard deviation of the error, proving an improved estimation consistency.
- The RI-EKF estimated lateral velocities more accurately than VALINOR, with a 20% smaller the error. But we believe this difference has lower significance due to the large value of the standard deviation.
- Both estimators showed similar performance for the estimation of the yaw and of vertical translation and velocity.

As shown in Figure 5, we noticed that both estimators drifted upwards during the walk on flat ground, which involved longer walks and many more steps than the multi-contact motion. One of the reasons, inherent to Leg odometry, is its reliance on the quality of the contact detection. For example, detecting contacts consistently too early (when they are not perfectly still yet), would cause the estimate to drift on each step. Conversely, drift can also occur when detecting contacts consistently too late, due to the robot's structural deformation under the contact force, etc. While investigating, we could indeed reduce the vertical drift by tuning the thresholds of the Schmitt Trigger used for contact detection. However, we decided to leave them changed to ensure a fair comparison. The other main factor explaining the vertical drift is the quality of the pitch estimate, which significantly affects vertical translation estimates during long forward walks.

8. CONCLUSION

In this paper, we proposed VALINOR, a lightweight estimator for humanoid robots. By combining Leg odometry with the Tilt Observer, we could perform accurate odometry whose accuracy rivalizes with the state-of-the-art method. We notably improved significantly the estimation of the lateral translations and of the tilt, while being over 7.5 times faster. Through this contribution, we aim to highlight the importance and the effectiveness of lightweight state estimation methods, especially in the current state of control for legged robots where computational efficiency is becoming a predominant challenge. We also seek to drive interest towards methods with mathematical guarantees, which we believe are essential for the safe industrial and societal deployment of legged robots. Further work would aim at improving the coupling between the Inertial and the Leg odometry, notably such that the position and yaw estimates leverage better IMU measurements. Other extensions could investigate the incorporation of other sensors into the framework.

Table 1. Mean and standard deviation (in parentheses) of the error on the estimations by VALINOR and the RI-EKF, during the two experiments on real robots. The 1 m and the 0.3 m Relative Errors are presented for the walk on flat ground and the multi-contact motion, respectively. The best result for each metric is highlighted in bold.

		RE Translation [m]		RE Orientation [°]		Linear velocity [m.s^{-1}]	
		Lateral $\{x, y\}$	Vertical z	Tilt	Yaw	Lateral $\{x, y\}$	Vertical z
Walk on flat ground	VALINOR (Proposed)	**0.032** (0.021)	0.015 (0.012)	**0.49** (0.29)	1.14 (1.17)	0.019 (0.021)	**0.007** (0.008)
	RI-EKF [27]	0.047 (0.057)	**0.011** (0.026)	0.68 (0.49)	**0.96** (0.90)	**0.015** (0.017)	0.008 (0.010)
Multi-contact motion	VALINOR (Proposed)	**0.007** (0.005)	**0.002** (0.002)	**0.23** (0.17)	**0.40** (0.32)	0.010 (0.011)	**0.004** (0.006)
	RI-EKF [27]	0.012 (0.016)	0.004 (0.006)	0.57 (1.02)	0.47 (0.65)	**0.009** (0.009)	0.005 (0.007)

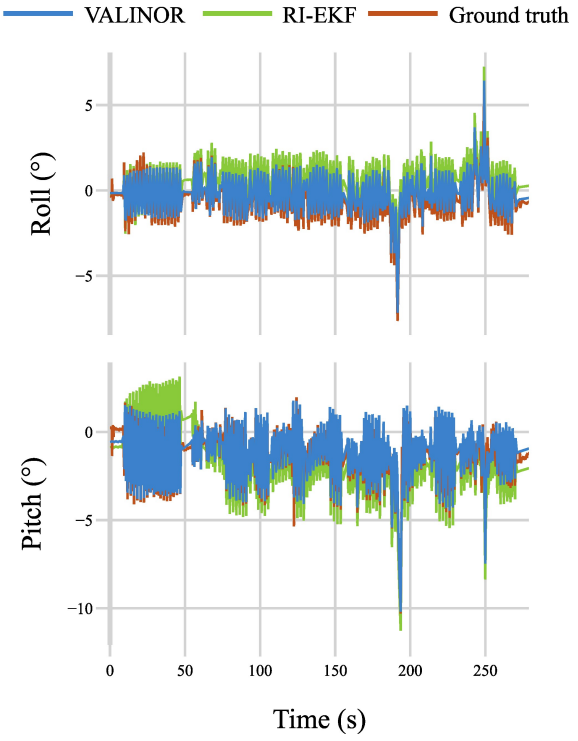

Fig. 4. Estimated and ground truth roll and pitch angles of the robot's pelvis during the first trial of the walk on flat floor.

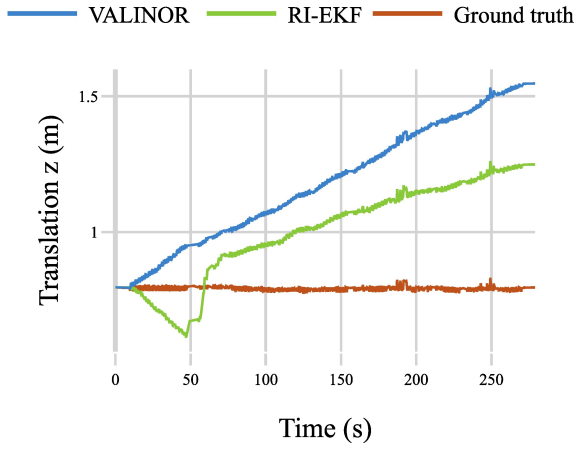

Fig. 5. Estimated and ground truth height of the robot's pelvis during the first trial of the walk on flat floor.

DECLARATIONS

Conflict of Interest

The authors declare that there is no competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

Authors' Contributions

Arnaud Demont and Mehdi Benallegue conceived the study and developed its theoretical foundations. Abdelaziz Benallegue supervised the study and validated the theoretical framework. Arnaud Demont and Thomas Duvinage conducted the experiments. Arnaud Demont implemented the work in C++ and processed the experimental results. Arnaud Demont and Mehdi Benallegue wrote the manuscript in consultation with Thomas Duvinage and Abdelaziz Benallegue.

Funding

This paper is based on results obtained from a project of Programs for Bridging the gap between R&D and the IDeal society (society 5.0) and Generating Economic and social value (BRIDGE)/Practical Global Research in the AI x Robotics Services, implemented by the Cabinet Office, Government of Japan, and partially funded by the Japan Science and Technology Agency (JST) with the JST-Mirai Program, grant number JPMJMI21H4.

REFERENCES

[1] 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," *Autonomous Robots*, vol. 40, 07 2015.

[2] T. Sugihara and M. Morisawa, "A survey: dynamics of humanoid robots," *Advanced Robotics*, vol. 34, no. 21-22, pp. 1338–1352, 2020.

[3] A. Chlipala, *Certified programming with dependent types: a pragmatic introduction to the Coq proof assistant*. MIT Press, 2013.

[4] S. Katayama, M. Murooka, and Y. T. and, "Model predictive control of legged and humanoid robots: models and algorithms," *Advanced Robotics*, vol. 37, no. 5, pp. 298–315, 2023.

[5] E. Dantec, M. Naveau, P. Fernbach, N. Villa, G. Saurel, O. Stasse, M. Taix, and N. Mansard, "Whole-body model predictive control for biped locomotion on a torque-controlled humanoid robot," in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, Nov 2022, pp. 638–644.

[6] A. Dallard, M. Benallegue, N. Scianca, F. Kanehiro, and A. Kheddar, "Robust Bipedal Walking with Closed-Loop MPC: Adios Stabilizers," *hal preprint*, Feb. 2024.

[7] J. Peters, S. Vijayakumar, and S. Schaal, "Reinforcement learning for humanoid robotics," *Proceedings of the third IEEE-RAS international conference on humanoid robots*, pp. 1–20, 01 2003.

[8] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for versatile, dynamic, and robust bipedal locomotion control," *The International Journal of Robotics Research*, vol. 44, no. 5, pp. 840–888, 2025.

[9] NVIDIA, :, J. Bjorck, F. Castañeda, N. Cherniadev, X. Da, R. Ding, L. J. Fan, Y. Fang, D. Fox, F. Hu, S. Huang, J. Jang, Z. Jiang, J. Kautz, K. Kundalia, L. Lao, Z. Li, Z. Lin, K. Lin, G. Liu, E. Llontop, L. Magne, A. Mandlekar, A. Narayan, S. Nasiriany, S. Reed, Y. L. Tan, G. Wang, Z. Wang, J. Wang, Q. Wang, J. Xiang, Y. Xie, Y. Xu, Z. Xu, S. Ye, Z. Yu, A. Zhang, H. Zhang, Y. Zhao, R. Zheng, and Y. Zhu, "Gr00t n1: An open foundation model for generalist humanoid robots," *arXiv preprint*, 2025.

[10] K. Kawaharazuka, T. Matsushima, A. Gambardella, J. Guo, C. Paxton, and A. Zeng, "Real-world robot applications of foundation models: A review," *Advanced Robotics*, vol. 38, no. 18, pp. 1232–1254, 2024.

[11] M. N. Zeilinger, D. M. Raimondo, A. Domahidi, M. Morari, and C. N. Jones, "On real-time robust model predictive control," *Automatica*, vol. 50, no. 3, pp. 683–694, 2014.

[12] R. Findeisen and F. Allgöwer, "Computational delay in nonlinear model predictive control," *IFAC Proceedings Volumes*, vol. 37, no. 1, pp. 427–432, 2004.

[13] P. Thodoroff, W. Li, and N. D. Lawrence, "Benchmarking real-time reinforcement learning," in *NeurIPS 2021 Workshop on Pre-registration in Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 181. PMLR, 13 Dec 2022, pp. 26–41.

[14] R. Firoozi, J. Tucker, S. Tian, A. Majumdar, J. Sun, W. Liu, Y. Zhu, S. Song, A. Kapoor, K. Hausman, B. Ichter, D. Driess, J. Wu, C. Lu, and M. Schwager, "Foundation models in robotics: Applications, challenges, and the future," *The International Journal of Robotics Research*, vol. 44, no. 5, pp. 701–739, 2025.

[15] D. Wisth, M. Camurri, and M. Fallon, "Vilens: Visual, inertial, lidar, and leg odometry for all-terrain legged robots," *IEEE Transactions on Robotics*, vol. 39, no. 1, pp. 309–326, 2022.

[16] M. Fallon, "Accurate and robust localization for walking robots fusing kinematics, inertial, vision and lidar," *Interface focus*, vol. 8, no. 4, p. 20180015, 2018.

[17] Y. Kuang, T. Hu, M. Ouyang, Y. Yang, and X. Zhang, "Tightly coupled lidar/imu/uwb fusion via resilient factor graph for quadruped robot positioning," *Remote Sensing*, vol. 16, no. 22, 2024.

[18] R. Mahony, T. Hamel, and J.-M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on automatic control*, vol. 53, no. 5, pp. 1203–1218, 2008.

[19] M. Benallegue, R. Cisneros, A. Benallegue, Y. Chitour, M. Morisawa, and F. Kanehiro, "Lyapunov-Stable Orientation Estimator for Humanoid Robots," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6371–6378, Oct. 2020.

[20] P.-C. Lin, H. Komsuoglu, and D. Koditschek, "A leg configuration measurement system for full-body pose estimates in a hexapod robot," *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 411–422, June 2005.

[21] 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," *Robotics*, vol. 17, pp. 17–24, 2013.

[22] P. Wawrzyński, J. Możaryn, and J. Klimaszewski, "Robust estimation of walking robots velocity and tilt using proprioceptive sensors data fusion," *Robotics and Autonomous Systems*, vol. 66, pp. 44–54, 2015.

[23] K. Masuya and T. Sugihara, "Dead reckoning for biped robots that suffers less from foot contact condition based on anchoring pivot estimation," *Advanced Robotics*, vol. 29, no. 12, pp. 785–799, 2015.

[24] T.-Y. Lin, R. Zhang, J. Yu, and M. Ghaffari, "Legged robot state estimation using invariant kalman filtering and learned contact events," in *Proceedings of the 5th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 08–11 Nov 2022, pp. 1057–1066.

[25] M. Maravgakis, D.-E. Argiropoulos, S. Piperakis, and P. Trahanias, "Probabilistic contact state estimation for legged robots using inertial information," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 163–12 169.

[26] Z. Yoon, J.-H. Kim, and H.-W. Park, "Invariant smoother for legged robot state estimation with dynamic contact event information," *IEEE Transactions on Robotics*, 2023.

[27] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizzle, "Contact-aided invariant extended kalman filtering for robot state estimation," *The International Journal of Robotics Research*, vol. 39, no. 4, pp. 402–430, 2020.

[28] A. Demont, M. Benallegue, A. Benallegue, P. Gergondet, A. Dallard, R. Cisneros, M. Murooka, and F. Kanehiro, "The Kinetics Observer: A Tightly Coupled Estimator for Legged Robots," *hal preprint*, Dec. 2024.

[29] R. Grandia, E. Knoop, M. Hopkins, G. Wiedebach, J. Bishop, S. Pickles, D. Müller, and M. Bächer, "Design and control of a bipedal robotic character," in *Robotics: Science and Systems XX*, ser. RSS2024. Robotics: Science and Systems Foundation, Jul. 2024.

[30] M. Tsuru, A. Escande, I. Kumagai, M. Murooka, and K. Harada, "Online multi-contact motion replanning for humanoid robots with semantic 3d voxel mapping: Exoctomap," *Sensors*, vol. 23, no. 21, 2023.

[31] M. Benallegue, G. Lorthioir, A. Dallard, R. Cisneros-Limón, I. Kumagai, M. Morisawa, H. Kaminaga, M. Murooka, A. Andre, P. Gergondet, K. Kaneko, G. Caron, F. Kanehiro, A. Kheddar, S. Yukizaki, J. Karasuyama, J. Murakami, and M. Kamon, "Humanoid robot rhp friends: Seamless combination of autonomous and teleoperated tasks in a nursing context," *IEEE Robotics and Automation Magazine*, vol. 32, no. 1, pp. 79–90, March 2025.

[32] K. Kaneko, H. Kaminaga, T. Sakaguchi, S. Kajita, M. Morisawa, I. Kumagai, and F. Kanehiro, "Humanoid robot hrp-5p: An electrically actuated humanoid robot with high-power and wide-range joints," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1431–1438, April 2019.

[33] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2018, pp. 7244–7251.