

VALINOR: a Lightweight Leg Inertial Odometry for Humanoid Robots

Arnaud Demont*^{ID}, Mehdi Benallegue^{ID}, Thomas Duvinage^{ID}, and Abdelaziz Benallegue^{ID}

Abstract: This article presents VALINOR (Velocity-Aided Leg Inertial Nonlinear Odometry and Registration), a method for lightweight yet accurate Leg-Inertial odometry for humanoid robots. VALINOR builds upon the Tilt Observer, a complementary filter that estimates the IMU's tilt and linear velocity with strong mathematical convergence guarantees, associating it with Leg odometry. We introduce an axis-agnostic fusion method that coherently combines the Tilt Observer's tilt estimate with the Leg odometry's yaw. We argue that this method is less arbitrary and more mathematically sound than those based on other orientation representations, especially on Euler angles. We validate VALINOR on real-world experiments with two humanoid robots. Results show that, while being 7.5 times faster than the state-of-the-art method used for comparison, VALINOR achieves estimation performance on par with it. This makes VALINOR a strong candidate for feedback in balance control, as well as for task planning and execution within the robot's local workspace. Beyond this contribution, we hope to highlight the broader need for estimation methods that are not only accurate and efficient, but also grounded in strong mathematical guarantees, towards their safe deployment around humans.

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, 5], reinforcement learning-based controllers [6, 7], and foundation models [8, 9] are becoming more prevalent due to their versatility and to their ability to handle long-horizon planning, complex dynamics. 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 [9–12]. 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. 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. [13, 14]. Such methods show remarkable accuracies. However, they also tend to be more and more computationally expensive (e.g. the order of the millisecond for [14]).

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. 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*, often under the as-

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.

sumption that accelerations are negligible compared to the gravitational acceleration (e.g. [15]). Unfortunately, this assumption cannot hold in the case of humanoids, especially 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. By modeling the contacts as fixed points in the environment, partial measurements of the robot’s kinematics in the world can be obtained from joint encoders. This consideration alone makes the tilt fully observable, however dynamic the motion is [16]. 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 [17]. This technique can be merged with IMU data and correct for the drift [18, 19]. This is called leg-inertial odometry.

Violations of slippage-free contact assumption can degrade the odometry performance. Work has thus been done to address the problem of contact slippage, for example by discarding unreliable contacts [20, 21], or by attempting to correct the contact reference position [18, 22, 23]. Despite that, the position and yaw orientation of the robot in the world frame remain unobservable [18]. 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 [24] or when task planning is also made in the robot’s local environment [25], 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 [16]. This complementary filter allows for a much faster computation than the conventionally used Kalman Filter (e.g. [22]), 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 estimator is 7.5 times faster than the state-of-the-art method for proprioceptive odometry, while achieving odometry accuracy on par with it.
- 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.

- We provide a ROS wrapper¹ for the open-source Tilt Observer².

2. GENERAL NOTATIONS

- The general notation for kinematic variables is ${}^1\bigcirc_2$, expressing the kinematics of a frame 2 in a frame 1. To simplify notation, kinematics in the world frame are written without the \mathcal{W} symbol: ${}^{\mathcal{W}}\bigcirc_2 = \bigcirc_2$.
- We note $\hat{\mathbf{x}}$ the estimate of the variable \mathbf{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.

3. DEFINITION OF THE ANCHOR POINT

We define here the notion of anchor point, since it will be used extensively in the following sections. It is notably essential for the design of the Tilt Observer, as explained in Section 4.2.

The anchor point, denoted \mathcal{A} , is defined as a point attached to the robot that is considered to have zero *instantaneous* 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 relative to 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, a contact is more likely to slip when contact forces are closer to Coulomb’s friction cones.

To address these two requirements we determine the kinematics of the anchor point through a weighted average of the kinematics of the n_c current contacts, as shown in Figure 1b. The weighting coefficients are defined such that weaker contacts, which are prone to violate Coulomb’s inequality and thus to slip, contribute less to the anchor point’s kinematics computation. Given mg_0 the weight of the robot, we first compute for each contact i the ratio u_i of the normal force to the tangential force, measured by the force / torque sensor at the contact. The term ε , arbitrarily small, allows to deal with the case where the tangential

¹https://github.com/ArnaudDmt/state_observation_ros.git

²<https://github.com/ArnaudDmt/state-observation>

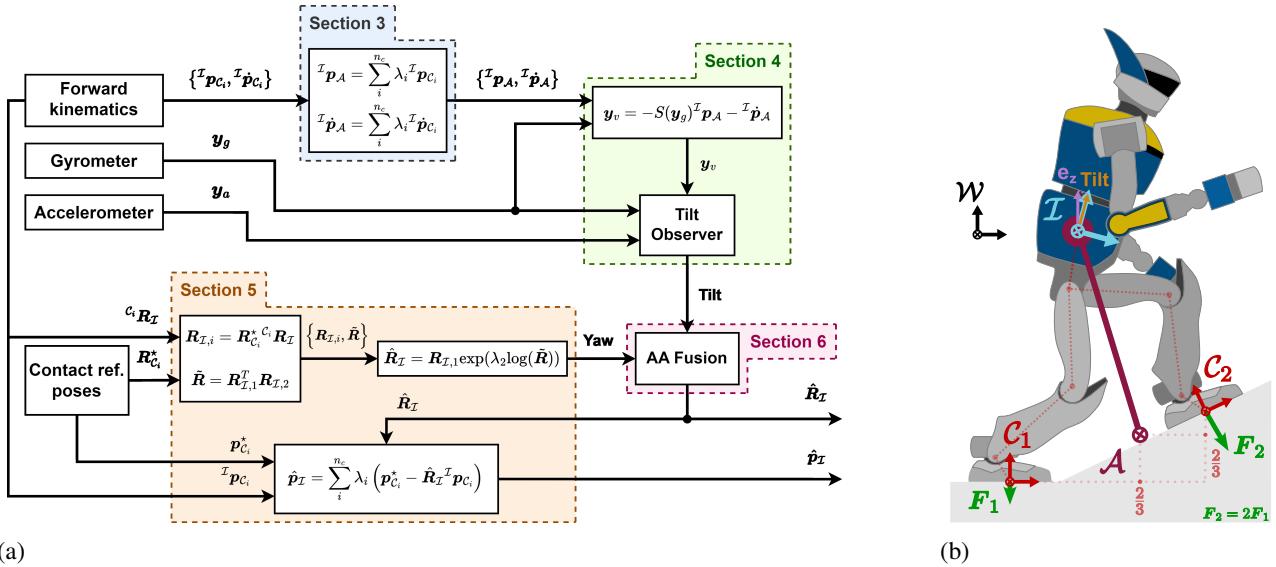


Fig. 1. (a) Summary of VALINOR's pipeline. (b) Illustration of the anchor point position computation and of the reference frames used in VALINOR. \mathcal{W} : world frame; \mathcal{I} : IMU frame; \mathcal{C}_i : frame of the i -th contact. For clarity, only the weighted average of the contact positions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

force is zero. The weighting coefficient λ_i of the contact in the average is then computed:

$$u_i = \frac{F_{i,z}}{\sqrt{F_{i,x}^2 + F_{i,y}^2} + \varepsilon mg_0}, \quad \lambda_i = \frac{u_i}{\sum_{j=1}^{n_c} u_i}. \quad (1)$$

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

$$\mathcal{I}^T \mathbf{p}_{\mathcal{A}} = \sum_i^{n_c} \lambda_i \mathcal{I}^T \mathbf{p}_{C_i}, \quad \mathcal{I}^T \dot{\mathbf{p}}_{\mathcal{A}} = \sum_i^{n_c} \lambda_i \mathcal{I}^T \dot{\mathbf{p}}_{C_i}, \quad (2)$$

with $\mathcal{I}^T \mathbf{p}_{C_i}$ the position and $\mathcal{I}^T \dot{\mathbf{p}}_{C_i}$ 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.

$\mathcal{I}^T \mathbf{p}_{\mathcal{A}}$ and $\mathcal{I}^T \dot{\mathbf{p}}_{\mathcal{A}}$ are updated at each iteration. As explained in Section 4.2, by combining them with the zero-velocity assumption of the anchor point in the world frame and the gyrometer measurement, we obtain an algebraic estimate of the IMU velocity in the world frame.

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

4. TILT OBSERVER WITH PROOF OF CONVERGENCE

VALINOR relies on a highly accurate estimate of the IMU's tilt provided by a complementary filter, introduced in [16], which we will call *Tilt Observer*.

4.1. Definition of the State Variables

The Tilt Observer provides estimates of:

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

We thus define the state variables:

$$\mathbf{x}_1 \triangleq \mathbf{v}_{\mathcal{I},l}, \quad \mathbf{x}_1 \in \mathbb{R}^3, \quad (3)$$

$$\mathbf{x}_2 \triangleq \mathcal{R}_{\mathcal{I}}^T \mathbf{e}_z, \quad \mathbf{x}_2 \in \mathbb{S}^2. \quad (4)$$

The set $\mathbb{S}^2 \subset \mathbb{R}^3$ is the unit sphere centered at the origin, and defined as $\mathbb{S}^2 = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| = 1\}$.

4.2. Definition of the Measurements

The measurements required by the Tilt Observer are:

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

Since no sensor provides a direct measurement of $\mathbf{v}_{\mathcal{I},l}$, we obtain \mathbf{y}_v from an algebraic combination of gyrometer signal \mathbf{y}_g and the position and velocity of the anchor point in the IMU's frame given by Eq. (2):

$$\mathbf{y}_v = -\mathbf{y}_g \times \mathcal{I}^T \mathbf{p}_{\mathcal{A}} - \mathcal{I}^T \dot{\mathbf{p}}_{\mathcal{A}} \quad (5)$$

4.3. Definition of the Filter

The state dynamics of the system can be written as:

$$\dot{\mathbf{x}}_1 = -\mathbf{y}_g \times \mathbf{x}_1 - g_0 \mathbf{x}_2 + \mathbf{y}_a, \quad (6)$$

$$\dot{\mathbf{x}}_2 = -\mathbf{y}_g \times \mathbf{x}_2, \quad (7)$$

with g_0 the gravitational acceleration constant. A complementary filter can then be defined, using the system's

dynamics as a feed-forward, and correcting the estimated state using the velocity measurement \mathbf{y}_v :

$$\begin{cases} \dot{\hat{\mathbf{x}}}_1 = -\mathbf{y}_g \times \hat{\mathbf{x}}_1 - g_0 \hat{\mathbf{x}}'_2 + \mathbf{y}_a + \alpha_1 (\mathbf{y}_v - \hat{\mathbf{x}}_1), \end{cases} \quad (8)$$

$$\begin{cases} \dot{\hat{\mathbf{x}}}'_2 = -\mathbf{y}_g \times \hat{\mathbf{x}}_2 - \frac{\alpha_2}{g_0} (\mathbf{y}_v - \hat{\mathbf{x}}_1), \end{cases} \quad (9)$$

$$\begin{cases} \dot{\hat{\mathbf{x}}}_2 = -(\mathbf{y}_g - \gamma \hat{\mathbf{x}}_2 \times \hat{\mathbf{x}}'_2) \times \hat{\mathbf{x}}_2. \end{cases} \quad (10)$$

$\hat{\mathbf{x}}_1$ is the estimate of \mathbf{x}_1 and $\hat{\mathbf{x}}'_2$ is an intermediate estimate of \mathbf{x}_2 , which converges exponentially, but is not restricted to lie on the sphere \mathbb{S}^2 . This estimate drives the convergence of the tilt's final estimate $\hat{\mathbf{x}}_2$, which is explicitly projected onto \mathbb{S}^2 so as to respect its Lie group structure. The positive scalar gains α_1 , α_2 and γ define the convergence rates of the three state estimates. Insights on their tuning are provided in [26]. Note that the filter's convergence guarantees hold regardless of the gains' tuning.

4.4. Advantages of the Tilt Observer

Using a complementary filter for the Tilt Observer offers key advantages over common approaches such as the Kalman Filter. First, it naturally operates in the frequency domain, which is particularly suitable for our model since the assumption of fixed contacts used to obtain the velocity measurement \mathbf{y}_v is more valid at low frequencies. In Eq. (8), IMU measurements thus provide the high frequency feed-forward of $\hat{\mathbf{x}}_1$, while \mathbf{y}_v ensures its low frequency correction. Second, one iteration of the filter is computationally extremely cheap since it only consists in computing three equations, which involve no matrix multiplication or inversion and is thus computationally extremely cheap. 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 [16] that:

- The dynamics of the estimation error is autonomous, and thus does not depend on the motion of the robot.
- The intermediate estimator $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}'_2\}$ 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. When a contact i is created, its *reference pose* $\{\mathbf{p}_{C_i}^*, \mathbf{R}_{C_i}^*\}$ in the world is obtained via forward kinematics from the estimated IMU frame pose

$$\{\hat{\mathbf{p}}_{\mathcal{I}}, \hat{\mathbf{R}}_{\mathcal{I}}\} :$$

$$\mathbf{R}_{C_i}^* = \hat{\mathbf{R}}_{\mathcal{I}}^T \mathbf{R}_{C_i}, \quad (11)$$

$$\mathbf{p}_{C_i}^* = \hat{\mathbf{p}}_{\mathcal{I}} + \hat{\mathbf{R}}_{\mathcal{I}}^T \mathbf{p}_{C_i}, \quad (12)$$

with $\{\mathbf{p}_{C_i}^*, \mathbf{R}_{C_i}^*\}$ the pose of the i -th contact in the IMU's frame provided by the robot's geometrical model and joint encoders. The contact's reference pose is considered constant to enforce its anchoring role, and is 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 the contacts to the leg odometry is weighted to favor those less prone to slippage, mitigating its effect.

First, we estimate the IMU's orientation from contact information. For the two most reliable contacts³, we compute the IMU frame's orientation from the contact's reference orientation. For each contact i , we write:

$$\mathbf{R}_{\mathcal{I},i} = \mathbf{R}_{C_i}^{* \mathcal{C}_i} \mathbf{R}_{\mathcal{I}}, \quad (13)$$

with $\mathcal{C}_i \mathbf{R}_{\mathcal{I}}$ 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:

$$\tilde{\mathbf{R}} = \mathbf{R}_{\mathcal{I},1}^T \mathbf{R}_{\mathcal{I},2} \quad (14)$$

$$\hat{\mathbf{R}}_{\mathcal{I}} = \mathbf{R}_{\mathcal{I},1} \exp(\lambda_2 \log(\tilde{\mathbf{R}})). \quad (15)$$

\exp and \log are the SO(3) *exponential* and *logarithm* maps. Note that for small angles $\log(\tilde{\mathbf{R}}) \simeq \frac{1}{2}(\tilde{\mathbf{R}} - \tilde{\mathbf{R}}^T)$.

Once the IMU's orientation $\hat{\mathbf{R}}_{\mathcal{I}}$ 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 and its mathematical convergence guarantees with little computations. The obtained orientation is then used to compute the IMU frame's position, together with the n_c contact reference poses:

$$\hat{\mathbf{p}}_{\mathcal{I}} = \sum_i^{n_c} \lambda_i (\mathbf{p}_{C_i}^* - \hat{\mathbf{R}}_{\mathcal{I}}^T \mathbf{p}_{C_i}). \quad (16)$$

Doing so, our position estimate thus takes full profit from the accurate tilt provided by the Tilt Observer.

6. TILT-YAW FUSION

The Tilt Observer of Section 4 provides $\hat{\mathbf{x}}_2$, an accurate estimate of the IMU *tilt* noted $\ell = \mathbf{R}_1^T \mathbf{e}_z$, while the

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

contact-based Leg odometry of Section 5 supplies an orientation $\mathbf{R}_2 = \hat{\mathbf{R}}_{\mathcal{I}}$, whose yaw reflects the robot’s heading. We now merge these two pieces into a single rotation \mathbf{R} , which would preserve our tilt estimate, especially its mathematical convergence guarantees.

6.1. Why not Euler angles?

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

$$\mathbf{R}_{\text{cl}} = R_z(\psi_2)R_y(\theta_1)R_x(\phi_1). \quad (17)$$

This solution has two properties

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

Despite its simplicity, Eq. (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 \mathbf{e}_x only. This makes sense when \mathbf{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, as in VALINOR where yaw information comes from Leg odometry (Section 5), not from that axis. Treating \mathbf{e}_x as “special” therefore introduces a systematic bias.
- Because (17) uses a ZYX decomposition of \mathbf{R}_1 , it inherits the Euler-angle singularity at $\theta_1 = \pm 90^\circ$. Exactly there the pitch axis aligns with \mathbf{e}_z , roll and yaw collapse, and the horizontal projection of \mathbf{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 ℓ , allows any horizontal vector to encode yaw, and remains regular for all configurations except the physically undefined case $\ell = -\mathbf{e}_z$.

The construction is detailed next.

6.2. Yaw without Euler angles

The yaw contained in \mathbf{R}_2 can be extracted without decomposing the matrix into Euler angles. Let \mathbf{r}_2 denote the third column of \mathbf{R}_2 . This column describes the world coordinates of the body z -axis, i.e $\mathbf{r}_2 = \mathbf{R}_2 \mathbf{e}_z$, which is not the tilt of \mathbf{R}_2 . It is used to build the horizontal vector \mathbf{m}_2 :

$$\mathbf{m}_2 = \begin{cases} \mathbf{e}_x, & \text{if } r_{2,x}^2 + r_{2,y}^2 \simeq 0, \\ \frac{1}{\sqrt{r_{2,x}^2 + r_{2,y}^2}} [r_{2,y} \ 0]^\top, & \text{otherwise.} \end{cases} \quad (18)$$

where $r_{2,x}$ and $r_{2,y}$ are the first and second components of \mathbf{r}_2 . The condition $r_{2,x}^2 + r_{2,y}^2 \simeq 0$ occurs when the tilt $\mathbf{R}_2^T \mathbf{e}_z$

is vertical (upward or downward), so we take $\mathbf{m} = \mathbf{e}_x$ to avoid the singularity. Otherwise, the vector \mathbf{m}_2 always exists and is unique.

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

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

$$\theta = \text{atan2}\left(m_{2,x}\tilde{m}_{2,y} - m_{2,y}\tilde{m}_{2,x}, m_{2,x}\tilde{m}_{2,x} + m_{2,y}\tilde{m}_{2,y}\right), \quad (19)$$

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 $\mathbf{R}_2^T \mathbf{e}_z$ is vertical, any horizontal vector is invariant. In such a case, taking $\mathbf{m} = \mathbf{e}_x$ amounts at mimicking Euler yaw angle. When this tilt is downward, meaning \mathbf{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 $\mathbf{m}_2 = \mathbf{e}_x$ 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 ℓ and the second matrix \mathbf{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 subtle than just identifying the yaw of a single matrix, because the two rotations could have a very different tilt. This could lead the tilt vector ℓ to be nearly colinear with the rotated horizontal invariant vector $\mathbf{R}_2^T \mathbf{m}_2$. 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 $\mathbf{r} = \mathbf{R}_2 \ell = [v_x \ v_y \ v_z]^\top$. Consequently, a proper reference vector can be defined as

$$\mathbf{m} = \begin{cases} \mathbf{m}_2, & \text{if } v_x^2 + v_y^2 \simeq 0, \\ \frac{1}{\sqrt{v_x^2 + v_y^2}} [v_y \ -v_x \ 0]^\top, & \text{otherwise.} \end{cases} \quad (20)$$

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

6.3.2 Build two right-handed bases.

Using cross products, form

$$\mathbf{B}_1 = [\mathbf{m} \times \mathbf{e}_z \ \mathbf{e}_z \times \mathbf{m} \times \mathbf{e}_z \ \mathbf{e}_z], \quad (21)$$

$$\mathbf{B}_2 = \begin{bmatrix} \mathbf{m}_l \times \ell & \ell \times \mathbf{m}_l \times \ell & \ell \end{bmatrix}. \quad (22)$$

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 $\{\mathbf{m}_2, \cdot, \ell\}$ to $\{\mathbf{m}, \cdot, \mathbf{e}_z\}$:

$$\mathbf{R} = \mathbf{B}_1 \mathbf{B}_2^\top. \quad (23)$$

6.4. Properties

The rotation obtained with Eq. (23) has the following properties:

- *Tilt preservation*: $\mathbf{R}^\top \mathbf{e}_z = \ell$.
- *Optimal horizontal alignment*: for any rotation $\hat{\mathbf{R}}$ satisfying $\hat{\mathbf{R}}^\top \mathbf{e}_z = \ell$, the choice (23) minimises $\|\hat{\mathbf{R}}^\top \mathbf{m} - \mathbf{m}_l\|^2$. The proof follows exactly the one-parameter maximisation used for the Euler fusion, but with the task-specific direction \mathbf{m} in place of the body axis \mathbf{e}_x .
- *Singularity*: the construction is undefined only when $\ell = -\mathbf{e}_z$; 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 \mathbf{e}_x , takes into account the difference of tilt and avoids the gimbal-lock issues of Eq. (17).

7. EXPERIMENTAL EVALUATION

We identified the following criteria that a state estimator should satisfy to serve in our intended use case, namely, providing feedback for remote control [24] or task planning and execution in the robot's local workspace [25]:

Criterion ① High computation speed, to leave more processing time for the rest of the pipeline.

Criterion ② Accurate tilt and local lateral (horizontal) velocity estimation, which are crucial for stabilization.

Criterion ③ Low local drift (over small displacements) in translation and yaw. This would correlate with lower absolute drift, but the latter cannot be used as a metric because of its lower determinism by proprioceptive odometry.

The experiments presented in this Section assess VALINOR based on these criteria. We compare it with the Right-Invariant EKF (RI-EKF) from [22], 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 for remote control [24]. It is therefore particularly relevant to compare ourselves to this method.

7.1. Description of the experiments

VALINOR and the RI-EKF have been evaluated across two experimental scenarios on two different humanoid robots⁴:

- Experiment 1: A walk on a flat ground over about 18 meters with the robot RHP Friends [27], as shown in Figure 2a. 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 [28]. This motion involved an additional contact at the robot's left hand, and non-coplanar contacts on tilted obstacles, as shown in Figures 2b and 2c. This experiment was repeated 4 times for a total distance of about 8 meters.

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 (involving three contacts), to that of the RI-EKF. Both estimators were run on the same laptop, equipped with an Intel Xeon E-2176M CPU (2.70GHz x 12). In average, an iteration of VALINOR was computed in 2.547 μ s, against 19.315 μ s for the RI-EKF. The proposed method is thus more than 7.5 times faster than the state-of-the-art estimator. Across all sequences, the average computation times had a standard deviation of 0.043 μ s, indicating consistency from one sequence to another. We also note that an additional contact in VALINOR implies only minor additional computations in Eqs. (1), (2), and (16), making our method highly scalable with respect to the number of contacts.

7.2.2 Estimation Accuracy Evaluation

In both experiments, the odometry performance of VALINOR and of the RI-EKF are evaluated using the ground truth trajectory of the robot's pelvis (to which the IMU is fixed on both robots), provided by a motion capture system (OptiTrack, 16 Prime^X with 13 cameras). The pose of the IMU's frame, estimated by both estimators, is transformed into that of the pelvis using simple rigid body transformation. For fairness in the comparison, both estimators receive the same contact state information, from a Schmitt Trigger on the Ground Reaction Force (with thresholds set to 10% and 15% of the robot's weight).

⁴This dataset was also used in [23]

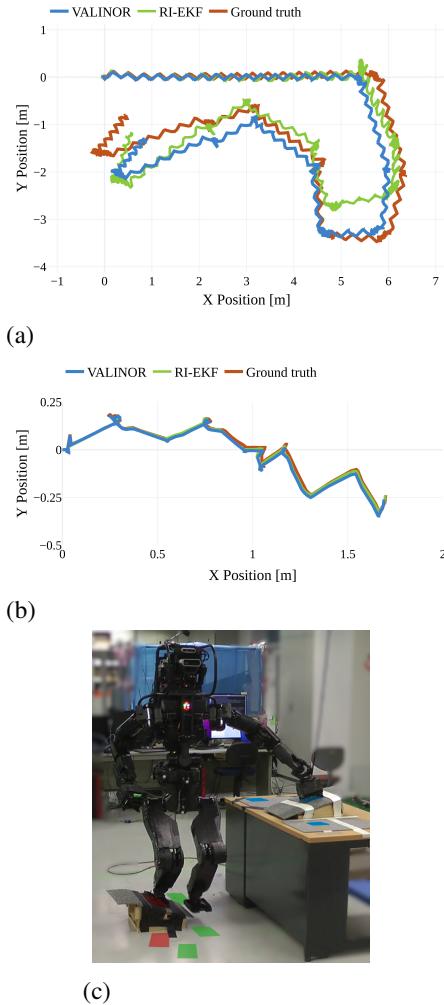


Fig. 2. Top views of the ground truth / estimated robot's pelvis trajectories during: (a) the first walk on flat ground (b) the multi-contact motion (c) HRP-5P performing the multi-contact motion. Contact imprints are represented by colored trapezoids. In red: Right foot. In green: Left foot. In blue: Left hand.

Since the focus is on accuracy in local translations and yaw (Criterion ③), performance is assessed using the Relative Error (RE), as defined in [29]. The latter evaluates the estimation drift over segments of fixed distances. It is particularly relevant here, as it provides valuable and easily interpretable insight into the estimator's expected local accuracy without being affected by global drift. The error e_ℓ on the tilt estimate is computed using $e_\ell = \arccos(\ell_{gt}^T \hat{\ell})$, with ℓ_{gt} the ground truth tilt, and $\hat{\ell}$ the estimated one. Finally, local linear velocity estimation is assessed using the Mean Absolute Error between the estimated velocity and the ground truth, the latter obtained by finite differences from the ground-truth position, filtered with a zero-phase low-pass filter.

Table 1 summarizes the evaluation results during the

two experiments on real robots. The presented errors are computed over all sequences in each scenario: 5 for the walk on flat ground, and 4 for the multi-contact motion. Based on the obtained results, we observe that the relative strengths of both estimators were consistent across the two experiments. In the following, we detail these strengths, indicating the number of the criterion to which each is favorable in a circle.

- VALINOR estimated local lateral translations ③ more accurately than the RI-EKF in both experiments, improving the estimation by over 30% and 40%, respectively. Additionally, as visible in Figure 3, it improved the estimation of the tilt ② 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 local lateral velocities ② and yaw ③ more accurately than VALINOR during Experiment 1, with 20% and 16% smaller errors, but both estimators displayed comparable performance in both estimates during Experiment 2.
- Both estimators showed similar performance in vertical translation ③ and velocity estimation.

To elaborate on the vertical translation results, one can notice in Figure 3 that both estimators drifted during the walk on flat ground, which involved a large number of steps. One contributing factor is the quality of the pitch estimation, an error in pitch estimation causing drift during forward walking. Another factor, inherent to Leg odometry, is its reliance on the quality of the contact detection. For example, systematically detecting contacts too early or too late can cause drift on each step, notably due to the robot's structural deformation under contact forces. In general use, this vertical drift could be reduced by tuning the contact detection threshold, which is arbitrarily set. However, to ensure a fair comparison, we kept the threshold unchanged and left this point for discussion. Moreover, absolute drift is not a reliable metric for proprioceptive odometry (see Criterion ③).

To conclude, both VALINOR and the RI-EKF displayed relative strengths with respect to each other, with regard to the criteria we defined. Overall, they provided highly satisfactory estimation performance which makes them well suited for local task planning and execution. Without the use of any exteroceptive sensors, their errors in translation estimates remain of the order of the centimeter per meter walked, those in orientation estimates remain around 1° per meter walked, and finally those in velocity estimates remain below 2 cm.s^{-1} .

It is noteworthy that, while reducing the computation time by more than a factor 7.5, VALINOR could consistently improve tilt and lateral translation estimation significantly.

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	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 [22]	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	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 [22]	0.012 (0.016)	0.004 (0.006)	0.57 (1.02)	0.47 (0.65)	0.009 (0.009)	0.005 (0.007)

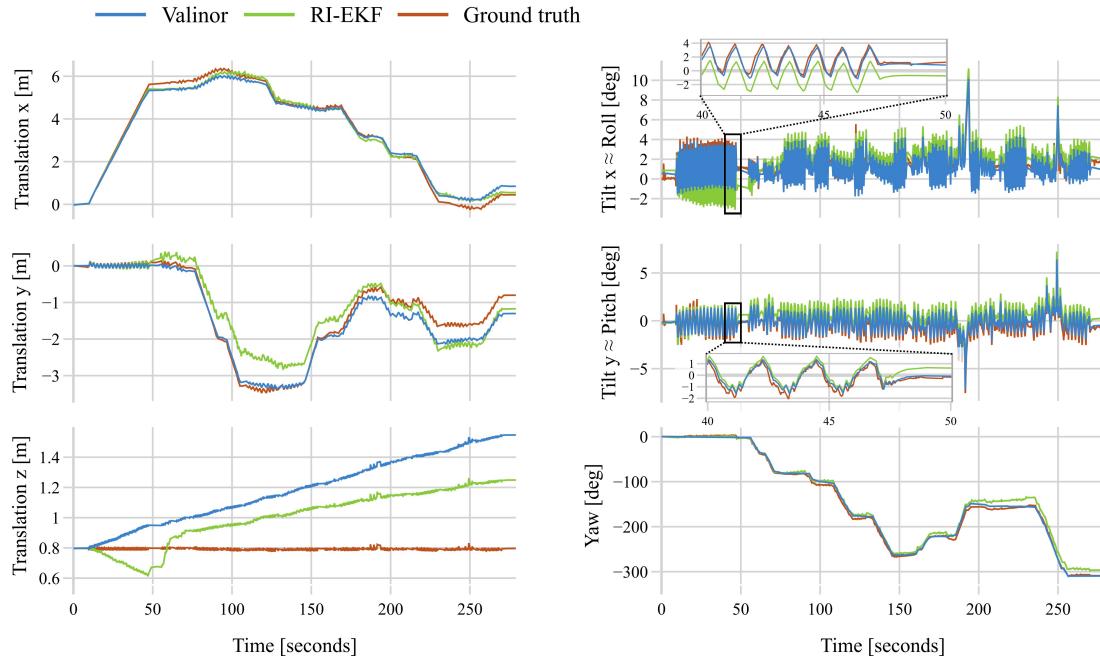


Fig. 3. Estimated and ground truth pose of RHP-Friends during the first walk on flat ground. Zoomed-in views are provided for the dense tilt plots, with a black rectangle indicating the corresponding area in the original plot.

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 proprioceptive odometry whose accuracy rivals with the state-of-the-art method, 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.

We note that by relying on Leg odometry and thus on the assumption of fixed contacts in the world, position and yaw estimation would drift from the actual ones in case of contact slippage [18], as visible in Figure 2a. Our method is particularly well suited to our intended use case of high-

speed feedback for control in a local task space, but for tasks requiring minimal drift in absolute pose, methods using exteroceptive measurements [13, 14], although slower, would of course be more appropriate.

Further work would aim at reducing the sensitivity to slippage and other failure by improving the coupling between the Inertial and the Leg odometry, while preserving high computational efficiency. Other extensions could consider the incorporation of other sensors into the framework. Finally, we could investigate the handling of biases in the measurements, which is known to be a considerable challenge in proprioceptive odometry, either in theoretical terms or practical performances.

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 developed the proposed estimator's theoretical foundations. Abdelaziz Benallegue supervised the study and validated the theoretical framework. Arnaud Demont and Thomas Duvinage conducted the experiments. Arnaud Demont implemented the estimator 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] A. Dallard, M. Benallegue, N. Scianca, F. Kanehiro, and A. Kheddar, “Robust bipedal walking with closed-loop mpc: Adios stabilizers,” *IEEE Transactions on Robotics*, vol. 41, pp. 4679–4698, 2025.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] R. Findeisen and F. Allgöwer, “Computational delay in nonlinear model predictive control,” *IFAC Proceedings Volumes*, vol. 37, no. 1, pp. 427–432, 2004.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.

- [22] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizelle, “Contact-aided invariant extended kalman filtering for robot state estimation,” *The International Journal of Robotics Research*, vol. 39, no. 4, pp. 402–430, 2020.
- [23] 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,” Dec. 2024, working paper or preprint.
- [24] 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.
- [25] M. Tsuru, A. Escande, I. Kumagai, M. Murooka, and K. Harada, “Online multi-contact motion replanning for humanoid robots with semantic 3d voxel mapping: Exocetomap,” *Sensors*, vol. 23, no. 21, 2023.
- [26] M. Benallegue, A. Benallegue, R. Cisneros, and Y. Chitour, “Velocity-aided imu-based tilt and attitude estimation,” *IEEE Transactions on Automatic Control*, vol. 68, no. 10, pp. 5823–5836, 2023.
- [27] 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 hrp friends: Seamless combination of autonomous and tele-operated tasks in a nursing context,” *IEEE Robotics and Automation Magazine*, vol. 32, no. 1, pp. 79–90, March 2025.
- [28] 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.
- [29] 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.



Arnaud Demont received the M.S. degree in mechanical engineering from the National Institute of Applied Sciences of Lyon, France, and a second M.S. degree in automation and robotics in intelligent systems from the University of Technology of Compiègne, France, in 2021 and 2023 respectively. He is currently pursuing the PhD degree of the Université Paris-Saclay, France, within the CNRS-AIST Joint Robotics Laboratory in Tsukuba, Japan. His research interests include state estimation for legged robots, multi-sensor fusion, and mobile robot perception and autonomous navigation.



Mehdi Benallegue obtained his engineering degree from the National Institute of Computer Science (INI) in Algeria, in 2007. He earned his MSc from the University of Paris 7, France, in 2008, and a Ph.D. from the University of Montpellier 2, France, in 2011. His research took him to the CNRS AIST Joint Robotics Laboratory (JRL), Japan, and to INRIA Grenoble, France. He also worked as a postdoctoral researcher at the Collège de France and at LAAS CNRS, France. Currently, he is a Senior Researcher with JRL, Japan. His research interests include robot estimation and control, legged locomotion, biomechanics, neuroscience, and computational geometry.



Thomas Duvinage received the M.S. degree in computer science with a specialization in robotics and computer vision from the University of Technology of Belfort-Montbéliard, France. He also holds a second M.S. degree in industrial and technology entrepreneurship from the same university. Since 2024, he has been serving as a CNRS delegate research engineer at the CNRS-AIST Joint Robotics Laboratory in Tsukuba, Japan. His work focuses on software integration for robotic systems, with research interests including robotics software architecture, system integration, and real-time robotic applications.



Prof. Abdelaziz Benallegue received the B.S. degree in electronics engineering from Algiers National Polytechnic School, Algeria in 1986 and both the M.S. and Ph.D. degrees in automatic control and robotics from University of Pierre and Marie Curie, Paris 6 (currently Sorbonne University), France in 1987 and 1991 respectively. He was Associate professor in Automatic Control and Robotics at the University Pierre et Marie Curie, Paris 6 (currently Sorbonne University) from 1992 to 2002. In September 2002, he joined the University of Versailles St Quentin as full Professor assigned. He was a CNRS delegate at JRL-AIST, Japan for three years, between 2016 and 2022. His research activities are mainly related to linear and non-linear estimation and control theory (adaptive control, robust control, neural learning control, observers, multi-sensor fusion, etc.) with applications in robotics (humanoid robots, aerial robots, manipulator robots, etc.).

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.