2024 I-RIM Conference
October 25-27, Rome, Italy
ISBN: 9788894580556
10.5281/zenodo.14731124

Multi-Sensor Fusion for Quadruped Robot State Estimation on Challenging Terrain

Ylenia Nisticò
Dynamic Legged Systems
Istituto Italiano di Tecnologia
University of Genoa
Genoa, Italy
ylenia.nistico@iit.it

João Carlos Virgolino Soares
Dynamic Legged Systems
Istituto Italiano di Tecnologia
Genoa, Italy
joao.virgolino@iit.it

Geoff Fink
Dynamic Legged Systems
Istituto Italiano di Tecnologia
Thompson Rivers University
Kamloops, Canada
gefink@tru.ca

Claudio Semini
Dynamic Legged Systems
Istituto Italiano di Tecnologia
Genoa, Italy
claudio.semini@iit.it

***Abstract*—This paper presents the preliminary results of MUSE, a MUlti-sensor State Estimator for quadruped robots, combining data from IMU, encoders, and cameras to accurately estimate pose and velocity, even in challenging environments such as uneven terrain. Experiments on a Unitree Aliengo robot, tested on stairs, rocks, and slippery surfaces, show MUSE's superior performance compared to using only a T265 tracking camera, providing reliable and high-frequency state estimation.**

***Index Terms*—legged robots, sensor fusion, localization**

I. Introduction

Accurate state estimation is crucial for enabling quadruped robots to perform robust locomotion, because it provides essential information about the robot's position, orientation, and velocity in the environment. This research introduces MUSE (MUlti-sensor State Estimator), a state estimator tailored for quadruped robots, based on [1]. MUSE integrates data from proprioceptive sensors (e.g., IMU, encoders) and exteroceptive sensors (e.g., cameras), combining their strengths to deliver high-frequency and low-drift state estimation even in challenging and unstructured environments. It also incorporates a slip detection module from [2] to further enhance performance in difficult terrain. The system was validated through offline evaluation in various indoor scenarios, using the Aliengo robot on uneven terrain, demonstrating its effectiveness in maintaining accuracy under real-world conditions.

II. MUSE Architecture

The main components of MUSE are shown in Fig. 1, comprising camera odometry, slip detection, leg odometry, an attitude observer, and a sensor fusion algorithm, that outputs the robot's pose and twist. MUSE receives as inputs the measurements from three proprioceptive sensor types (PS): IMU, encoders, force/torque sensors, and one exteroceptive sensor (ES), a camera.

A. Camera Odometry

We obtain visual odometry from an Intel Realsense T265 binocular visual-inertial tracking camera, already embedded in the robot, and equipped with an IMU and two fisheye lenses with 163° of field of view. The camera can provide pose and twist estimates at up to 200 Hz, used to update the robot's state.

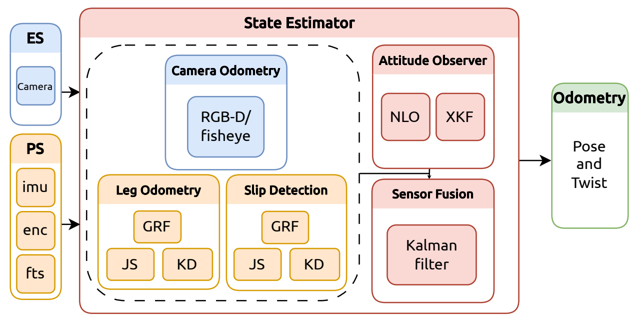


Fig. 1. Structure of the MUSE state estimator. The Slip Detection and Leg Odometry modules include Joint State (JS), robot Kinematics/Dynamics (KD), and Ground Reaction Forces (GRF). Together with the Camera Odometry, these modules send their outputs to the Attitude Observer, which includes a nonlinear observer (NLO) and an eXogenous Kalman Filter (XKF), and to the Sensor Fusion module, which utilizes a Kalman filter to estimate odometry.

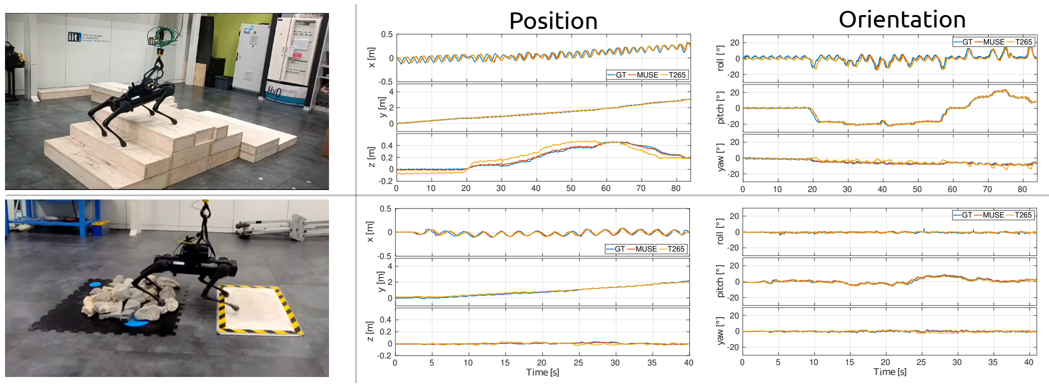
B. Leg Odometry and Slip Detection

Leg odometry estimates the linear velocity of the floating base from the forward kinematics of the legs in stable contact with the ground. The base velocity given by the leg odometry is $\dot{\mathbf{x}}^{\mathbf{b}} = \frac{1}{\mathbf{n_s}} \sum_{\ell}^{\mathbb{L}} \dot{\mathbf{x}}^{\mathbf{b}}_{\ell}$, where $\mathbb{L}$ is the set of legs, $\mathbf{n_s} = \sum_{\ell}^{\mathbb{L}} \alpha_\ell$ is the number of stance legs, and α_ℓ is a boolean that indicates if the leg ℓ is in contact with the ground ($\alpha_\ell = 1$). The contribution of each leg $\ell \in \mathbb{L}$ to the overall velocity of the base is $\dot{\mathbf{x}}^{\mathbf{b}}_{\ell} = -\alpha_\ell(\mathbf{J}_\ell(\mathbf{q}_\ell)\dot{\mathbf{q}}_\ell - \omega^{\mathbf{b}} \times \mathbf{x}^{\mathbf{b}}_{\ell})$, where $\mathbf{J}_\ell(\mathbf{q}_\ell)$ is the Jacobian of the leg ℓ, $\mathbf{q}_\ell$ is the joint position, $\dot{\mathbf{q}}_\ell$ is the joint velocity, $\omega^{\mathbf{b}}$ is the angular velocity of the base, and $\mathbf{x}^{\mathbf{b}}_{\ell}$ is the position of the leg ℓ in the base frame.

Furthermore, we used the slip-detection algorithm, presented in our previous work [2], to compensate for the drift that can occur in leg odometry when walking on uneven and slippery ground. Once slippage is detected, either in one or multiple legs, we increase the covariance of the leg odometry to prevent it from negatively affecting pose and velocity estimates.

C. Attitude Observer

To estimate the robot's orientation, a cascaded structure is used, combining a Nonlinear Observer (NLO) from [3], and an eXogeneous Kalman Filter (XKF) from [4]. The XKF linearizes based on the globally stable signal from the NLO,



ATE and RPE statistics over 1m

RMSE	ATE [m]	RPE [m]	RPE [°]
T265	0.089	0.081	7.792
MUSE	**0.045**	**0.077**	**6.529**

Camera ON

RMSE	ATE [m]	RPE [m]	RPE [°]
T265	0.097	0.100	2.151
MUSE	**0.096**	**0.092**	**1.897**

Camera OFF

RMSE	ATE [m]	RPE [m]	RPE [°]
NO SD	0.912	0.893	1.903
SD	**0.425**	**0.120**	**1.899**

Fig. 2. Aliengo Experiments: GT vs. T265 vs. MUSE positions. The Tables on the right show the translation ATE and the translation and rotational RPE - **Top**: **Walking up and down stairs**. MUSE corrected the significant error in the z-position of the T265 estimate. The table on the right confirms the improvements in ATE and RPE. **Bottom**: **Walking over uneven and slippery terrain**. In this experiment, starting from $\approx 6.5s$ at least one of the legs was always in contact with uneven or slippery terrain. On the right, the upper table shows the results of comparing the T265 and MUSE estimates. The lower table compares the proprioceptive MUSE estimates without and with the slip detection module.

ensuring global stability [4] while preserving the near-optimal performance of the Kalman Filter.

The observer state $\mathbf{x} = [\mathbf{q}^\top \mathbf{b}^\top]^\top \in \mathbb{R}^\mathbf{7}$ includes the quaternion $\mathbf{q} \in \mathbb{R}^\mathbf{4}$ and the gyroscope bias $\mathbf{b} \in \mathbb{R}^\mathbf{3}$, while the measurement vector is computed using the orientation provided by the camera in the sensor frame $\mathcal{S}$, to rotate a vector that is constant in the navigation frame $\mathcal{N}$, in the body frame $\mathcal{B}$. The measurement vector is $\mathbf{z} = [\mathbf{f_b}^\top \mathbf{m_b}^\top]^\top \in \mathbb{R}^\mathbf{6}$ where $\mathbf{f_b} = \mathbf{R_i^b f_n} \in \mathbb{R}^\mathbf{3}$ is the force given by the accelerometer in $\mathcal{B}$, and $\mathbf{m_b} = \mathbf{R_s^b}[\mathbf{1}\ \mathbf{0}\ \mathbf{0}]^\top \in \mathbb{R}^\mathbf{3}$, is the "pseudo-magnetometer measure", in which $[\mathbf{1}\ \mathbf{0}\ \mathbf{0}]^\top$ is constant in $\mathcal{N}$ and pointing to a "pseudo" North. It is important to underline that camera odometry is used in place of magnetometer data, because the last is often unreliable in quadruped robots due to interference from motors and metal objects. Finally, the complete equations of the XKF and NLO are in [1].

D. Sensor Fusion

The inertial measurements are fused with the leg odometry and the camera odometry in a Kalman Filter (KF), with the dynamic described in [1]. The state $\mathbf{x} = [\mathbf{x^n}^\top \mathbf{v^n}^\top]^\top \in \mathbb{R}^\mathbf{6}$ is the position and velocity of the base, the input $\mathbf{u} = (\mathbf{R_b^n f^b} - \mathbf{g^n}) \in \mathbb{R}^\mathbf{3}$ is the acceleration of the base, and $\mathbf{z} = [\mathbf{R_b^n \dot{x}_\ell^b}^\top\ \mathbf{R_b^n \dot{x}_c^b}^\top\ \mathbf{R_b^n x_c^b}^\top]^\top \in \mathbb{R}^\mathbf{9}$ is the vector of measurements, given by the velocity from the leg odometry, and the velocity and position from the camera odometry. In our formulation, $\mathbf{K} \in \mathbb{R}^{\mathbf{6\times 9}}$ is the Kalman gain, $\mathbf{P} \in \mathbb{R}^{\mathbf{6\times 6}}$ is the covariance matrix, $\mathbf{Q} \in \mathbb{R}^{\mathbf{6\times 6}}$ is the process noise, and the measurement noise covariance matrix $\mathbf{R}$ is a diagonal block matrix, assuming that the measurements are uncorrelated: $\mathbf{R} = \mathbf{diag}(\mathbf{R_1}, \mathbf{R_2}, \mathbf{R_3})$ where $\mathbf{R_1} \in \mathbf{R}^{\mathbf{3\times 3}}$ is the covariance of the leg odometry, and its values are updated in case of slippage, while $\mathbf{R_2} \in \mathbf{R}^{\mathbf{3\times 3}}$ and $\mathbf{R_3} \in \mathbf{R}^{\mathbf{3\times 3}}$ are the covariance of the camera velocity and position.

III. Experimental Results

We conducted two experiments using the Aliengo robot, with ground truth (GT) data collected with a *Vicon* motion capture system. Sensor frequencies are 250 Hz for the IMU and leg kinematics, and 200 Hz for the camera. MUSE operated at 250 Hz. The accuracy of the estimated trajectories is assessed using Absolute Trajectory Error (ATE) and Relative Pose Error (RPE), as shown in Fig. 2.

In the first experiment, Aliengo used a crawling gait to walk up and down stairs. The results demonstrated that the estimate combining leg kinematics was more accurate than using only the T265 camera, showing that leg kinematics improve the trajectory estimation over just visual and inertial data.

In the second experiment, Aliengo walked over uneven terrain, including a pile of rocks and a soap-covered plastic sheet. Here, the T265 camera odometry and slip detection compensated for the leg odometry drift, making the benefit of slip detection less apparent. To highlight its importance, the camera was deactivated, relying only on proprioception and slip detection. The results confirmed that slip detection significantly improved estimation when camera data was unavailable (last table in Fig. 2).

IV. Conclusion

MUSE integrates data from IMU, encoders, and camera, to provide reliable pose and motion estimation, particularly valuable for handling complex scenarios such as uneven and slippery terrains. Benchmarking against a Visual-Inertial SLAM, we demonstrated that adding leg-kinematics and slip detection allows to improve the performance and to reach high-frequency state estimation.

References

[1] G. Fink and C. Semini, "Proprioceptive sensor fusion for quadruped robot state estimation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10 914–10 920, DOI: 10.1109/IROS45743.2020.9341521.

[2] Y. Nisticò, S. Fahmi, L. Pallottino, C. Semini, and G. Fink, "On slip detection for quadruped robots," *Sensors*, vol. 22, no. 8, 2022, DOI: 10.3390/s22082967.

[3] H. F. Grip, T. I. Fossen, T. A. Johansen, and A. Saberi, "Globally exponentially stable attitude and gyro bias estimation with application to GNSS/INS integration," *Automatica*, vol. 51, pp. 158–166, 2015, DOI: 10.1016/j.automatica.2014.10.076.

[4] T. A. Johansen and T. I. Fossen, "The eXogenous Kalman filter (XKF)," *International Journal of Control*, vol. 90, no. 2, pp. 161–167, 2017, DOI: 10.1080/00207179.2016.1172390.