Heterogeneous Sensor Fusion for Accurate State Estimation of Dynamic Legged Robots

Simona Nobili*[1,3], Marco Camurri*[2], Victor Barasuol[2], Michele Focchi[2], Darwin G. Caldwell[2], Claudio Semini[2] and Maurice Fallon[3]

[1] School of Informatics, University of Edinburgh, UK

[2] Dept. of Advanced Robotics, Istituto Italiano di Tecnologia, Italy

[3] Oxford Robotics Institute, University of Oxford

***Abstract*—We will present a system for the state estimation of a dynamically walking and trotting quadruped. The approach fuses four heterogeneous sensor sources (inertial, kinematic, stereo vision and LIDAR) to maintain an accurate and consistent estimate of the robot's base link velocity and position. The work focuses on the complexity of consistent estimation of pose and velocity states, as well as the fusion of multiple exteroceptive signal sources at largely different frequencies and latencies, in a manner which is acceptable for a quadruped's feedback controller. A substantial experimental evaluation is demonstrated.**

I. Introduction and Related Work

For legged robots to be useful and eventually autonomous, they must be able to reliably walk and trot over a variety of terrains and in the presence of disturbances such as slips or pushes. They must also be able to perceive their environment and to avoid collisions with obstacles and people. Accurate and reliable state estimation is essential to achieve these capabilities.

The motivation for proprioceptive state estimation is somewhat different for legged control system. Notably, Bloesch et al. [1] presented a treatment of leg kinematics and IMU fusion focusing on consistency.

The work of Ma et al. [4] is most closely related to ours. Their system was designed to function as a modular sensor head fusing a tactical grade inertial measurement unit with stereo visual odometry.

II. Experimental Scenario

Our experimental platform is a torque-controlled Hydraulic Quadruped robot (HyQ, Figure 1). The system is $1\,\mathrm{m}$ long, and weighs approximately $85\,\mathrm{kg}$. A summary of the core sensors on the robot is provided in Table I. The robot's main exteroceptive sensor is the Carnegie Robotics Multisense SL which is composed of a stereo camera and a Hokuyo UTM-30LX-EW planar ranging laser (LIDAR).

*Both authors contributed equally to this work. A more complete version of this work appears in [5]

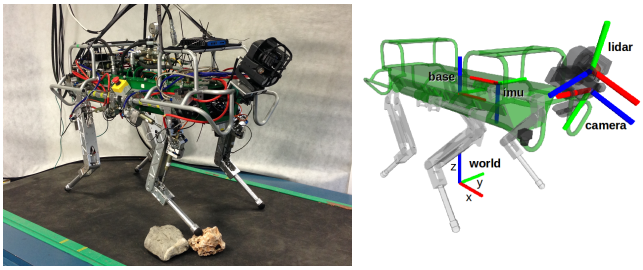

Fig. 1: *Left:* Hydraulic Quadruped robot (HyQ) with the Carnegie sensor head. *Right:* Main coordinate frames used in this paper.

Sensor	Sensor Freq.	Sensor Latency	Integration Freq.	Integration Latency	Variables Measured
IMU	1000	< 1	1000	n\a	${}_b\boldsymbol{\omega}_b$ ${}_b\ddot{\mathbf{x}}_b$
Joint Encoders	1000	< 1	1000	< 1	${}_b\dot{\mathbf{x}}_b$
LIDAR	40	10	0.2-0.5	600	${}_w\mathbf{x}_b$ ${}_w\boldsymbol{\theta}_b$
Stereo	10	125	10	42	${}_w\mathbf{x}_b$

TABLE I: Frequency (Hz) and latency (ms) of the main sensors.

III. Approach

We build upon an inertial-kinematic estimator recently described in [2]. In this section, we overview the core approach and use the same notation introduced therein. The 15 elements of the robot's base link state vector are defined by:

$$\mathcal{X} = \begin{bmatrix} {}_w\mathbf{x}_b & {}_b\dot{\mathbf{x}}_b & {}_w\boldsymbol{\theta}_b & \mathbf{b}_a & \mathbf{b}_\omega \end{bmatrix} \tag{1}$$

where the base velocity ${}_b\dot{\mathbf{x}}_b$, is expressed in the base frame b, while the position ${}_w\mathbf{x}_b$ and orientation ${}_w\boldsymbol{\theta}_b$ are expressed in a fixed world frame w (the list of frames and their location on HyQ is depicted in Figure 1). The orientation is expressed as quaternion, but the attitude uncertainty is tracked by the exponential coordinates of the perturbation rotation vector. The state vector is completed by IMU acceleration and angular velocity biases $\mathbf{b}_a$ and $\mathbf{b}_\omega$, which is updated by an EKF.

IMU biases are typically estimated when the robot is stationary and held static thereafter, as they are difficult to infer on a dynamic robot[1]. When operating, the robot drift of the yaw estimate is a significant issue.

A. Leg Odometry Module

Joint sensing contributes through a Leg Odometry (LO) module [2], which runs at $1\,\mathrm{kHz}$. During the filter update step, a measure for the base velocity ${}_b\dot{\mathbf{x}}_b$ is computed as a combination of the individual velocity measurements ${}_b\dot{\mathbf{x}}_{b_f}$ from each in-stance foot f, as follows

$${}_b\dot{\mathbf{x}}_{b_f} = -{}_b\dot{\mathbf{x}}_f - {}_b\boldsymbol{\omega}_b \times {}_b\mathbf{x}_f, \tag{2}$$

where ${}_b\dot{\mathbf{x}}_f$ and ${}_b\mathbf{x}_f$ are the velocity and position of foot f in the base frame, respectively.

As the robot is not equipped with contact sensors, we use the probabilistic contact classifier described in [2] to infer the

[1] In [4] the robot was commanded to stand still occasionally to back out rotation rate bias estimates.

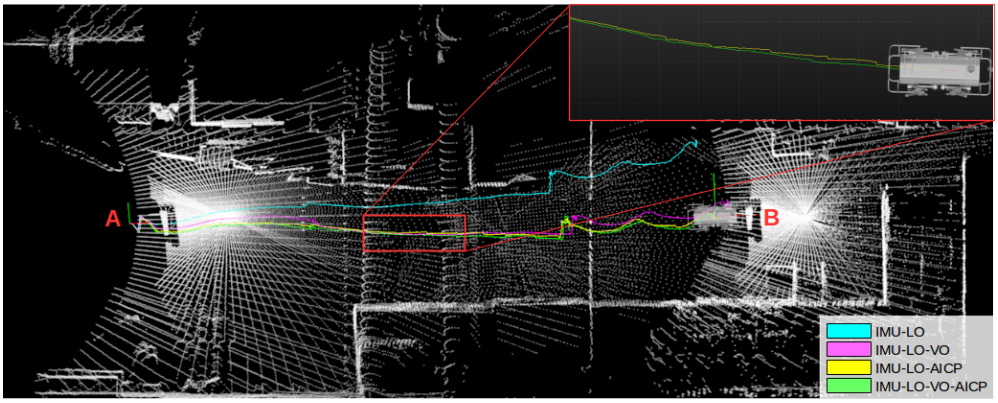

Fig. 2: A LIDAR map during a 17 m trot across the test arena. Also shown are the trajectories for the 4 estimator configurations discussed in this paper. The final combination (IMU-LO-VO-AICP) produces a smooth trajectory with continuously accurate localization (inset).

set of feet which are in stable and reliable contact. The core estimator drifts at 1-3 cm/m traveled, depending on gait.

B. Visual Odometry Module

Our visual odometry pipeline uses the open source implementation of FOVIS [3]. It tracks FAST features in a key-frame approach so as to estimate incremental camera motion, from image frame $k-1$ to frame k which we denote ${}_c\hat{T}_c{}^{k-1:k}$, where c indicates the camera frame. Using the known camera-to-body frame transformation, ${}_bT_c$, this can be expressed at the corresponding estimate of the motion of the body frame from $k-1$ to k as

$$ {}_b\hat{T}_b{}^{k-1:k} = {}_bT_c \; {}_c\hat{T}_c{}^{k-1:k} ({}_bT_c)^{-1} \tag{3} $$

Taking the posterior estimate of the EKF filter corresponding to time t_{k-1}, a measurement of the pose of the body at time t_k can be computed as follows:

$$ {}_w\hat{T}_b{}^{k} = {}_wT_b{}^{k-1} \oplus {}_b\hat{T}_b{}^{k-1:k} \tag{4} $$

C. LIDAR-based Localization

We use a strategy for non-incremental 3D scene registration, called Auto-tuned Iterative Closest Point (AICP) [6], to estimate the robot's pose. AICP extended a baseline ICP implementation to more reliably register point clouds with reduced overlap by automatically tuning an outlier-rejection filter to account for the degree of overlap of the sensors footprint. In our experiments, we could reliably register point clouds with only 11% overlap, which corresponded to a position offset of approximately 13 m.

The implementation of the filter maintains a history of measurements so as to enable asynchronous corrections with significant latency — specifically the ICP and VO corrections.

IV. Experimental Results

To validate the described system, we performed experiments in two different scenarios. A video to accompany this paper is available at `youtube.com/watch?v=k76tIJApJXI`

A. Experiment 1: Validation and Repeatability

The robot was commanded to continuously trot forward and backward to reach a fixed target. Robot position and velocity estimates are used by the controller to stabilize the robot motion while tracking the desired position. The results presented here show that incorporating VO reduces the drift rate relative to the baseline system, while adding AICP achieves localization relative to a fixed map.

B. Experiment 2: Comparing Variants in a Realistic Scenario

The robot explores a $20 \times 5\,\mathrm{m}^2$ industrial area. Lighting conditions vary dramatically during data recording, from bright light to strong shadows and from day to night-time. The map in Figure 2 shows the improvement in navigation performance with the addition of new sensor sources.

V. Discussion

The current system limitations are: a) the incremental error introduced by updates of the reference cloud, b) the frequency of the LIDAR sensor and resulting point cloud accumulation, and c) the susceptibility of the VO system to occasionally fail during short periods of poor lighting and the absence of visual features.

References

[1] M. Bloesch, M. Hutter, M. A. Höpflinger, S. Leutenegger, C. Gehring, C. D. Remy, and R. Siegwart. State Estimation for Legged Robots - Consistent Fusion of Leg Kinematics and IMU. In *Robotics: Science and Systems (RSS)*, Sydney, Australia, July 2012.

[2] M. Camurri, M. Fallon, S. Bazeille, A. Radulescu, V. Barasuol, D. G. Caldwell, and C. Semini. Probabilistic Contact Estimation and Impact Detection for State Estimation of Quadruped Robots. *IEEE Robotics and Automation Letters*, 2(2):1023–1030, April 2017.

[3] A.S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy. Visual odometry and mapping for autonomous flight using an RGB-D camera. In *Proc. of the Intl. Symp. of Robotics Research (ISRR)*, Flagstaff, USA, August 2011.

[4] J. Ma, M. Bajracharya, S. Susca, L. Matthies, and M. Malchano. Real-time pose estimation of a dynamic quadruped in GPS-denied environments for 24-hour operation. *Intl. J. of Robotics Research*, 35(6):631–653, May 2016.

[5] S. Nobili, M. Camurri, V. Barasuol, M. Focchi, D. Caldwell, C. Semini, and M. Fallon. Heterogeneous sensor fusion for accurate state estimation of dynamic legged robots. In *Robotics: Science and Systems (RSS)*, Cambridge, MA, August 2017.

[6] S. Nobili, R. Scona, M. Caravagna, and M. Fallon. Overlap-based ICP tuning for robust localization of a humanoid robot. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Singapore, May 2017.