INTERFACE FOCUS

rsfs.royalsocietypublishing.org

Research

Cite this article: Fallon M. 2018 Accurate and robust localization for walking robots fusing kinematics, inertial, vision and LIDAR. *Interface Focus* **8**: 20180015.
http://dx.doi.org/10.1098/rsfs.2018.0015

Accepted: 9 May 2018

One contribution of 12 to a theme issue 'Understanding images in biological and computer vision'.

Subject Areas:
mathematical physics

Keywords:
robotics, computer vision, sensing, estimation, locomotion

Author for correspondence:
Maurice Fallon
e-mail: mfallon@robots.ox.ac.uk

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

Accurate and robust localization for walking robots fusing kinematics, inertial, vision and LIDAR

Maurice Fallon

Oxford Robotics Institute, University of Oxford, Oxford, UK

MF, 0000-0003-2940-0879

In this article, we review methods for localization and situational awareness of biped and quadruped robotics. This type of robot is modelled as a free-floating mechanical system subject to external forces and constrained by whole-body distributed rigid contacts. Measurements of the state of the robot can be made using a variety of sensor information—such as kinematics (the sensing of the joint angles of the robot), contact force (pressure sensors in the robot's feet), accelerometers and gyroscopes as well as external sensors such as vision and LIDAR. This high-frequency state estimate is then passed to the control system of the robot to allow it to traverse terrain or manipulate its environment. In this article, we describe the development of an estimator for the Boston Dynamics Atlas humanoid robot. It was later adapted to the HyQ2 quadruped, developed by the Istituto Italiano di Tecnologia. Some discussion is given as to future trends while also considering briefly the relationship with biological systems.

1. Introduction

Dynamic locomotion of legged robotic systems remains an open and challenging research problem whose solution will enable humanoids to perform tasks in and reach places inaccessible to wheeled or tracked robots. Several research institutions are developing walking and running robots with a range of form factors, from all-terrain quadrupeds operating outdoors to experimental bipedal running machines being developed in the laboratory.

Within the control system, the role of the state estimator is to combine the sometimes disparate sensor measurements of the input and output of the system to create an overall estimate of the state (often including a measure of uncertainty) which is then used as input to the controller which commands the robot—namely, the 6 d.f. pose and velocity of its pelvis, as well as the configuration of its joints. An accurate and timely estimate of the robot state not only facilitates effective control for dynamic whole-body motions such as walking and manipulation, but also enables a greater degree of task autonomy by maintaining a consistent knowledge of the surrounding environment and the locations of objects within it.

1.1. Related work

One common class of estimation method is based on dynamics (e.g. [1]), and relies on knowledge of the controller outputs and a motion model of the humanoid to infer the state of the robot's centres of mass and pressure. Errors in link[1] centre-of-mass modelling and the presence of unpredictable forces (e.g. from the robot's support/power tether or external contacts) are accounted for by appending an additional process model for each class of disturbance.

Xinjilefu *et al*. [2] extend this approach and apply it to the Atlas robot (which we are also using in our work). They discuss the computational challenges of formulating a single extended Kalman filter (EKF) [3] for a humanoid with many degrees of freedom, and propose instead to estimate the pelvis position and joint dynamics in separate filters.

THE ROYAL SOCIETY PUBLISHING

© 2018 The Author(s) Published by the Royal Society. All rights reserved.

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

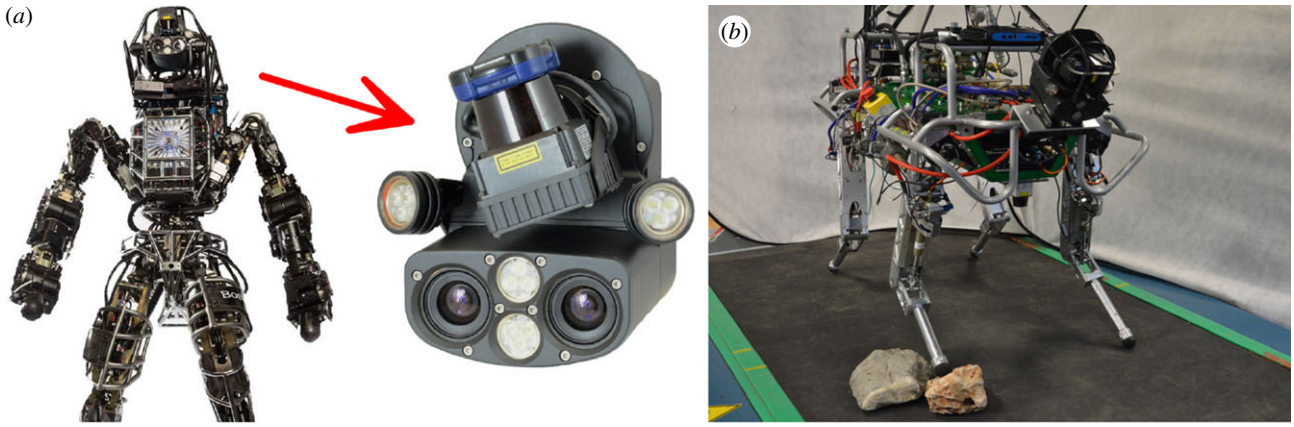

Figure 1. (*a*) Boston Dynamics' Atlas robot (2015 version) contains 28 hydraulically actuated joints. (*b*) IIT's HyQ2 robot contains four identical legs containing three hydraulic joints each. Both robots use the Carnegie Robotics Multisense SL sensor head which is equipped with a rotating LIDAR scanner and a stereo camera. (photo credits: Boston Dynamics, IIT and CRL). (Online version in colour.)

An EKF-based estimator is presented in [4] for a quadruped that uses a sensor-based prediction model and creates filter corrections using foothold measurements. This approach incorporates the positions of footholds into the state vector (using a point model for each foot) and gives particular consideration to consistency and observability analysis. Recently, this approach was extended to bipedal locomotion [5] with results presented on a simulated SARCOS robot. The primary contribution was extension of the algorithm to a biped with a full foot plate, which requires a 6 d.f. constraint on the foot frame.

Finally, there has also been work in coupling odometry estimates to a higher-level navigation system. Localization of a quadruped (in 3 d.f.) against a prior terrain map was explored by Chitta *et al*. [6] using a particle filter [7].

All of the above works use only proprioceptive sensing. While visual mapping is an active area of research—for instance Stasse *et al*. [8] demonstrated loop-closing and visual keyframe registration to known landmarks in a laboratory setting—it has not been widely adopted for field operations. High computational cost, latency and sensitivity to environmental conditions (e.g. illumination and visual texture content) all present substantial challenges for robust vision-based mapping onboard a humanoid. Hornung *et al.* [9] used a LIDAR sensor to localize an Aldebaran NAO robot within a multi-level environment, thus circumventing many problems inherent to visual sensing; our work also adopts LIDAR as the primary exteroceptive sensor for localization.

1.2. Overview

In this work, we review a state estimation algorithm, first presented in [10], that combines measurements from three distinct sensing modalities—inertial, leg kinematics and LIDAR—into a single consistent estimate of the robot's pelvis link via probabilistic fusion. The estimator does not require elaborate dynamics models; like [5], we couple foot placements and leg kinematics with inertial predictions in an EKF framework, and like [2], we consider pelvis pose separately from joint states. Our primary novel contributions are (i) incorporation of exteroceptive sensing to achieve reliable drift-free alignment to a prior map while walking using a Gaussian particle filter (GPF) to apply position corrections derived from each LIDAR scan and (ii) extensive experimental validation of the algorithm on a real humanoid robot.

The estimator is demonstrated using the Boston Dynamics (BDI) Atlas humanoid (figure 1) provided to our research team for the ongoing DARPA Robotics Challenge (DRC). The LIDAR-based localization component was specifically motivated by the slow locomotion rates achieved in the DRC Trials in December 2013: due to position drift of the order of 2 cm per footstep, teams typically took just two steps at a time to traverse uneven terrain, pausing periodically to manually relocalize the robot and to create new motion plans with respect to the environment. As we push towards enabling greater autonomy in task execution (e.g. walking for several minutes at a time with manipulation actions interspersed) a continuous localization capability becomes critically important, as it allows the robot to retain accurate and consistent reference to terrain maps and objects of interest in its vicinity.

In §2, we present an overview of our requirements for state estimation and discuss two different use cases for our approach. Then in §§3.1–3.3 we discuss how each sensor stream can be abstracted to a basic probabilistic measurement suitable for fusion.

Finally in §3.4, we benchmark the performance of the algorithm and present results from a series of extended duration walking experiments, executed passively with BDI's native controller totalling approximately 1 h. A clear operational benefit is demonstrated: alignment to the prior model enables the robot to continuously traverse uneven terrain without stopping, and thus operate up to four times more quickly than previously possible.

2. Requirements

The ultimate goal of our research efforts [11] is to develop a system that enables a humanoid robot to operate at a semi-autonomous level with human interaction at a task level, as depicted in figure 2: 'walk over to the drill and use it to cut a circular hole in the wall'.

Executing such actions requires the ability to precisely and continuously localize, thus enabling the robot to walk to a goal without interruption. An alternative would be to encode safety sequences such as stopping short of a goal and then stepping conservatively into position which, as well as being inefficient, adds complexity to an otherwise simple action.

To achieve this, we have developed a navigation system that comprises two concurrent state estimators within our

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

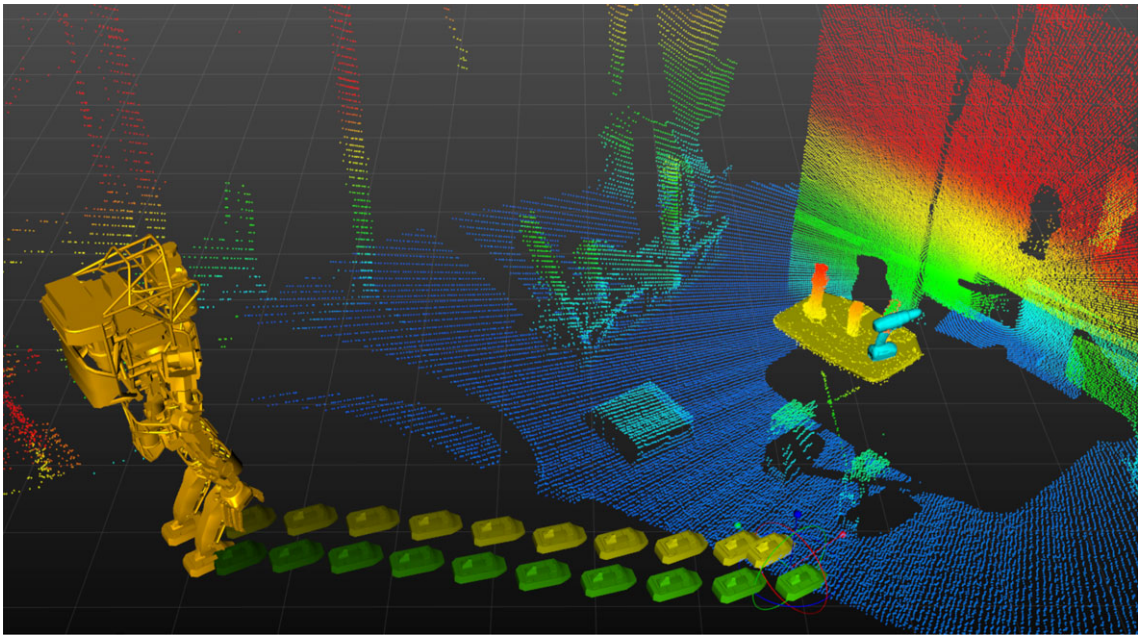

Figure 2. A three-dimensional rendering of the robot with a footstep plan leading towards a table to grasp a drill. (Online version in colour.)

closed-loop locomotion controller [12], produces a stable estimate of position and velocity with high rate (333 Hz) and low latency, and, crucially, is without any discontinuities such as would be produced by a map alignment correction.

However, proprioception is subject to drift and thus is not amenable to task-level autonomy, which requires constant long-term localization within the environment but can better tolerate small instantaneous adjustments. We therefore maintain a second localization estimate that incorporates exteroception data from LIDAR (or vision sensors) to remove global drift, but can allow discontinuous corrections of the robot's position. (These corrections would typically be 1 cm or less, and applied at the 40 Hz framerate of the LIDAR.) This two-tiered approach, which is in the spirit of [13,14], maintains reference to higher-level features of interest used by our robot/operator team. Here, we focus primarily on this second, exteroceptive, localization mode; however, in practice we use the same algorithms and software framework for both modes.

3. Biped state estimation

3.1. Integration

Both estimators use the same integrator algorithm, originally described in [15] and used for a micro aerial vehicle (MAV), with simple software configuration flags enabling or disabling the various input sensors.

Following the notation described therein, we wish to estimate the position and orientation of the robot's kinematic root link, the pelvis, as well as its linear and angular velocities. The full-state vector is defined as $x = [w_b^{\mathrm{T}}\ v_b^{\mathrm{T}}\ R\ \triangle^{\mathrm{T}}]^{\mathrm{T}}$ and each component is as follows:

— angular velocity, $w_{b,b} \in \mathbb{R}^3$
— linear velocity, $v_{b,b} \in \mathbb{R}^3$
— orientation, $R_{b,w} \in SO_3$
— position, $\triangle_{b,w} \in \mathbb{R}^3$.

Both velocity components are estimated in the (body) pelvis frame, while the position and orientation of the pelvis are expressed in a fixed world frame. We exclude both the robot's contact foot position and the joint states from this state vector and filter them separately (as do others, including Xinjilefu *et al*. [2]).

The pelvis of the Atlas humanoid contains the robot's primary Inertial Measurement Unit (IMU), located 9 cm behind the pelvis link position and rotated by 45°. The sensor is a KVH 1750-IMU composed of Fiber Optic Gyroscopes (FOG) and Micro Electromechanical Systems (MEMS) accelerometers of tactical grade.

The state estimate and its associated covariance are updated using an EKF. The prior distribution is propagated using a process model driven by the IMU measurements: the rotation rates w_b and the accelerations a_b are both sensed in the IMU frame and transformed to the body frame[2] before being integrated. As discussed in [15], orientation uncertainty is expressed in exponential coordinates around the body frame.

The IMU sensor provides very accurate raw measurements, but vibrations induced by the hydraulic pressurizer within the robot's torso corrupt the signal. We therefore apply a cascading set of IIR-notch filters to dampen the 85 Hz vibrational component and its harmonics.

Additional states of the EKF are used to maintain rotation rate and acceleration bias estimates, which are computed while the robot stands still at the start of an experiment. Although the estimator supports online bias updates, we typically retain these initial values as they tend to remain consistent during a typical experiment.

In the subsequent sections, we will describe how each individual sensing modality is used to form Kalman measurement updates to the state vector. This information is summarized in table 1.

3.2. Leg kinematics

The robot has two legs, each with six joints: three at the hip, a knee joint and two angle joints. As with many leg kinematic integration algorithms [1,4], our approach assumes that the robot's stance foot maintains non-slipping contact with the

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

Table 1. Contribution of various sensors to the filtered state estimate. Modes of integration found to be useful are marked ✓ and those not used here (for a variety of reasons) are indicated ✘.

	pos	orient	velocity	ang rate	accel
dimension	$\triangle$	R	v_b	w_b	a_b
accelerometers					✓
gyroscopes				✓	
leg sensing	✘	✘	✓	✘	
LIDAR	✓	✓			

ground during part of the gait and that this foot is stationary. This allows instantaneous velocity and position measurements of the robot's pelvis to be inferred via forward kinematics. Of course in practice perfectly clean and stable ground contact is seldom achieved, but we assert that for short periods (the sample time of our sensors) these assumptions are reasonable.

3.2.1. Contact classification

At the base of the locomotion algorithm, a gait transition detector infers the current stage of the walking motion and then decides which of the feet has stationary contact with the ground. We use a Schmitt trigger with a threshold of 575 N to classify contact forces sensed by the robot's 3-axis foot force-torque sensors and detect whether either foot is in contact. A simple state machine then decides which foot is the most reliably in contact and thus will provide the basis for kinematic measurements.

In the specific case of walking up stairs, the toe of the trailing foot can be used to push the robot's pelvis forward and upward while not being in stationary contact (known as 'toe off'). In this situation, we ensure that the leading foot is assigned to be the primary fixed foot.

We also classify other events in the gait cycle. Striking contact is determined when a rising force of 20–30 N is maintained for more than 5 ms. Breaking contact is determined when force falls below 275 N in the opposite direction. Because these events create unrealistic measurements, the EKF integrates these measurements with higher measurement covariance.

We note that when the robot is in a double support stance, information from both legs could be leveraged to provide additional kinematic measurements. For simplicity, we currently neglect this information.

3.2.2. Kinematic measurements

Given the accuracy of IMU orientation sensors, we choose to use joint sensing to measure the linear position and velocity of the pelvis only. Our input information is the current pelvis orientation estimate in the world frame $R^t_{b,w}$, the previous position of the stationary foot $\triangle^{t-1}_{f,w}$, and the position of the foot relative to the pelvis, as produced by forward kinematics $T^t_{f,b}$.

Dropping the time index, consider the pelvis located at the origin of a convenience coordinate frame aligned with the true world frame, $T_{p,w'} = \begin{bmatrix} R_{b,w} & 0_{3,1} \\ 0 & 1 \end{bmatrix}$. Within this convenience frame, the orientation of the foot is given by

$$T_{f,w'} = T_{b,w'} T_{f,b} = \begin{bmatrix} R_{f,w'} & \triangle_{f,w'} \\ 0 & 1 \end{bmatrix}. \tag{3.1}$$

By effectively assuming that the pelvis orientation estimate has zero covariance, the foot orientation in the world frame is then simply $R_{f,w} = R_{f,w'}$.

Finally, pairing this orientation with the fixed foot position $\triangle^{t-1}_{f,w}$, the position of the pelvis in the world frame can be recovered via

$$T_{p,w} = T_{f,w} T_{p,f} = T_{f,w} (T_{f,p})^{-1}, \tag{3.2}$$

where $T_{f,w} = \begin{bmatrix} R_{f,w} & \triangle_{f,w} \\ 0 & 1 \end{bmatrix}$.

Two types of filter measurement could be formulated using this position estimate. The simplest approach would be directly applying this as a position measurement within the EKF. However, because of the inconsistencies in joint sensing, and because the robot's foot does not make and maintain perfectly clean static contact with the ground, we avoid this approach.

Alternatively, we can use the difference between consecutive position estimates over a short period of time to create a velocity measurement of the pelvis frame

$$\hat{v}^t_b = \frac{\triangle^t_b - \triangle^{t-1}_b}{T_s}, \tag{3.3}$$

where T_s is the integration period, typically 3 ms.

This approach is more attractive because each resultant observation is a discrete measurement of the robot's velocity and does not retain any accumulated history (e.g. the effect of non-ideal ground contact, such as the footplate rolling or sliding). The influence of an erroneous velocity is transient and quickly corrected by subsequent observations.

Using both measurement types together would be comparable to the approach in [2], but we avoid doing so as it raises the possibility of creating inconsistencies, particularly when combined with position measurements derived from the LIDAR module (presented in the following section).

We note that our approach neither assumes knowledge of the terrain surface normal that robot is standing upon nor attempts to maintain a consistent estimate of the foot orientation over time.

Using VICON motion capture, typical pelvis velocity standard deviations, measured when standing still, are calculated as follows: (i) raw incoming kinematics: 7.6 cm s^{-1}; (ii) after joint level filtering: 2.3 cm s^{-1}; and (iii) after EKF integration: 1.4 cm s^{-1}.

3.3. LIDAR measurements

As illustrated in figure 1, the robot is equipped with a Multisense SL sensor head designed by Carnegie Robotics, which combines a fixed binocular stereo camera with a Hokuyo UTM-30LX-EW planar LIDAR sensor mounted on a spindle that can rotate up to 30 r.p.m. (We typically operate the device at 5 r.p.m. to densely sample the terrain when walking.) The LIDAR captures 40 scan lines of the environment per second, each containing 1081 range returns out to a maximum range of 30 m. The entire head can pitch up and down (powered by a hydraulic actuator) but cannot yaw or roll.

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

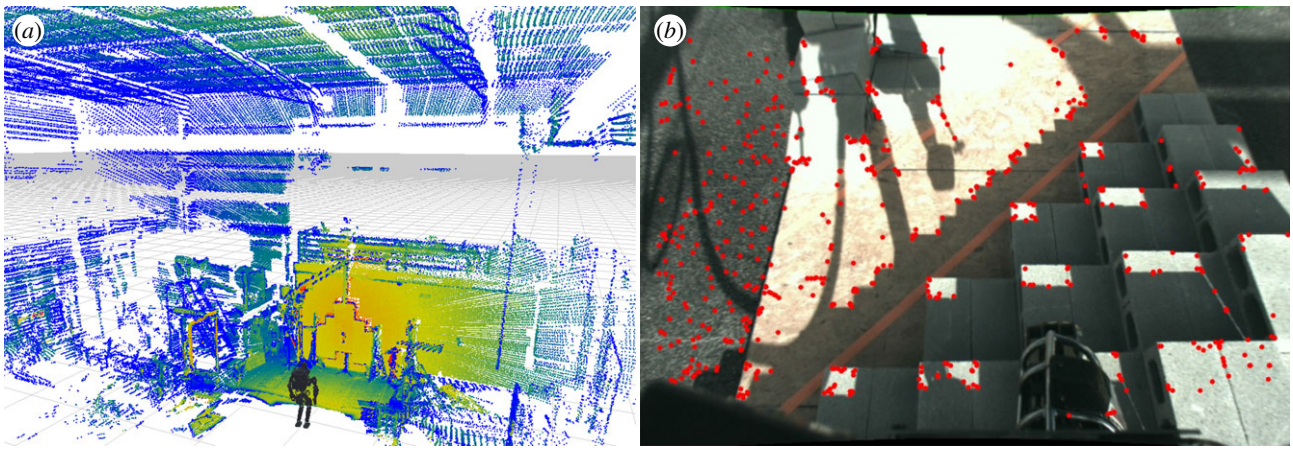

Figure 3. (*a*) For LIDAR localization the robot initially collects a static LIDAR point cloud of its environment, which is then converted into an occupancy map for subsequent localization. (*b*) The challenge of doing visual odometry in realistic field scenarios such as the DRC Trials, strong shadows and self-observations result in false visual features (shown in red). (Online version in colour.)

Our projection of LIDAR range returns as points in the three-dimensional workspace accounts for the robot's motion, and more importantly, the spindle rotation during the 1/40 s scanning period of the LIDAR's internal mirror. Neglecting this effect would result in misprojections of returns to the side of the robot by as much at 2.5 m at the highest spindle rotation speed. Accurate projection also requires precise calibration of the LIDAR sensor, as discussed in [11].

3.3.1. Contribution to estimation

Our strategy is to use the LIDAR to *continuously* infer the robot's position relative to a prior map while walking. We cannot assume that the sensor is oriented horizontally [16], nor can we afford time to stop moving and perform static three-dimensional registration (e.g. using an iterative closest point algorithm [17]). Instead we aim to incorporate information from each individual LIDAR scan into the state estimate using a GPF, as originally described in [15].

In typical operation, the robot is first commanded to stand still for between 10 and 30 s, while it collects a full three-dimensional point cloud of its environment (figure 3). This cloud is then converted into a probabilistic occupancy grid (OctoMap [18]) against which efficient localization comparisons are later performed. While the MAV experiments presented in [15] required offline mapping with a separate sensor, our legged humanoid and actuated LIDAR with 30 m range permit the map to be constructed immediately prior to operation and used during the entire task.[3] This makes our approach practical both for highly variable laboratory experiments and for field trials in which the environment is initially unknown. Furthermore, if the robot were to approach the map boundary, online construction of a new map could easily be performed *in situ*.

3.3.2. Latency and computation

To use this sensor modality in a real-time system requires careful consideration of latency. The LIDAR range measurements require significantly more time to be sensed and processed, which introduces significant latency relative to the 1 kHz kinematic and inertial information. These latencies are shown in table 2 for a 3.3 GHz 12-core desktop PC.

Table 2. Contribution of the various sensors to the state estimator indicating their computation latency and frequency.

component	latency	frequency
lower joint Kalman filters	0.16 ms	1 kHz
pose EKF	0.54 ms	333 Hz
LIDAR data transmission	7 ms	40 Hz
GPF processing time	11.4 ms	40 Hz
Overall LIDAR latency	18.4 ms	40 Hz

The values and the experiments presented in §3.4 use 1000 GPF samples, although reliable performance (and reduced latency) is possible with just 300 samples.

We use a multi-process messaging architecture to parallelize computation, with the GPF algorithm requiring a single CPU core. Within the estimator, the EKF retains a 1 s history of measurements to accommodate the LIDAR/GPF latency with the corrections made to the appropriate filter state followed by refiltering of all newer kinematic and inertial measurements.

3.3.3. Reliability and practicality

We also considered the reliability of each modality within the DRC competition context. A practical issue with the inertial sensor is that the robot must be completely stationary during initialization. BDI's estimator requires initialization as soon as the robot is first powered on, typically with its feet solidly contacting the ground. Our proposed estimator provides greater flexibility, as it can be initialized at any point when the robot is standing and deemed to be stationary.

While the time taken to construct the LIDAR map (10–30 s) is a minor inconvenience, in all datasets and task scenarios available to us there was sufficient stationary structure within sensor range for localization to operate reliably. Since the algorithm uses measurements of the entire environment, movement of a few objects or people near the robot has no notable effect on performance. However, in scenarios containing substantial background motion (such as the crowds attending the DRC Trials), special consideration may be required to disregard portions of the map with significant activity.

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

3.4. Biped robot experiments

In this section, we demonstrate the performance of the state estimator through a variety of experiments.

In each case our team's footstep planner (described in [11] and further developed in [19]) creates a kinematically feasible footstep sequence that reaches a goal position and orientation while minimizing the number of steps taken. The experiments involving uneven terrain traversal require human input to ensure correct initial footstep placement.

In our first set of experiments, the system sequentially feeds footsteps to the BDI walking controller (with its own internal state estimate) for execution while continually modifying their positions during execution so that the controller achieves the intended motion.

The kinematic-only estimates (both our own and the manufacturer's) are seen to continuously drift, typically at 1.2–1.5 cm per step. This drift rate generally increases when the locomotion is atypical: up inclines, more dynamic or with extended length steps. Orientation estimation performance is comparable across the different estimators. (Note that the precision of the ground truth orientation, via VICON, is of the order of 1°, so a precise comparison is not possible.)

Walking for just 10 min, the kinematic-only estimators drift by as much as a metre while the LIDAR-aided approach remains accurate to within 2 cm throughout.

A high-level overview video showing the robot operating using the estimator is available at https://www.youtube.com/watch?v=RuuZoChAtjQ.

4. Quadruped state estimation

Using the core estimator described above, subsequent research in collaboration with the Dynamic Legged Systems Group of the Istituto Italiano di Tecnologia worked to develop a suitable estimator for the HyQ2 robot shown in figure 1*b*.

While similar to Atlas in its level of dynamics and torque generation, a quadruped's foot makes only a point contact with the ground. This means that one cannot assume that the entirety of the foot is rigidly in contact and thus only partial information about contact is available at any time. As a result the focus on quadruped estimation is reliable dynamic contact estimation of these impacting feet.

4.1. Contact estimation

We define the contact status for a foot belonging to leg $l \in \{\mathrm{LF}, \mathrm{RF}, \mathrm{LH}, \mathrm{RH}\}$ as $S_l \in \{0, 1\}$, where 1 indicates a reliable stance (i.e. with no motion relative to the ground) and 0 indicates swing or slipping contact. The letters indicate combinations of left/right and front/hind. Let $\mathbf{f}_l$ be the GRF for leg l. This is either measured with some degree of uncertainty or, in our case, computed from the joint position $\mathbf{q}_l$, velocity $\dot{\mathbf{q}}_l$ and effort $\boldsymbol{\tau}_l$, as follows:

$$\mathbf{f}_l = -(J_l^T(\mathbf{q}_l))^{-1}(\tau_l - \mathbf{h}_l(\mathbf{q}_l, \dot{\mathbf{q}}_l, {}_b\mathbf{g})), \tag{4.1}$$

where $J_l^T(\mathbf{q}_l)$ is the Jacobian transpose that maps from joint to Cartesian space and $\mathbf{h}_l(\mathbf{q}_l, \dot{\mathbf{q}}_l, {}_b\mathbf{g})$ is the vector of centrifugal/Coriolis/gravity torques for leg l, computed using recursive Newton–Euler algorithms, as described in [20]. Given the small mass of HyQ2's legs compared with the torso, we assume the effect of inertial torques as negligible, compared to the numerical error of computing $\ddot{\mathbf{q}}$.

Given $\mathbf{f}_l = (f_{l,x}, f_{l,y}, f_{l,z})$ and following the definition from [21], the quantity

$$\mu_f = \frac{\sqrt{f_{l,x}^2 + f_{l,y}^2}}{f_{l,z}}, \quad \forall f_{l,z} > 0, \tag{4.2}$$

defines a metric to evaluate the robustness of a foothold, in terms of contact stability. This metric is equal to the actual static friction coefficient μ_s when the lateral components of the GRF (denoted with $f_{l,x}, f_{l,y}$) have a value beyond which the foot would start slipping. Although μ_s is unknown, any value of $\mu_f < \mu_s$ would yield a stable contact, and in particular, the smaller μ_f is, the more likely the foot is firmly on the ground. Hence, the quality of contact for a foot related to leg l at time k is nonlinearly proportional to $\mathbf{f}_l^k$. For simplicity and numerical stability, instead of accounting for all the terms of μ_s, we ignore the lateral components of $\mathbf{f}_l^k$, and assume that, above a certain threshold of $f_{l,z}^k$, the frictional force will be sufficient to produce a stable, reliable contact.

To learn this threshold, we model the probability of a reliable ground contact P_k using a discriminative logit model,

$$\mathrm{P}_k(S_l = 1 \mid \mathbf{f}_l^k) = \frac{1}{1 + \exp(-\beta f_{z,l}^k - \beta_0)}, \tag{4.3}$$

where $f_{z,l}^k$ is the normal component of the GRF at time k for leg l, while β and β_0 can be regarded as the weights of a logistic regression classifier. The weights are computed by maximum-likelihood estimation on a training set of data collected from characteristic motions, as described next.

4.1.1. Fitting with simulated data

We performed preliminary tests of our approach on data generated by simulation, with contact ground truth. The presented methodology was applied for two distinctive locomotion styles: a static crawl and a dynamic trot.

The crawl gait was obtained using a deliberate planning controller, as depicted in the experiments in [22]. The trot gait was generated using the reactive controller framework presented in [23]. In our experiments the movement was generated using a step frequency of 1.7 Hz, a duty factor of 0.5 and a leg stiffness of $8.55 \times 10^3\,\mathrm{N\,m^{-1}}$.

4.1.2. Fitting with real data

To test the classifier in a real scenario, we performed training on half of a trot log and half of a crawl log from our dataset, and we used the rest of the dataset as a test set for the learned model.

As no suitable commercial solution for contact sensing was available on our hardware, the ground truth for the training was defined as the time sequence of stance leg combinations that minimized the error between estimated and true base velocities. Additional post-processing was applied to maintain the continuity of swing and stance intervals. This approach also has the advantage that the classifier tends to learn the force threshold beyond which the associated velocity measurement produced by the foot in question becomes reliable.

Figure 4 displays, for a crawl gait, the GRF signal (top plot) and the fitting of the model against the ground truth for the test set (bottom plot). As in the simulation, the threshold for contact activation in the trot gait is higher, with a larger difference compared with the simulated gaits.

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

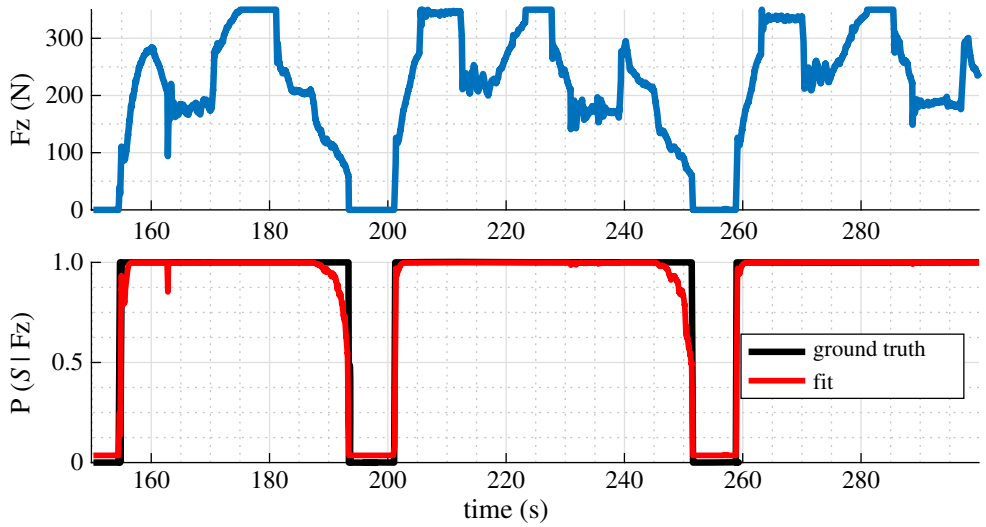

Figure 4. Crawl gait experiment. Learned stance probability and ground truth. (Online version in colour.)

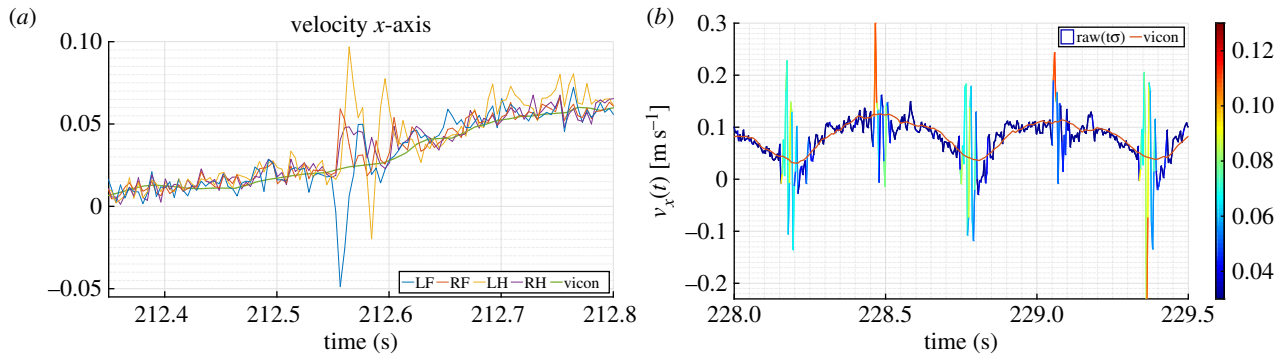

Figure 5. (*a*) Effect of impulsive force on estimated velocities during a crawl gait. The left hind (LH) leg strikes the ground at time 212.55 s, producing unrealistic velocity estimates for that leg—as well as for the other legs, due to propagation of the impact on the rest of the structure. (*b*) Raw velocity on *x*-axis compared with ground truth during a trot motion. The standard deviation associated with the velocity samples is shown with a colour scale, ranging from dark blue (0.03 m s^{-1}) to dark red (0.13 m s^{-1}). (Online version in colour.)

4.2. Velocity estimation

Given an estimate of which feet are likely to be in reliable contact, we compute a velocity estimate of the base ${}_b\dot{\mathbf{x}}_b$ and its associated covariance matrix $\Sigma_v = \mathrm{diag}(\sigma_x^2, \sigma_y^2, \sigma_z^2)$ using kinematic sensing. This will then be used as a measurement in the EKF update step. To compute the estimate we use the contact estimate introduced in §4.1, while to compute the covariance we leverage the knowledge about the consistency between the velocity contributions of the stance legs and the detection of impacts.

4.3. Velocity computation

To produce a base velocity update for the filter we combine the individual base velocity estimates produced by each leg. We use the probability of a given foot related to leg l being in contact at time k as a weighting criteria as follows:

$$ {}_b\dot{\mathbf{x}}_b(k) = \frac{\sum_{l\in C} \mathrm{P}_k(S_l = 1 \mid \mathbf{f}_l^k)\, {}_b\dot{\mathbf{x}}_{b_l}(k)}{\sum_{l\in C} \mathrm{P}_k(S_l = 1 \mid \mathbf{f}_l^k)}, \tag{4.4} $$

and C is the set of feet that exceed the 0.5 threshold of the logistic regressor. Among the feet that are detected to be in contact, weight is given to be proportional to the probability of contact.

4.3.1. Covariance estimation

Correctly estimating the covariance of these velocity contributions is particularly important. The robot executes different types of dynamic gaits and creates entirely unrealistic velocity updates when a foot strikes the ground.

To compute the covariance matrix Σ_v associated with each velocity update, we considered two factors: consistency between each contribution ${}_b\dot{\mathbf{x}}_{b_l}$ and impact forces. For each coordinate $r \in \{x, y, z\}$ we compute the corresponding variance at a given instant as

$$ \sigma_r^2(k) = \sigma_0^2 + (\alpha_1 \mathrm{std}\,[{}_b\dot{\mathbf{x}}_{b_{l\in C}}(k)]_c + (1-\alpha_1)\alpha_2 |\Delta \bar{f}_z^k|)^2, \tag{4.5} $$

where

$$ |\Delta \bar{f}_z^k| = \frac{1}{\dim(C)} \sum_{l\in C} |f_{z,l}^k - f_{z,l}^{k-1}|, \tag{4.6} $$

is the average of the absolute difference between the current and previous normal component of the GRF. We use this value as an indicator of an impact event. σ_0 is the baseline standard deviation for velocity, $\mathrm{std}\,[{}_b\dot{\mathbf{x}}_{b_{l\in C}}]$ is the rth component of the standard deviation of the velocity contributions among stance legs, α_1 is a factor that balances the effects of leg consistency and impacts (we use 0.5) and α_2 is a normalization factor, computed as the ratio between typical velocity error and $|\Delta \bar{f}_z|$ at the same instant.

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

The middle term of equation (4.5) incorporates the fact that legs deemed to be in contact should provide consistent estimates for the same base velocity. The last term takes into consideration the effect of impact forces, which propagate throughout the system and also affect legs that are already in contact (figure 5).

In figure 5*b*, we show an example of the adaptive covariance described in this section, on data from a trot log. We compare the raw (i.e. not yet processed by the EKF) base velocity computed from equation (4.4) and the ground truth, on the *x*-axis. The velocity is coloured proportionally to the standard deviation $\sigma_x(k)$ extracted from equation (4.5). Note the change of colour in the proximity of feet contact transitions and impacts, where the standard deviation is increased from $0.02\,\mathrm{m\,s^{-1}}$ up to $0.15\,\mathrm{m\,s^{-1}}$. During these intervals, the confidence in the velocity updates processed by the EKF is reduced.

This estimator is now commonly used to operate the HyQ2 robot as described in [24]. A video showing this work, as well as further extensions from [25] to include LIDAR and stereo vision, are shown in https://www.youtube.com/watch?v=k76tIJApJXI.

5. Discussion and future directions

The control bandwidth of a locomoting system is of the order of 10 Hz, which places a high performance bar on the estimator and allows little margin for error when introducing latency or delay in the filtering/estimation process. It is of little use to detect and correct for a misstep or stumble hundreds or thousands of milliseconds after it has occurred.

Much of the latitude for improvement lies in the development of novel contact sensors. In the upcoming EU H2020 THING project, we intend to develop novel sensing on the robot's foot to determine slip and sheer at high frequency (1 kHz). In addition, this detection would result in low-level reactive reflex with minimal latency, probably by leveraging reflex.

Low latency image sensing is also being considered. Event-based cameras or dynamic vision sensors (DVS) have been developed with sub-millisecond delay and very high frequency. DVS cameras work by measuring *changes* in illumination and report the pixels which have changed rather than a full texture image. Typically, these are the pixels where we would wish to focus our image processing. Currently, these cameras are prototypes with low resolution, but in the future we would like to explore carrying out robot velocity estimation with nearly no latency using such a camera.

A final topic of interest is the more general topic of introspection: that is detecting or inferring discord between different sensor modalities and using this to isolate failing subsystems within the estimator. Specifically, when a robot walks, it builds a terrain height map of its environment. However, when it walks upon on this terrain which has been mapped, we currently do not consider if the action achieved (the step) matches what we had expected (the map). More attention should be given to appropriately estimating and reacting to a foot passing through the ground surface, which would indicate mud, water or leaves. Similarly, we are interested in data-driven methods which use terrain material prediction based on learned material properties—drawing a physical connection between bushes and deep grass, for example, and the increased complexity of traversing that type of material.

5.1. Relationship to and inspiration from biology

This article has a specific engineering focus, but much of what has developed has been motivated or inspired by biology in one way or another. It is impossible to deny the connections and relationship between humanoid walking machines and the humans and animals which inspire them.

There is an aspiration to develop walking machines which are as efficient and as mechanically minimal as the human. The low cost of transport achieved by a human or an animal such as a dog (the amount of energy required to move the robot's mass a certain distance, by walking) is a key target which robotic systems fall some way short of achieving. It could be argued that walking robots are over-engineered—highly controlling pre-defined walking trajectories and motions, rather than embracing the natural dynamics and elastic material as humans do. In the author's opinion that is due to sub-optimal representations and inefficient algorithms to preclude rich and novel motions to be generated on the fly.

Our perception systems are also over-engineered—for example, building centimetre-accurate reconstructions of the terrain rather than conserving power and merely finding the portions of the terrain which are suitable to walk upon. This is often the case because it is easier to develop a one-size-fits-all three-dimensional mapping system rather than something tailored to the scenario. In addition, vision sensors which can be used reliable during the types of motion a walking robot generates—sharp accelerations and foot impacts—lack the robustness needed to accurately track the motion of the robot. Camera blur during these events is a particular problem, with basic feature tracking failing, causing visual mapping to fail. Our recent work builds a system from the bottom up, which expects these types of failures and instead focuses on the parsimonious use of visual motion cues *when suitable* [25].

These reflections are but a few thoughts about what is a complex problem, often underestimated by the average person. One thought which often occurs is that the development of general purpose robotic helpers working in the human vicinity will require robots with (i) a high centre of mass (to use human tools), (ii) a small support polygon (to fit within human spaces) and (iii) dexterity and compliance (for safe coexistence). All of these properties tend towards a robot dynamically balancing, and probably doing so on feet.

6. Conclusion

This article reviewed sensor fusion approaches for biped and quadruped robots for localization and floating base state estimation. There are a variety of other problems relevant to this domain which have not been discussed, such as joint torque estimation, contact force estimation and contact position estimation. Outside of sensor accuracy, computational complexity, observability analysis and modelling accuracy can have a big impact on performance both of the estimator and of the control system using the input from the estimator. We concluded by discussing some future challenges being explored by the community so as to overcome these estimation challenges.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Downloaded from https://royalsocietypublishing.org/ on 04 March 2022

Funding. M.F. is supported by a Royal Society University Research Fellowship.

Acknowledgements. This article corresponds to an abridged overview of various works cited in the above text [10,11,24,25]. It involves significant research carried out by the MIT DRC team, as well as the Oxford Dynamic Robot Systems Group. In particular, I would like to acknowledge the recent collaboration with Dr Marco Camurri at the Dynamic Legged Systems Group of Istituto Italiano di Tecnologia led by Dr Claudio Semini.

Endnotes

[1]A kinematic chain is made up of links and joints. The former are the physical components (such as a thigh or forearm) which rotate around joints such as a hip or elbow.

[2]The manufacturer-provided orientation estimate was not used but is compared with in §3.4.

[3]The DRC Trials terrain course was approximately 15 m in length.

References

1. Stephens BJ. 2011 State estimation for force-controlled humanoid balance using simple models in the presence of modeling error. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, *Shanghai, China*, *9–13 May 2011*, pp. 3994–3999. Piscataway, NJ: IEEE.
2. Xinjilefu X, Feng S, Huang W, Atkeson C. 2014 Decoupled state estimation for humanoids using full-body dynamics. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, *Hong Kong, China*, *31 May–7 June 2014*. Piscataway, NJ: IEEE.
3. Kalman R. 1960 A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45. (doi:10.1115/1.3662552)
4. Bloesch M, Hutter M, Hoepflinger MA, Leutenegger S, Gehring C, Remy CD, Siegwart R. 2012 State estimation for legged robots—consistent fusion of leg kinematics and IMU. In *Robotics: Science and Systems*, vol. VIII (eds N Roy, P Newman, S Srinivasa), p. 504. Cambridge, MA: MIT Press.
5. Rotella N, Bloesch M, Righetti L, Schaal S. 2014 State estimation for a humanoid robot. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*, *14–18 September 2014*, *Chicago, IL*. Piscataway, NJ: IEEE.
6. Chitta S, Vernaza P, Geykhman R, Lee DD. 2007 Proprioceptive localization for a quadrupedal robot on known terrain. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, *10–14 April 2007*, *Rome, Italy*, pp. 4582–4587. Piscataway, NJ: IEEE.
7. Doucet A, Godsill S, Andrieu C. 2000 On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **10**, 197–208.
8. Stasse O, Davison AJ, Sellaouti R, Yokoi K. 2006 Real-time 3D SLAM for a humanoid robot considering pattern generator information. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, *9–15 October 2006*, *Beijing, China*. Piscataway, NJ: IEEE.
9. Hornung A, Wurm KM, Bennewitz M. 2010 Humanoid robot localization in complex indoor environments. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (*IROS*), *18–22 October 2010*, *Taipei, Taiwan*. Piscataway, NJ: IEEE.
10. Fallon MF, Antone M, Roy N, Teller S. 2014 Drift-free humanoid state estimation fusing kinematic, inertial and LIDAR sensing. In *14th IEEE-RAS Int. Conf. on Humanoid Robots* (*Humanoids*), *18–20 November 2014*, *Madrid, Spain*, pp. 112–119. Piscataway, NJ: IEEE.
11. Fallon M *et al.* 2014 An architecture for online affordance-based perception and whole-body planning. *J. Field Robotics* **32**, 229–254. (doi:10.1002/rob.21546)
12. Kuindersma S, Permenter F, Tedrake R. 2014 An efficiently solvable quadratic program for stabilizing dynamic locomotion. *arXiv* 1311.1839.
13. Moore DC, Huang AS, Walter M, Olson E, Fletcher L, Leonard J, Teller S. 2009 Simultaneous local and global state estimation for robotic navigation. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, *12–17 May 2009*, *Kobe, Japan*. Piscataway, NJ: IEEE.
14. Marder-Eppstein E, Berger E, Foote T, Gerkey B, Konolige K. 2010 The Office Marathon: Robust navigation in an indoor office environment. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, *3–7 May 2010*, *Anchorage, Alaska*. Piscataway, NJ: IEEE.
15. Bry A, Bachrach A, Roy N. 2012 State estimation for aggressive flight in GPS-denied environments using onboard sensing. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, *14–18 May 2012*, *Saint Paul, Minnesota*, pp. 1–8. Piscataway, NJ: IEEE.
16. Dellaert F, Fox D, Burgard W, Thrun S. 1999 Monte Carlo localization for mobile robots. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, *10–15 May 1999*, *Detroit, Michigan*. Piscataway, NJ: IEEE.
17. Besl PJ, McKay ND. 1992 A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 239–256. (doi:10.1109/34.121791)
18. Wurm KM, Hornung A, Bennewitz M, Stachniss C, Burgard W. 2010 OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems. In *Proc. of the ICRA 2010 Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation*. See http://ais.informatik.uni-freiburg.de/publications/papers/wurm10octomap.pdf.
19. Deits RLH, Tedrake R. 2014 Computing large convex regions of obstacle-free space through semi-definite programming. In *Algorithmic foundations of robotics XI* (eds H Akin, N Amato, V Isler, A van der Stappen), pp. 109–124. Cham, Switzerland: Springer.
20. Featherstone R. 1987 *Robot dynamics algorithms*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
21. Hoepflinger MA, Hutter M, Gehring C, Bloesch M, Siegwart R. 2013 Unsupervised identification and prediction of foothold robustness. In *IEEE Int. Conf. on Robotics and Automation* (*ICRA*), *6–10 May 2013*, *Karlsruhe, Germany*, pp. 3293–3298. Piscataway, NJ: IEEE.
22. Havoutis I, Ortiz J, Bazeille S, Barasuol V, Semini C, Caldwell DG. 2013 Onboard perception-based trotting and crawling with the hydraulic quadruped robot (HyQ). In *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), *3–7 November 2013*, *Tokyo, Japan*, pp. 6052–6057. Piscataway, NJ: IEEE.
23. Barasuol V, Buchli J, Semini C, Frigerio M, De Pieri ER, Caldwell DG. 2013 A reactive controller framework for quadrupedal locomotion on challenging terrain. In *IEEE Int. Conf. on Robotics and Automation* (*ICRA*), *6–10 May 2013*, *Karlsruhe, Germany*, pp. 2554–2561. Piscataway, NJ: IEEE.
24. Camurri M, Fallon M, Bazeille S, Radulescu A, Barasuol V, Caldwell DG, Semini C. 2017 Probabilistic contact estimation and impact detection for state estimation of quadruped robots. *IEEE Robot. Autom. Lett.* **2**, 1023–1030. (doi:10.1109/LRA.2017.2652491)
25. Nobili S, Camurri M, Barasuol V, Focchi M, Caldwell D, Semini C, Fallon M. In press. Heterogeneous sensor fusion for accurate state estimation of dynamic legged robots. Paper presented at Robotics: Science and Systems, 12-16 July 2017, Cambridge, MA.