2021 IEEE International Conference on Robotics and Automation (ICRA 2021)
May 31 - June 4, 2021, Xi'an, China

State Estimation for Hybrid Wheeled-Legged Robots Performing Mobile Manipulation Tasks

Yangwei You*, Samuel Cheong, Tai Pang Chen, Yuda Chen, Kun Zhang,
Cihan Acar, Fon Lin Lai, Albertus Hendrawan Adiwahono, Keng Peng Tee

***Abstract*— This paper introduces a general state estimation framework fusing multiple sensor information for hybrid wheeled-legged robots performing mobile manipulation tasks. At the core of the state estimator is a novel unified odometry for hybrid locomotion which can seamlessly maintain tracking and has no need to switch between stepping and rolling modes. To the best of our knowledge, the proposed odometry is the first work in this area. It is calculated based on the robot kinematics and instantaneous contact points of wheels with sensor inputs from IMU, joint encoders, joint torque sensors estimating wheel contact status, as well as RGB-D camera detecting geometric features of the terrain (e.g. elevation and surface normal vector). Subsequently, the odometry output is utilized as the motion model of a 3D Lidar map-based Monte Carlo Localization module for drift-free state estimation. As part of the framework, visual localization is integrated to provide high precision guidance for the robot movement relative to an object of interest. The proposed approach was verified thoroughly by two experiments conducted on the Pholus robot with OptiTrack measurements as ground truth.**

I. INTRODUCTION

With the advancement of technologies in torque-controlled actuators, legged robots are progressively becoming a practical option for locomotion in addition to their wheeled counterparts, and related research is becoming more intense. Combining the best of both worlds – the agility of legged robots on rough/uneven terrain and the efficiency of wheeled robots on smooth terrain – hybrid wheeled-legged robots [1], [2], [3] have been developed recently as a versatile solution for complex environments and scenarios, including search and rescue, payload delivery, exploration and inspection. Motivated by this, our work in this paper was done on a dual-arm quadrupedal wheeled-legged Pholus robot, as shown in Fig. 1, and focused on the state estimation module targeted at mobile manipulation tasks.

A. Related work

State estimation is a method of estimating the internal state of a dynamic system based on available measurement data. It is indispensable in many applications, such as drones [4], unmanned vehicles [5], or Augmented reality(AR)/Virtual Reality(VR) [6]. Here, *state* refers to a set of physical quantities that completely describe the robot's movement over time, including position, attitude, velocity, angular velocity, and the orientation to describe the robot's behavior and motion.

The authors are all with Institute for Infocomm Research, Agency for Science, Technology and Research (A*Star), 138632, Singapore
* Corresponding author. Email:youy@i2r.a-star.edu.sg

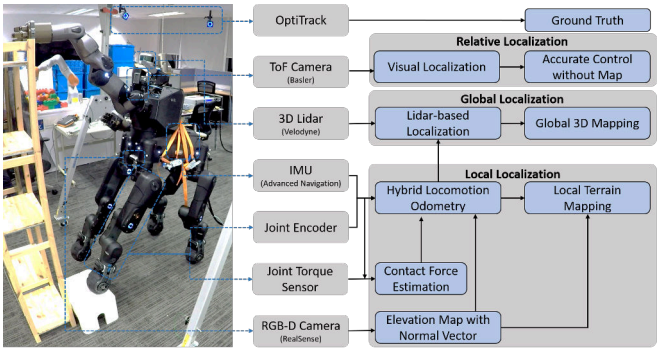


Fig. 1: The dual-arm wheeled quadrupedal robot Pholus tried to grasp a bottle. The proposed state estimation framework is shown on the right side linking to the sensor suit of Pholus.

State estimation on legged robots has been widely addressed in the literature[7], [8], [9]. A drift-free solution for humanoid robots was provided in [10]. For quadruped robots, accurate state estimation for dynamic high-speed gaits [11] and unstable terrains [12] were proposed. Besides legged robots, localisation on wheeled robots is well-studied using wheel odometry, Lidar and camera information [13], [14], [15]. Based on classic Lidar localization approaches, several works were proposed to improve the state estimation performance by utilizing unique structures such as planes and edges [16], [17]. These papers provide standalone methods of state estimation for either wheeled or legged robots, individually, but do not integrate them together. This paper aims to provide a combined state estimator for hybrid wheeled-legged robots that can seamlessly maintain tracking during transitions between walking and driving motions. To the best of our knowledge, this is the first state estimation method developed for hybrid wheeled-legged robots.

As part of the application for the proposed state estimator, this paper will also demonstrate its usage in mobile manipulation tasks. Unlike conventional manipulators, mobile manipulators have virtually unlimited workspace, enabling them to be used in diverse environments, including urban (e.g. homes, offices, hospitals), industrial and hazardous environments. With mobility and many Degrees of Freedom (DoFs) at their disposal, they are able to accomplish a wide range of tasks such as avoidance of obstacles, pushing, picking and placing tasks, and opening doors, among others [18]. In wheeled-legged robots, since both wheels and legs can be utilized for manipulation tasks, an even wider range of tasks can be performed [19].

978-1-7281-9077-8/21/$31.00 ©2021 IEEE

To achieve complex mobile manipulation tasks demanding high accuracy and robustness, one typical solution is to guide the robotic operation with visual localization or visual servoing at the last centimeter[20], [21], [22]. Image retrieval techniques were adopted in [23], [24] to identify the most relevant to the query to estimate the pose from 2D-3D machines. Learning-based methods are another popular solution to realize visual localization which either predicts matches for pose estimation [25] or directly regresses the pose [26].

B. Contribution

Compared with existing works on state estimation of pure legged or wheeled robots, the main contributions of this paper are two-fold:

1) Novel hybrid locomotion odometry is developed for wheeled-legged robots based on robot kinematics, contact force estimation and geometric features of terrain.
2) Unified state estimation framework is proposed to fuse different sensor readings for long-range yet accurate operations so as to complete general mobile manipulation tasks.

This paper is organized as follows. In Section II, the robot system we target on and the whole state estimation framework are introduced. Thereafter, different modules under the framework are elaborated in Section III, especially the novel odometry for hybrid wheeled-legged locomotion. Section IV presents two experiments running on the real robot and the evaluation results of our method. The paper ends with conclusions and an outlook on future research.

II. System & Framework

Before going into technical details of the proposed approach, we provide a broad view of our methodology by introducing the robot platform we develop on and the general state estimation framework.

A. Robot Platform

The robot platform we use to evaluate our algorithm is a dual-arm wheeled-legged quadrupedal robot named *Pholus*, developed by Istituto Italiano di Tecnologia (IIT). Pholus is an improved version of the *Centauro* robot [3] and has the same 41 DoFs. The robot was designed with toughness and strength in mind to perform demanding mobile manipulation tasks. It contains multiple sensors for perception, including an RGB-D camera, a Time-of-Flight (ToF) camera and a 3D Lidar, and is also equipped with an inertial measurement unit (IMU) as well as torque sensors at every joint (except the two for rotating the ToF camera and Lidar).

B. State Estimation Framework

To fulfill complex mobile manipulation tasks, state estimation is one of the keystones which allows the robot to locate itself in the surrounding environment. As shown in Fig. 1, the hybrid locomotion odometry lies at the heart of the framework. It mainly relies on proprioceptive sensors which promise high-frequency, smooth and robust estimation [27].

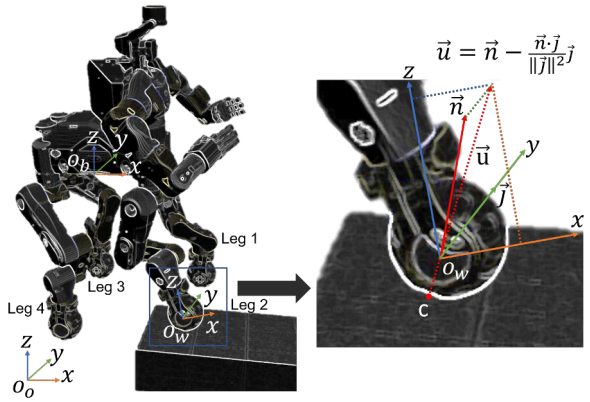


Fig. 2: Sketch of the wheeled-legged robot Pholus traversing uneven terrain. Three coordinates are depicted: fixed odometry frame O_o, body link frame O_b attached at the bottom center of pelvis and the wheel frame O_w with Y axis pointing along the rotation vector $\vec{j}$. $\vec{n}$ stands for the ground normal, and $\vec{u}$ is its projection in the wheel rotation plane. The instantaneous contact point c is defined as the lower intersection point between the wheel contour and $\vec{u}$.

Similar to existing works on the odometry of legged robots, we need to estimate the contact status of the wheels. Further considerations are given to wheel rolling and the terrain surface. In order to mitigate the issue of drift caused by dead-reckoning, we include exteroceptive sensors (e.g. Lidar) to improve localization within a prior map. However, even with the help of Lidar localization, robots may fail in some mobile manipulation scenarios which require a high degree of accuracy. Complementing our state estimation approach, we integrate visual localization into the framework to address this difficulty. More details will be given in the following section.

III. Methodology

In this section, we elaborate on the three main modules of the state estimation framework, namely the Hybrid Locomotion Odometry (HLO), Lidar-based localization and visual localization.

A. Hybrid Locomotion Odometry

Based on the Leg Odometry (LO) described in [11], the foot at the end of one supporting leg is assumed to be fixed to the ground. However, for hybrid wheeled-legged locomotion concerned in this paper, there is no such fixed point in the legs. Instead, by using the Instantaneous Contact Point (ICP) of supporting wheels, we can obtain the following kinematic equation, which is similar to that in LO:

$$ {}^{b}\dot{\mathbf{X}}_b + {}^{b}\dot{\mathbf{X}}_c + {}^{b}\boldsymbol{\omega}_b \times {}^{b}\mathbf{X}_c = \mathbf{0} \tag{1} $$

where b, c in the subscripts and superscripts stand for the robot base link and ICP respectively; ${}^{b}\mathbf{X}_c$ is the position of the ICP with respect to the base link frame and ${}^{b}\dot{\mathbf{X}}_c$ the corresponding velocity; ${}^{b}\dot{\mathbf{X}}_b$, ${}^{b}\boldsymbol{\omega}_b$ are, respectively, the linear and angular velocities of the base link relative to a fixed odometry frame represented in the coordinates of base link.

Since there may be multiple wheels supported by the ground, we use contact forces to weight the estimations from different legs. This stems from a reasonable assumption that a larger contact force generally implies greater contact stability and reliability, and therefore, better estimation, while swing legs with no contact forces should be neglected. Similar handling has also been adopted in [11], [28]. The resulting equation to combine all four legs are:

$$\begin{bmatrix} {}^b\dot{\mathbf{X}}_b \\ {}^b\boldsymbol{\omega}_b \end{bmatrix} = -\left(\mathbf{w} \begin{bmatrix} \mathbf{I} & \times {}^b\mathbf{X}_{c1} \\ \mathbf{I} & \times {}^b\mathbf{X}_{c2} \\ \mathbf{I} & \times {}^b\mathbf{X}_{c3} \\ \mathbf{I} & \times {}^b\mathbf{X}_{c4} \end{bmatrix} \right)^{\dagger} \mathbf{w} \begin{bmatrix} {}^b\dot{\mathbf{X}}_{c1} \\ {}^b\dot{\mathbf{X}}_{c2} \\ {}^b\dot{\mathbf{X}}_{c3} \\ {}^b\dot{\mathbf{X}}_{c4} \end{bmatrix} \tag{2}$$

$$\mathbf{w} = \begin{bmatrix} w_{c1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_{c4} \end{bmatrix}$$

where the subscripts $c1$ $c4$ specifically indicates the ICP of four legs as shown in Fig. 2, $\dagger$ the pseudoinverse operator, and $\mathbf{w}$ a diagonal weight matrix calculated based on contact force $\boldsymbol{\lambda}$ along the z direction. The weight assigned for the ith leg can be computed as:

$$w_{ci} = \frac{\boldsymbol{\lambda}_i^z}{\sum_{n=1}^{4} \boldsymbol{\lambda}_n^z} \tag{3}$$

Because our robot is not mounted with any contact sensors to directly measure contact force, here we estimate it through the robot dynamic model and joint torque readings:

$$\boldsymbol{\lambda}_i = (\mathbf{J}_{wi}^T(\mathbf{q}))^{\dagger}(\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{h}(\mathbf{q}, \dot{\mathbf{q}}) - \mathbf{S}\boldsymbol{\tau}) \tag{4}$$

where $\mathbf{q}$ is a vector of joint angles (including the 6 floating base joints), $\mathbf{M}(\mathbf{q})$ the inertia matrix, $\mathbf{h}(\mathbf{q}, \mathbf{q})$ the centripetal, Coriolis and gravity forces, $\mathbf{S}$ the actuated joint selection matrix and $\boldsymbol{\tau}$ the actuated joint torques, and $\mathbf{J}_{wi}^T(\mathbf{q})$ the Jacobian at the center of the ith wheel.

Compared with LO, the most tricky part in formulating HLO is to locate the ICP ${}^b\mathbf{X}_c$ on the supporting wheels. As depicted in Fig. 2, ICP is considered as the lower intersection point between the wheel contour and the normal vector of the local ground surface:

$$\begin{aligned} \vec{u} &= \vec{n} - \frac{\vec{n} \cdot \vec{j}}{\|\vec{j}\|^2} \vec{j} \\ {}^w\mathbf{X}_c &= \frac{r}{\|\vec{u}\|} \vec{u} \end{aligned} \tag{5}$$

where $\vec{j}$ is the wheel rotation axis, $\vec{n}$ the normal vector of the local ground surface supporting under one wheel, $\vec{u}$ its projection in the wheel rotation plane, r the wheel radius, and ${}^w\mathbf{X}_c$ the ICP represented in the wheel coordinates.

Thereafter, the position of ICP in the base link frame ${}^b\mathbf{X}_c$ can be easily derived from ${}^w\mathbf{X}_c$ through forward kinematics, and its velocity can be computed through the Jacobian ${}^b\mathbf{J}_c$, i.e. ${}^b\dot{\mathbf{X}}_c = {}^b\mathbf{J}_c \cdot \dot{\mathbf{q}}$.

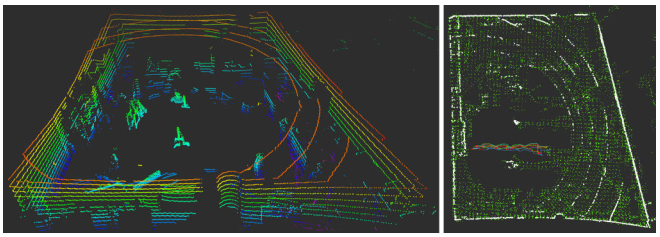


Fig. 3: Visualisation of Lidar-based Localization (Left: point cloud directly from 3D Lidar, color scaled with point height; Right: map matching of particle filter in the top-down view, greed particles are from prior map and white are the matched ones, trajectory tracked by MCL is depicted as red line while green for the ground truth and red from the odoemtry).

With the estimated linear and angular velocities of base link in Eq.(2), we can transform them into a fixed odometry frame and then integrate in time to obtain the pose. In addition, to make use of high-quality orientation estimation in the IMU, we implemented the extended Kalman filter to filter its measurement into the HLO.

B. Lidar-based Localization

In order to handle complex ground situations like slippery or soft areas, a 3D Lidar Map-Based Monte Carlo Localization (MCL) method was also implemented to provide robust drift-free state estimation in real environments.

MCL predicts the pose of a robot by evaluating the observation, Lidar scan in this case, at a set of hypothetical poses. Each hypothetical pose p_i with its associated weight ω_i is called a particle $\mathbf{P}_i$. MCL consists of mainly two steps: *prediction* of new hypothetical poses and *update* of particles at the predicted poses.

1) Prediction: Inputs to prediction step are the particles and HLO introduced above. The change of odometry u_t between last and current Lidar data is applied to previous pose $p_{i,t-1}$ with a random Gaussian noise ϵ_i to get the new pose $p_{i,t}$:

$$p_{i,t} = p_{i,t-1} \times u_t \times \epsilon_i \tag{6}$$

2) Update: In the update step, an observation model $P(z_i|p_i, m)$, where z_i is the observation and m stands for the map used to calculate the weight ω_i of each particle. Particles with weights lower than the median will be deleted on a percentile basis. The output pose p_{MCL} of MCL is the weighted average of the remaining particle poses. To maintain a fixed number of particles N, new particles are added by duplicating high weight particles. The update process can be written as:

$$p_{MCL} = \frac{\sum_{n=1}^{N} P(z_i|p_i, m) p_i}{\sum_{n=1}^{N} P(z_i|p_i, m)} \tag{7}$$

By repeating the two steps recursively, a continuous state estimation can be carried out based on the MCL method as shown in Fig. 3.

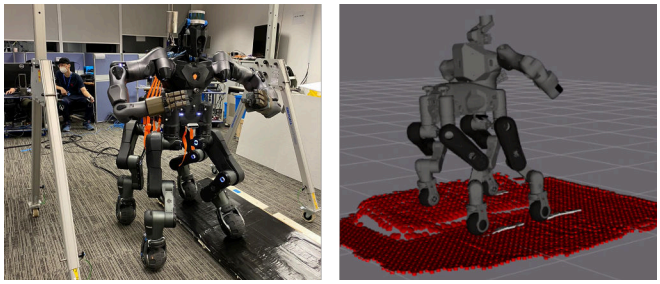


Fig. 4: Pholus stepping over a slanted platform (Left: experiment setup with a platform of 156cm length, maximum height of 15cm and an inclination angle of 10 degrees; Right: RVIZ visualization with red arrows representing the normal vectors of the detected terrain, and square marks in front of wheels representing predicted future foot prints.).

C. Visual Localization

For certain mobile manipulation tasks demanding highly accurate operation, even the aforementioned Lidar-based MCL may not meet the requirements. In this case, visual localization can be a good complement to our existing framework and relocate the robot precisely with respect to the environment. Here, a ToF camera and a color camera have been calibrated for image registration. The captured color image was used for object detection whereas the ToF camera was used to capture the depth image for constructing 3D information such as point cloud. A Single Shot Detector (SSD) [29] with MobileNet [30] was used for objection recognition. This deep learning network has been revised and fine-tuned using our data sets. The recognition module assigns each recognized object an attribute which comprises Region of Interest (ROI), object type and 3D key points related to specific tasks.

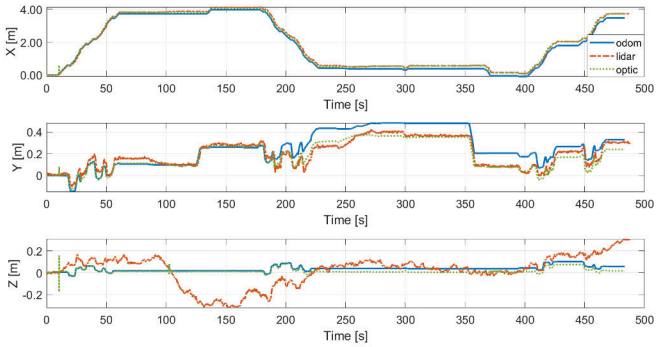
IV. Experiment

Two experiments were conducted on our robot Pholus to evaluate the performance of the proposed state estimation framework. Multiple OptiTrack Prime series cameras were mounted at the ceiling of the lab in a rectangular configuration to track the 10 markers attached on the pelvis of the robot which is considered as the ground truth in these two experiments.

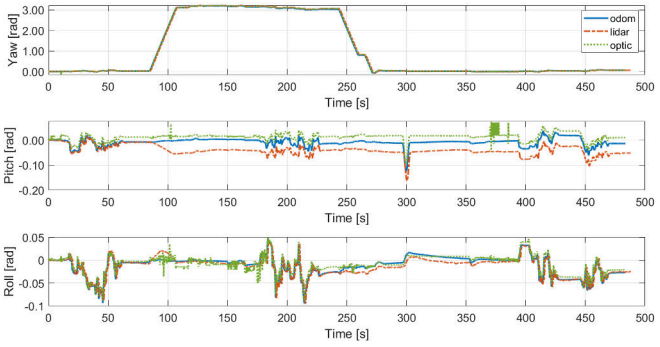
A. Stepping over Slanted Platform

In this experiment, Pholus stepped over a slanted platform repeatedly to test the accuracy of the hybrid state estimation as shown in Fig. 4. The detection of the inclined slope was obtained using the Intel Realsense D435 mounted at the pelvis of the Pholus. The collected point cloud was utilized to construct an elevation map [31] with surface normal vector which decides desired wheel poses and also contributes to HLO as shown in Fig. 4.

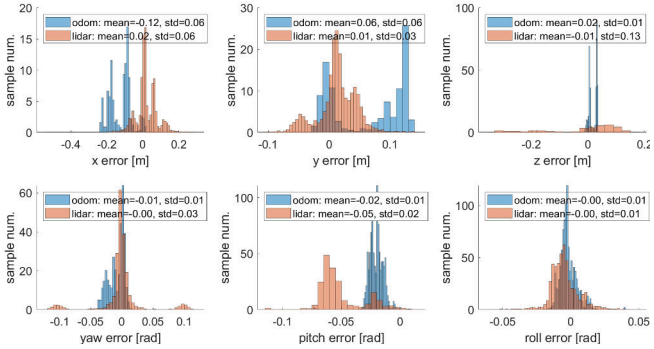
During this experiment, Pholus stepped over the platform for a total of 3 times taking around 10 minutes. The tracking results from HLO, MCL and OptiTrack are compared in Fig. 5. Overall, both HLO and MCL were able to maintain good tracking (around several centimeters) throughout the



(a) Position tracking along x, y, z directions over time



(b) Yaw, pitch and roll tracking over time



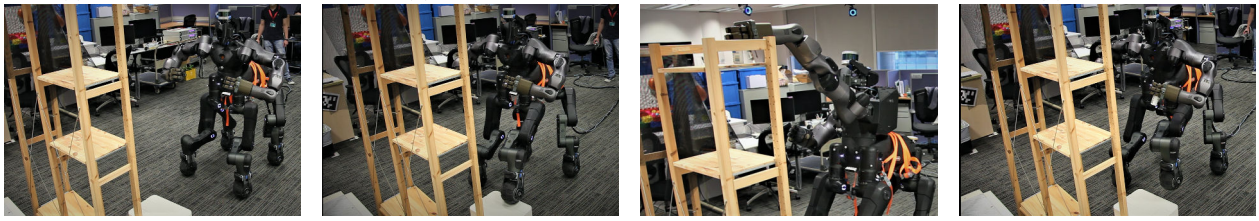
(c) Error distribution with respect to the ground truth OptiTrack

Fig. 5: Comparison of tracking results from HLO, MCL and OptiTrack for stepping over slanted platform. In the legends, odom, lidar, optic respectively stands for the HLO, MCL and OptiTrack tracking results.

whole experiment of stepping up and down and rolling. Compared with HLO, MCL showed less drift along X and Y direction while more noisy along Z direction. HLO showed some significant drifting along Y direction starting from 200s which should blame to the slipping during stepping down. The noisy Z tracking of MCL is quite common in Lidar localization methods because usually there are insufficient Lidar points observed on floor and ceiling to correct the pose in Z direction. And the orientation tracking are both very well thanks to the extended Kalman filtering of IMU.

Additionally, the contact force estimation of wheels is shown in Fig. 7 from where the contact switch among legs can be easily observed. In order to see more details clearly,

(a) Stool Detection (b) Pick Task (c) Reverse with Stool (d) Place Task



(e) Dock to Shelf (f) Step On (g) Grasp Bottle (h) Step Off

Fig. 6: Sub-task sequence for the mobile manipulation experiment.

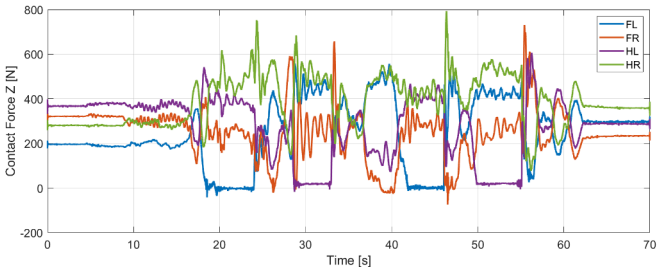


Fig. 7: The estimation results of contact force along Z direction for four legs (FL: front left leg; FR: front right leg; HL: hind left leg; HR: hind right leg.).

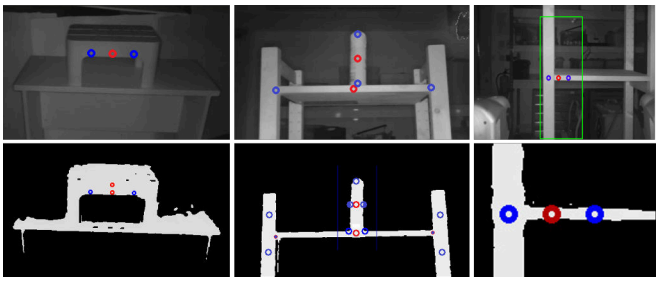


Fig. 8: Key-points detection for visual localization. Red circles are the target center to be published for docking, grasping or picking up while blue circles are reference points for computing the orientation of objects (Upper row: the infra-red images captured from the ToF camera; Lower row: segmented images which were generated by using depth-based segmentation.).

here we only shows data for one round of Pholus traversing through the platform. From the figure, we can tell that the legs on the left side were lifted up (almost zero contact force) and down in sequence to cross the platform (Step on: 15 35s; Step off: 40 60s). The total weight of Pholus is around 120 kg, so the sum of all four contact forces roughly matches. To be noted, the joint torque sensors haven't been properly calibrated for a long time so the contact force estimator was not working at its best condition.

B. Mobile Manipulation

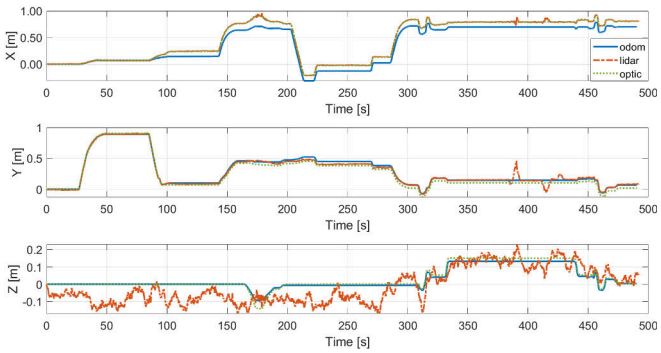
In this experiment, Pholus was tasked to grasp a bottle on top of a shelf. However, as the shelf is too high and the bottle out of reach when Pholus is standing on the ground, it needs to place a stool in front of the shelf, and step on it in order to reach higher to grasp the bottle. This can be considered as a representative and challenging mobile manipulation problem. As shown in Fig. 6, the whole procedure was divided into 8 sub-tasks which were triggered by a state machine in sequence. The main tasks involved were the picking and placing of a stool using dual-arms from a table to the shelf as shown in Fig. 6b, and the grasping of a bottle at the top of the shelf using left-arm and torso as shown in Fig. 6g. The stool used in this experiment has a dimension of 40x28x20cm, while the shelf is of 48x30x180cm.

The plot on Fig. 9 shows the tracking performance of the state estimator over the period of the mobile manipulation experiment. The histogram on Fig. 9c also shows the error distributions with a sampling rate of 10 samples/sec.

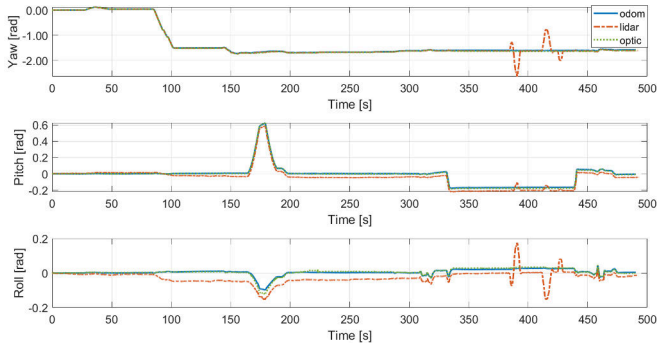
For the picking and placing task, the robot searched for the pose of the stool, and used visual servoing to approach the table while orientating itself into a suitable position for picking with both of its arms. Upon successfully picking up the stool, the robot then reversed and proceeded to dock toward the shelf using a combination of HLO and visual servoing. This is displayed during 0-100s on the plot, with a displacement in Y and a rotation in the yaw axis.

Once the robot docked, it placed the stool down in a suitable position with respect to the shelf position guided by visual servoing. The putting down task is observed at 150-200s when there is a shift in pitch angle which was successfully tracked by HLO with less than 0.02rad error.

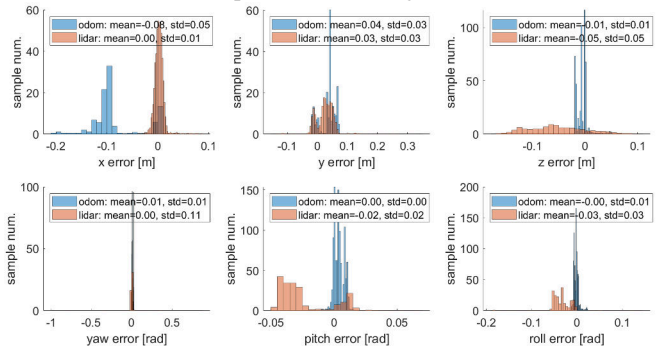
For the grasping task, Pholus had to step up on the stool and tilt its torso upwards, in order for the hand to be able to reach a height of 1.75 meters to reach the bottle. The stepping task is observed at 300-450s where there is a large increment in height on the Z plot. One of the feet was also



(a) Position tracking along x, y, z directions over time



(b) Yaw, pitch, roll tracking over time



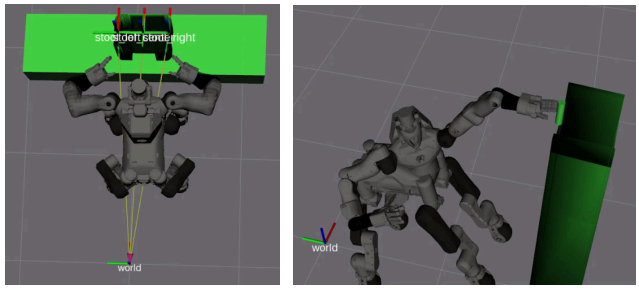
(c) Error distribution with respect to the ground truth OptiTrack

Fig. 9: Comparison of tracking results from HLO, MCL and OptiTrack for mobile manipulation experiment. In the legends, odom, lidar, optic respectively stands for the HLO, MCL and OptiTrack tracking results.

not in contact with the ground. The tracking error started to increase further during this period due to inaccuracies from the contact forces estimates and leg kinematics, which was also observed in the previous experiment.

Overall, the state estimator provided reasonably good tracking in comparison to the actual ground truth, with an average positional error of 8cm and orientation error of less than 0.01 rads, even as a series of complex maneuvers was performed. However, during 350-450s where the grasping task took place, it is also observed that there were some discrepancies from the Lidar tracking, due to poor response time in the Lidar-based localisation as the robot torso was abruptly rotated. This is more noticeable in the orientation plot on Fig. 9b.

Fig. 8 shows examples of key-points detection for visual



(a) Dual-Arm Pick Task (b) Grasp Task

Fig. 10: RVIZ visualisation of the mobile manipulation experiments. The objects detected by visual localization module was registered in the odometry frame.

localization. The upper row of images shows the infra-red images captured from the ToF camera and the lower row of images are segmented images which were generated by using our depth-based segmentation. Once the segmented image is generated, our key-points generator will base on the geometry of the object to extract the desired key-points for manipulation. For the extracted key-points, the center (red circle) is always the targeted position of the object to be published for docking, grasping or picking up, while the adjacent points around the center are reference points (blue circle) to compute the orientation of the object relative to the robot. When locating the stool/bottle, it will take at least 10 frames to compute the average key-points from the detectors to filter out detection noises. For the action of grasping a bottle on the wooden shelf, besides detecting the bottle, the key-points of the shelf were also detected for generation of the 3D model of the shelf for computing the collision avoidance. In this experiment, the 3D models of objects were pre-defined and imported into MoveIt[32] for motion planning as shown in Fig. 10.

V. CONCLUSION

This paper presented a state estimation framework for hybrid locomotion of wheeled-legged robots by fusing various sensor input. A novel odometery algorithm was introduced as the core of this framework which uniformed the calculation of both legged and wheeled motions, and 3D Lidar MCL and visual localization were implemented further to achieve drift-free and accurate state estimation. The proposed approach was evaluated in two experiments, stepping over a slanted platform and performing a challenging mobile manipulation task, and showed very promising results. The robot was able to grasp a bottle with two centimeters margin of finger openness when starting movements meters away.

To further improve the accuracy and robustness of this work, we may consider mounting contact force sensors directly on the wheels to eliminate the estimation error introduced by using joint torques. In addition, more effort will also be put to address the issue of noisy Z tracking from MCL.

REFERENCES

[1] G. Bellegarda, K. van Teeffelen, and K. Byl, "Design and evaluation of skating motions for a dexterous quadruped," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 1703–1709.

[2] M. Bjelonic, C. D. Bellicoso, Y. de Viragh, D. Sako, F. D. Tresoldi, F. Jenelten, and M. Hutter, "Keep rollin'—whole-body motion control and planning for wheeled quadrupedal robots," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2116–2123, 2019.

[3] N. Kashiri, L. Baccelliere, L. Muratore, A. Laurenzi, Z. Ren, E. M. Hoffman, M. Kamedula, G. F. Rigano, J. Malzahn, S. Cordasco *et al.*, "Centauro: A hybrid locomotion and high power resilient manipulation platform," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1595–1602, 2019.

[4] A. Bry, A. Bachrach, and N. Roy, "State estimation for aggressive flight in gps-denied environments using onboard sensing," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 1–8.

[5] S. Z. A. Mukhtar, V. B. Saputra, S. Verma, K. Zhang, and A. H. Adiwahono, "Sparse-3d lidar outdoor map-based autonomous vehicle localization," 2019.

[6] P. Li, T. Qin, B. Hu, F. Zhu, and S. Shen, "Monocular visual-inertial state estimation for mobile augmented reality," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2017, pp. 11–21.

[7] 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.

[8] M. Bloesch, "State estimation for legged robots-kinematics, inertial sensing, and computer vision," Ph.D. dissertation, ETH Zurich, 2017.

[9] N. Gupta, "State estimation for legged robots using proprioceptive sensors," Ph.D. dissertation, Carnegie Mellon University, 2019.

[10] M. F. Fallon, M. Antone, N. Roy, and S. Teller, "Drift-free humanoid state estimation fusing kinematic, inertial and lidar sensing," in *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2014, pp. 112–119.

[11] 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," *Robotics: Science and Systems (RSS)*, 2017.

[12] M. Bloesch, C. Gehring, P. Fankhauser, M. Hutter, M. A. Hoepflinger, and R. Siegwart, "State estimation for legged robots on unstable and slippery terrain," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 6058–6064.

[13] N. Ganganath and H. Leung, "Mobile robot localization using odometry and kinect sensor," in *2012 IEEE International Conference on Emerging Signal Processing Applications*, 2012, pp. 91–94.

[14] J. Levinson, M. Montemerlo, and S. Thrun, "Map-based precision vehicle localization in urban environments." in *Robotics: science and systems*, vol. 4, no. Citeseer. Citeseer, 2007, p. 1.

[15] L. de Paula Veronese, J. Guivant, F. A. A. Cheein, T. Oliveira-Santos, F. Mutz, E. de Aguiar, C. Badue, and A. F. De Souza, "A light-weight yet accurate localization system for autonomous cars in large-scale and complex environments," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 520–525.

[16] J. Saarinen, H. Andreasson, T. Stoyanov, and A. J. Lilienthal, "Normal distributions transform monte-carlo localization (ndt-mcl)," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 382–389.

[17] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time." in *Robotics: Science and Systems*, vol. 2, no. 9, 2014.

[18] . Brock, O. Khatib, and S. Viji, "Task-consistent obstacle avoidance and motion behavior for mobile manipulation," in *2002 International Conference on Robotics and Automation (ICRA)*. IEEE, 2002, p. 388–393.

[19] M. Schwarz, T. Rodehutskors, C. Droeschel, M. Beul, S. M, N. Araslanov, I. Ivanov, C. Lenz., J. Razlaw, S. Schuller, D. Schwarz, A. Topalidou-Kyniazopoulou, and S. Behnke, "Nimbro rescue : Solving disaster-response tasks through mobile manipulation robot momaro," *Journal of Field Robotics (JFR)*, vol. 34, no. 3, pp. 400–425, 2017.

[20] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *ieee Transactions on Robotics and Automation*, vol. 8, no. 3, pp. 313–326, 1992.

[21] P. I. Corke *et al.*, *Visual Control of Robots: high-performance visual servoing*. Research Studies Press Taunton, UK, 1996.

[22] J. L. Schönberger, M. Pollefeys, A. Geiger, and T. Sattler, "Semantic visual localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6896–6906.

[23] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[24] T. Sattler, A. Torii, J. Sivic, M. Pollefeys, H. Taira, M. Okutomi, and T. Pajdla, "Are large-scale 3d models really necessary for accurate visual localization?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1637–1646.

[25] J. Valentin, M. Nießner, J. Shotton, A. Fitzgibbon, S. Izadi, and P. H. Torr, "Exploiting uncertainty in regression forests for accurate camera relocalization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4400–4408.

[26] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2938–2946.

[27] G. Fink and C. Semini, "Proprioceptive sensor fusion for quadruped robot state estimation, in," *Las Vegas, Nevada*, 2020.

[28] 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*, vol. 2, no. 2, pp. 1023–1030, 2017.

[29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg, "SSD: single shot multibox detector," *CoRR*, vol. abs/1512.02325, 2015. [Online]. Available: http://arxiv.org/abs/1512.02325

[30] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017. [Online]. Available: http://arxiv.org/abs/1704.04861

[31] P. Fankhauser, M. Bloesch, C. Gehring, M. Hutter, and R. Siegwart, "Robot-centric elevation mapping with uncertainty estimates," in *International Conference on Climbing and Walking Robots (CLAWAR)*, 2014.

[32] S. Chitta, I. A. Sucan, and S. Cousins, "Moveit! [ROS topics]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012.