A standardized benchmark for humanoid whole-body manipulation

William Thibault[1] Francisco Javier Andrade Chavez[2] Katja Mombaur[3]

Abstract— **In this paper we focus on the evaluation of humanoid manipulation skills while balancing on two feet. This involves manipulation while standing and loco-manipulation where the object is being manipulated while taking steps. With this objective in mind, an initial study of whole-body manipulation in a box manipulation scenario with two different motions using the University of Waterloo's REEM-C, "Seven", is investigated to provide insight into a valuable setup, comprehensive test protocols and useful performance metrics based on real world data. The contribution of this paper is a proposed benchmark for whole-body manipulation consisting of the design of a test bed inspired by real use cases for humanoid whole-body manipulation tasks, the definition of a set of protocols to standardize the testing procedure and insightful key performance indicators (KPIs) based on this initial study with the real robot. The proposed benchmark for humanoid whole-body manipulation is part of the EUROBENCH project that aims at creating a benchmarking framework for robotic systems performing locomotion related tasks.**

I. INTRODUCTION

In order to enable human-centered robots such as humanoid and wearable robots to move out of the lab into the real world, it is important to assess their suitability for given real world tasks and application domains in advance. Benchmarking is considered as an important instrument for evaluating robot performance and predicting how robots will satisfy the specific needs of users. Recent benchmarking efforts included a number of international robotics competitions that received a lot of attention, such as the Cybathlon, the DARPA Robotic Challenge, RoboCup, RoCKIn and the European Robotics League. Often these competitions focus on a more qualitative performance execution of specific tasks as well as the required time in a game or race setting, but there still is a clear lack of unified benchmarking scenarios and quantified key performance indicators.

The research presented in this paper has been performed in collaboration with the European project EUROBENCH, which aims to create the first unified benchmarking framework for robotic systems in Europe [1] and to set up two benchmarking facilities accessible to the entire academic and industrial robotics community. The goal is to allow companies and researchers to test the performance of their robots at any stage of development and to get reliable information in advance about robots they may consider purchasing for given tasks. The proposed benchmarks all consist of standardized test beds, benchmarking protocols and key performance indicators which will be implemented in a benchmarking software. The focus is on locomotion related tasks, and the benchmarks measure elementary locomotion skills with respect to stability, robustness, and motions on different terrains with inspiration from real world use cases.

An inciting work for the EUROBENCH project was Torricelli et al. [2], where the general design of a benchmarking scheme for bipedal locomotion is described. The scheme involves classification of different bipedal balancing and locomotion tasks such as stair walking, balancing under disturbances like pushes or under constant weight and walking on compliant terrain. To be able to benchmark the motions, it proposes a benchmark scheme for motor abilities that is classified into performance and human likeness where metrics like success rate, energetic and mechanical cost of transport and joint torques are measured. This is followed by a scheme for defining benchmarking protocols that would allow collaborators to improve upon existing protocols or develop new ones. This work, supported by projects like H2R, BALANCE, KoroiBot, WALKMAN and BioMot, motivated the development of the unified benchmarking framework that EUROBENCH aims to create. This work along with the KoroiBot project inspired locomotion benchmarking of the HRP-2 robot [3]. A range of locomotion tasks [4] were considered including flat ground walking, walking on a beam, step stone walking and stair walking fulfilling a number of the bipedal locomotion skills from Torricelli et al. [2]. To measure the performance of the HRP-2 performance indicators included success rate, mechanical joint energy, actuators energy, cost of transport and duration of the experiment, again inspired from Torricelli et al. [2]. While bipedal locomotion tasks are well considered, there is a lack of consideration for balancing or locomotion tasks involving manipulation. This work will refer to these tasks as whole-body manipulation and loco-manipulation, respectively.

Benchmarks for bimanual manipulation have been proposed by Chatzilygeroudis et al. through the performance of tasks such as watchmaking and belt assembly of engines [5] while Sotiropoulos et al. proposes a method to evaluate soft hand grasping through the grasping of fruits and vegetables [6]. Other grasping and manipulation benchmarks include a benchmark for grasp planning [7], the box and blocks test for dexterity [8] and in-hand manipulation [9]. These are not very useful when trying to quantify performance for manipulation (typically bimanual) of larger objects which requires the robot to reach the limits of its kinematic workspace and

[1]William Thibault is with Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada wcthibau@uwaterloo.ca

[2]Francisco Javier Andrade Chavez is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

[3]Katja Mombaur is with the Department of Systems Design Engineering and the Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

to go to the edge of dynamic stability margins.

The balance of a humanoid robot while performing a manipulation task has been analyzed in simulation using a generalized zero-moment point approach to determine the robot's ability to balance during manipulation tasks like pushing or pulling objects [10]. Another metric proposed was the dynamic reconfiguration manipulability shape index, based on other manipulability metrics like dynamic and reconfiguration manipulability, that is used to evaluate a bipedal robot's walking posture for a greater dynamical shape changeability through a humanoid's redundancy while performing another task such as directing the head [11]. Whole-body balancing has also been considered by Sugihara and Nakamura using the COG Jacobian to allow the humanoid robot to balance while compensating for disturbances through cooperation of the whole-body [12]. Other approaches to complex whole-body problems, include human to humanoid motion re-targeting such as that done by Di Fava et al. for multi-contact motions between the robot and the environment using a multi-contact QP control formulated framework [13] or a guided manipulation planning approach such as that used by Dellin et al. for the DARPA Robotics Challenge [14]. A planning method that exploits constrained manifolds to increase planning speed has been successfully implemented in humanoid robots for loco-manipulation tasks [15]. The method was validated using two scenarios: opening a door and pushing a cart. The only comparison metric used was the time to complete the task. Similarly, others provide methods for solving complex loco-manipulation tasks with results in simulation only, but again the only metric provided is time [16], [17]. Benchmarking of loco-manipulation tasks with the Valkyrie were performed for box-picking and object grasping on flat ground, uneven terrain and in restricted space [18]. This benchmarking also used few metrics, only the success rate and average time, but also more from a planning perspective without strict protocols.

Most of the approaches mentioned above are able to evaluate the state of balance of the robot while performing a task. However this gives no objectively quantifiable data to evaluate if the task was successful and what performance level a given robot or control strategy was able to achieve in the given task. Others provide few metrics like time or success rate to compare and show improvement with respect to a previous method. This does not provide insight into how efficient the method is or how robust and repeatable it is. This information will be very important when choosing which robotic platform or algorithmic solution will be used in the real world.

In this paper we focus on the evaluation of humanoid manipulation skills while balancing on two feet. This involves manipulation while standing and loco-manipulation where the object is being manipulated while taking steps. To provide more insight into this task we present an initial study of a box manipulation scenario using the humanoid robot REEM-C with an initial set of performance metrics where two different control methods are compared. For the initial study, we have chosen a setup with a single

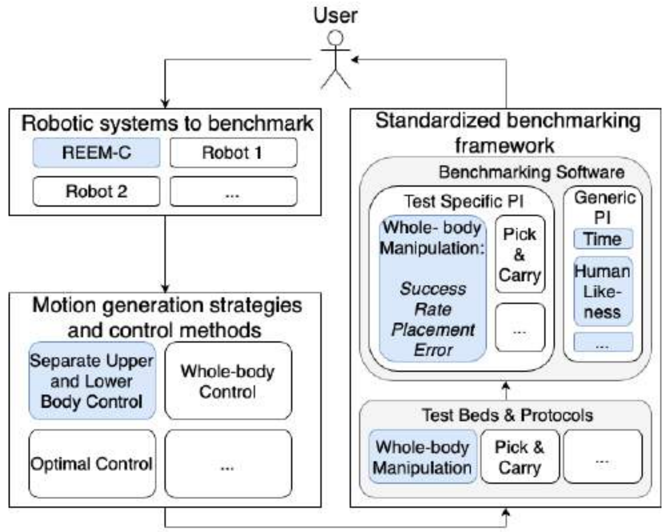


Fig. 1: EUROBENCH benchmarking work flow with blue highlighting the work in this paper.

shelving unit holding a box that must be manipulated in a precise manner while remaining balanced as this is the basis for many tasks that exist in warehouse and logistics scenarios. From this simple scenario, we then propose a design of a complete benchmarking scenario for whole-body manipulation of humanoid robots. The contribution of this paper is a proposed benchmark for whole-body manipulation consisting of the design of a test bed inspired by real use cases for humanoid whole-body manipulation tasks, the definition of a set of protocols to standardize the testing procedure and insightful key performance indicators (KPIs) based on experiments from an initial study of a whole-body box manipulation scenario. This benchmark is closely related to another loco-manipulation benchmark we are developing on object carrying between shelves at a larger distance, in which the locomotion component with a load becomes more important and which will be discussed in a later work. Both benchmarks will be part of the EUROBENCH facility for humanoid robots where they will be applied to general humanoid robots visiting the facility or available there. In this paper we propose a first version of the whole-body manipulation benchmark for humanoid robots, using simple manipulation tasks in the test bed by comparing two motion generation strategies. Figure 1 shows the EUROBENCH benchmarking flow with the blue showing what will be covered in this work.

Section II explains the initial study performed including the experimental setup, the performance metrics measured, the humanoid robot "Seven" and the motion control methods. Section III provides the results obtained in the initial study with a comparison of control methods and an analysis of the success at providing performance information on whole-body manipulation. Section IV describes the proposed benchmarking test bed, the protocols used in it and the KPIs after considering insights from the initial study. Finally, Section V contains the conclusions and future work.

II. An Initial Box Manipulation Scenario

A. Experimental Setup

As an initial study of whole-body manipulation tasks, a box manipulation scenario acts as a reasonable test. Typical humanoid application areas include manufacturing tasks such as in airplanes and on ships, household tasks, healthcare or support care and work in space to name a few. Whenever a humanoid interacts with the environment many manipulation scenarios can be broken down into a pick and place scenario. The combination of many pick and place actions can be used for more complex work such as in one of the previously mentioned applications. Therefore, a shelving unit with a box to be picked and placed from one shelf to another was selected for the initial study as it serves as an excellent example of a core whole-body manipulation task.

The shelving unit is an IKEA IVAR system that has a depth of 0.5 m. This shelving system is readily available worldwide and it is flexible as it allows for heights of shelves to be easily varied. In the configuration used, the shelving unit has three shelves where one is at a height below the waist (0.14 m), another is at a height between the waist and shoulders (0.780 m) and last at a height above the shoulders (1.42 m). The humanoid robot REEM-C was used for reference when gauging the heights as it is a human sized humanoid robot and is the standard test robot that is used in the EUROBENCH framework.

For the object to be manipulated, a regular cardboard box was selected. The box dimensions are 0.5 m wide by 0.16 m tall by 0.15 m deep and weighs 0.1 kg. This was selected as an appropriate box for the initial study as the size requires a bimanual manipulation treatment for the pick and place motion and is not exceedingly heavy for the robot.

The shelving unit with the box on the middle shelf can be seen in front of the robot "Seven" in Figure 2.

In this initial box manipulation study, the desired box manipulation is that the box is picked from the middle shelf and placed on the top shelf. The starting position of the box will be at the edge of the middle shelf, centered horizontally, and the ending position of the box will also be in a similar configuration except on the top shelf. The box is placed at the edge for this initial study to allow for easier access to the object to be manipulated. As some error is to be expected in the pick and place motion, the ending position has two marked boundaries for different levels of error in the placement of the box. The two error boundaries are for 2.5% error and 5% error in the box placement. This error margin is based on a percentage value of the shelf's dimensions. If the box is placed inside the marked boundary then the placement is successful for that level of error. See Figure 2 for the starting position of the box and the two error boundaries for the ending location where within the inner edge of the green tape indicates a placement accuracy of 2.5% and within the inner edge of the blue tape indicates a placement accuracy of 5%.

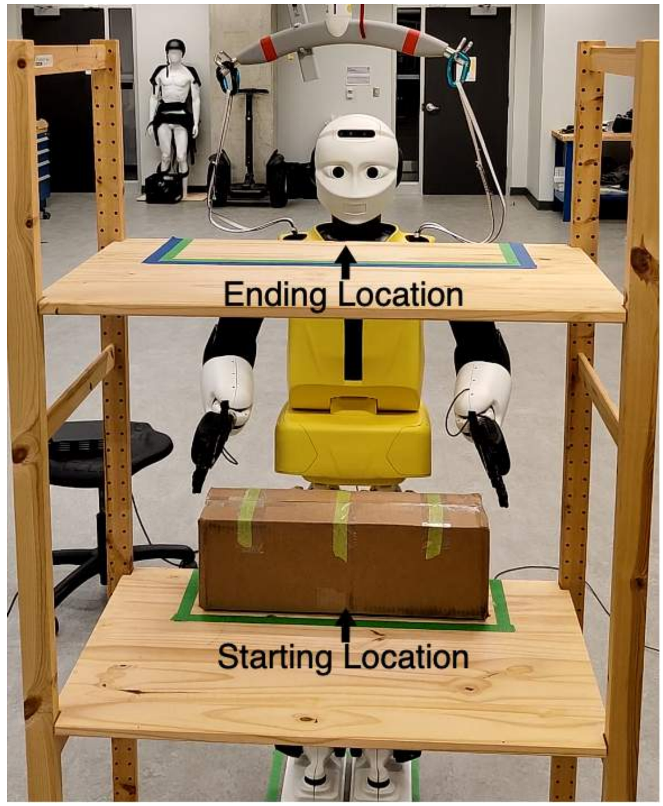


Fig. 2: Starting and Ending Locations for Box Manipulation

B. Performance Metrics

To measure the performance of the robot in the box manipulation scenario several performance metrics were selected:

- Time to perform box manipulation (s)
- Success rate of 2.5% Placement Accuracy
- Success rate of 5% Placement Accuracy
- Cost of Operation ($Amps \cdot s/kg$)
- Mechanical Work (J)

The cost of operation represents the current consumed by the joints during the box manipulation motion for a box of a given mass. These values are readily available from the motor currents provided by the joint states published by ROS. To calculate the cost of operation the following equation is used:

$$Cost = \sum_{j=1}^{n} \int_{0}^{T} \frac{1}{m_O} |i_j| dt \tag{1}$$

where n is the number of motorized joints consuming current, i_j is the current of the motorized joint j, T is the time to perform the entire motion (from start pose to rest pose) and m_O is the mass of the object being manipulated (0.1 kg).

The mechanical work represents the work done by the robot on the box to move it from the starting position to the ending position. This metric requires joint torques, which cannot always be provided by every robot, such as with the REEM-C. As the joint torques cannot be determined experimentally, they can be determined by performing a dynamic simulation of the robot such as in Gazebo. The mechanical work is calculated as follows:

$$W = \sum_{j=1}^{n} \int_{0}^{T} |\tau_j \cdot \dot{\phi}_j| dt \qquad (2)$$

where n is the number of joints, τ_j is the torque at joint j, $\dot{\phi}_j$ is the angular velocity at joint j and T is the time to perform the entire motion (from start pose to rest pose).

C. The Humanoid Robot "Seven"

Seven arrived at the University of Waterloo in the summer of 2020. The robot has 68 degrees of freedom (DoF), which includes each underactuated, 19 DoF hand. Seven weighs 80kg and is 1.64m tall. It has an IMU at the pelvis, four FT sensors (two at the wrists and two at the ankles), lasers in the feet, an Intel RealSense on the head as well as a NVIDIA Jetson TX2. The REEM-C series come with a set of skills that work right out of the box, such as walking, grasping, whole-body control and text to speech. It is fully ROS based. Another of the REEM-C series robots will be available at the EUROBENCH robotics facility where it will be used as well for benchmarking purposes.

D. Motion Generation and Control Methods

1) Separate Upper and Lower Body Controllers: The separate upper and lower body control method uses one controller to control the leg joints and another to control the torso and arm joints of the robot. In this control method, a walking controller and stabilizer is used for the lower body and only controls the leg joints. This controller is the default walking controller developed by PAL Robotics that implements Kajita walking pattern generation [19] with IK for the leg joints and a stabilizer. The upper body controller uses joint trajectory controllers that are controlled by MoveIt! sending joint trajectories based on calculated inverse kinematics results. The torso joints are held at zero to help keep the center of mass centered and the arms are used to manipulate the box. This torso configuration is preferred by the default walking controller for stability, otherwise the robot is likely to fall. To create valid motion plans MoveIt! was configured to use the TRAC-IK kinematics solver, as it handles joint limits better than KDL, using a threaded approach to solving the inverse kinematics with two different solvers and returning a valid solution from whichever converges first [20]. The first uses an inverse Jacobian method approach with Newton's method convergence that can avoid local minima that occur from joint limits and the second approach uses a sequential quadratic programming nonlinear optimization with quasi-Newton methods [20]. The default planner in MoveIt!, Open Motion Planning Library (OMPL), was used to generate motion plans for the arms moving from one set of joint positions to another. The joint positions for a given hand pose were pre-calculated using MoveIt! then mirrored for both arms as a bimanual motion planner was not yet developed.

To generate a motion with these controllers, MoveIt!'s kinematic planning can be used to see if the box can be reached given the robot's starting position. If the box cannot be reached then the robot can step forward to reach the box and pick it up. Then MoveIt! can be used to determine if the box can be placed on the top shelf without any collisions. If this cannot be done then the robot can take a step back to raise the box then step forward to place it. These motions are generated offline then the joint trajectories and step sequences are run on the controllers. Based on the simple motion generation approach described the motion resulted as shown in Figure 3.

As seen in Figure 3, the robot can be seen approaching and retreating from the shelf depending on the stage of the motion. This motion generation method and set of controllers leverages the ability of humanoid robots to walk to a space in which they can make use of a wider workspace. This makes it easier to find solutions that can reach the desired height of the objective while avoiding collisions with the environment. It allows decoupling of the upper body from the lower body for simpler planning while still finding a viable solution.

2) Whole-body Controller: The whole-body control method uses a single controller to control all the joints in the robot. In this control method a stack-of-tasks method is used to control all the joints according to a set hierarchy of tasks and plan a valid motion. To implement the whole-body control method the Cartesian control software CartesI/O was used to develop a hierarchical task description and specify constraints using the OpenSoT math-of-tasks approach with a quadratic programming solver [21]. This framework allowed the joint and velocity limits to be applied with the task list set to the following in descending order or priority:

1) Left and right foot Cartesian pose
2) Center of mass X and Y position, waist yaw
3) Left and right hand Cartesian pose, torso posture, left and right arm joint 2 posture
4) Head posture, left and right leg posture

To generate a motion with this controller, the left hand, right hand and center of mass can be sent Cartesian positions to move to. By leveraging the fact that all the joints can be used the robot can lean forward to grab the box then remove it from the shelf and raise it in such as way that no collision with the shelving unit occurs. These Cartesian trajectories for the hands and center of mass are sent to CartesI/O, which solves the stack-of-tasks online and the motion plan is executed on the robot with the joint trajectory controllers. Based on this simple Cartesian stack-of-tasks approach with CartesI/O the motion occurs as shown in Figure 4.

Figure 4 shows how the robot leverages the legs, torso and arms to reach for the object creating a crouching motion. The whole-body control method allows for a valid solution to be found in the constrained workspace, which was done in simulation. By using the stack-of-tasks method we can explore solutions that involve all the joints in the robot, increasing our solution search space. Note that no stabilization algorithm was used, but the center of mass was shifted forward and backward to maintain balance. Also, no collision avoidance was implemented, so the posture tasks were enabled and disabled depending on the phase to avoid collisions while increasing workspace. Although walking can

Fig. 3: Separate Upper and Lower Body Controllers Method Motion Sequence

Fig. 4: Whole-body Controller Method Motion Sequence

be performed with a stack-of-tasks method, it was not in this motion generation approach as it was not necessary to perform a valid motion and would require higher complexity.

III. EXPERIMENTAL RESULTS & DISCUSSION

To provide an initial study for the box manipulation, the two motion approachs were executed 10 times for the box placement from the front middle shelf to the front top shelf for the 0.1 kg box. See Figure 1 for benchmarking overview being performed. Table I shows the results for the initial performance metrics. Note that in the case of the time to perform the box manipulation, cost of operation and mechanical work average metrics are presented.

TABLE I: Manipulation Key Performance Indicators

Performance Metric	Separate Upper and Lower Body Control Method	Whole-body Control Method
Average Time (s)	52.96	62.41
2.5% Placement Accuracy Success Rate	60 %	100%
5% Placement Accuracy Success Rate	100%	100%
Average Cost of Operation ($Amps \cdot s/kg$)	10802.08	20245.68
Average Mechanical Work (J)	2585.25	561.69

Based on the experiments, the separate upper and lower body control method was on average 10 seconds faster than the whole-body control method. However, due to the swaying of the robot during the walking phases, the stepping strategy experienced a 40% lower placement accuracy success rate for the 2.5%. When considering the 5% margin error, both algorithms show equally full 100% success rate. The cost of operation shows better performance for the separate upper and lower body control method, since the values are integrated over time, the 10 second difference generates a considerable contrast in results. The whole-body control method shows almost double the cost of operation compared to the other control method. From the mechanical work point of view, using the whole-body instead of walking requires less mechanical work, even while taking longer to finish the task. The separate upper and lower body control method takes almost five times the mechanical work of the whole-body control strategy.

One consideration for benchmarking whole-body manipulation that was noticed during this initial study is the need for a greater variety of motions and test objects. While the initial study provided some insight into a humanoid performing a whole-body box manipulation, the variety of motions is fairly limited if only a single shelf is used as only vertical lifting motions are required. Additional shelving units could be useful to introduce lateral box manipulation motions and encourage more challenging manipulation combinations.

Improved testing for how low or high an object can be grasped from is another possible metric of interest. In this initial study the box motions were very prescribed, but more complex manipulations could be tested using a more random and automated pick and place indication system that the robot must discern in real time. Also, a greater grasping possibilities for the robot, such as handles to grip, would be useful along with varying weight. These are all important factors to consider when selecting a robot for a whole-body manipulation task as the robots ability to handle these varying motions or objects could affect the suitable applications.

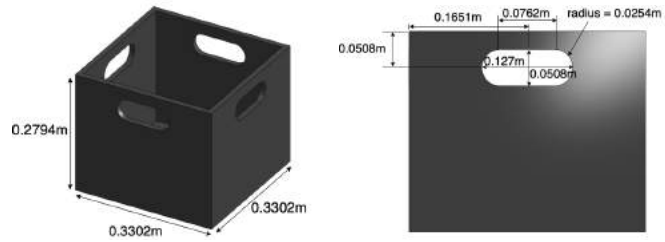
IV. A Complete Benchmark for Whole-body Manipulation

A. Test Bed

To benchmark the ability of a robot's whole-body manipulation abilities, a test bed composed of three shelving units surrounding a square floor space of 1 m^2 on three sides is proposed. This provides a constrained workspace similar to that in industrial or logistics settings. Note that the workspace is kept small as this benchmark aims at benchmarking manipulation abilities with little locomotion involved, such as only taking a few steps. The pick and carry test bed that will be developed in future work will include a larger workspace to better benchmark loco-manipulation. Each shelving unit contains three shelves at heights that can be varied due to the flexibility of the shelving unit. To benchmark whole-body manipulation motions similar to those a human would perform when moving objects around, heights below the waist, below the shoulders but above the waist and above the shoulders heights are selected. These heights are equally spaced and selected as 0.14m, 0.78m and 1.42m, respectively. This shelving setup allows for vertical, lateral and more complex manipulation combinations. Each shelf in the test bed contains a target for the location of the box to be picked from and placed at as well as an multi-colour LED strip along the front edge to indicate where the objects should be picked from and placed. Since the ability to replicate the setup is an important feature for benchmarking test beds, we propose using shelving units of the IKEA IVAR system with a depth of 0.5 m that are available worldwide, as previously noted.

The main object to be manipulated is an open top box, such as a standard milk crate, with the dimensions are 0.33 m wide by 0.33 m deep by 0.28 m tall. This object has handles that can be used for the robot to grasp or it can be grasped on either side with friction grip. Given that the top is open weights can be easily added to increase the mass of the object. The test object and dimensions can be seen in Figure 5.

A model of the test bed and key dimensions can be seen in Figure 6 with the standard milk crate for manipulation and REEM-C, the standard test robot for EUROBENCH. Further details are provided with the complete benchmark description.



(a) Outer Dimensions (b) Handle Dimensions

Fig. 5: Standard Whole-body Manipulation Benchmarking Open Top Box (Milk Crate)

B. Protocols

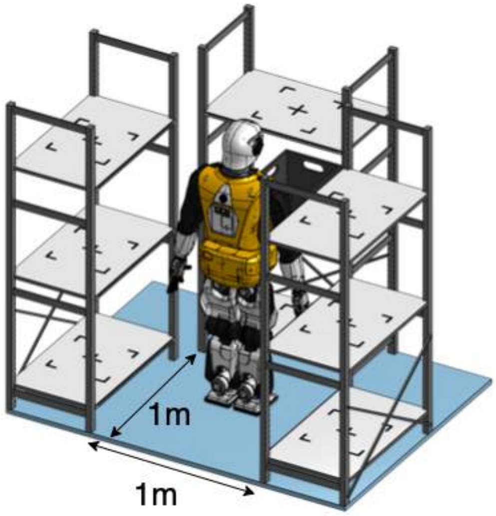
To provide a comprehensive set of tests on the robot's whole-body manipulation abilities, four protocols can be used to assess the robot's performance.

1) Protocol 1: Predefined Frontal Placement: The first protocol tests the ability for the robot to perform whole-body box manipulations on the shelf directly in front of the robot. This protocol involves picking an object from a shelf in front of the robot and placing it on another shelf in front in a predefined location for several runs. A single run contains 10 placements that are recorded for evaluation with respect to the key performance indicators defined for the test bed. After each successful completion of a run the weight of the object is increased according to the user's choice. When performing a run for any of the predefined protocols the robot has 15 minutes to complete the run. In this protocol the weight is added to the open top box described earlier.

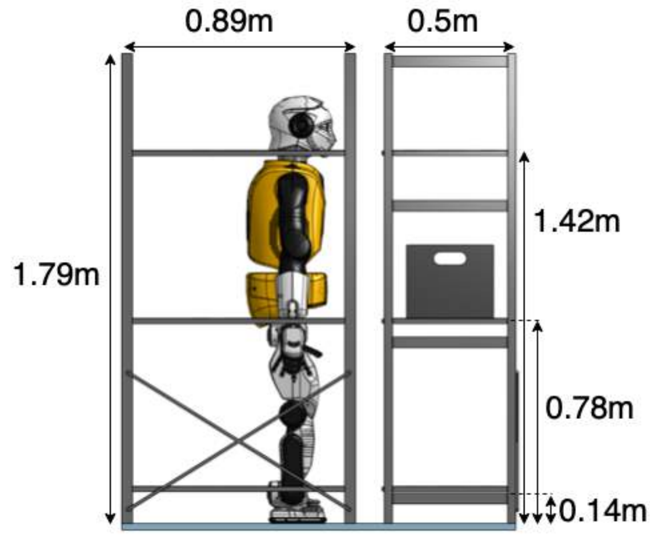
2) Protocol 2: Predefined Lateral Placement: The second protocol tests the ability for the robot to perform whole-body box manipulations on the shelves to the side of the robot. This protocol involves picking an object from a shelf in front or to the side of the robot and placing it on another shelf of the same height in front or to the side in a predefined location for several runs. In similar fashion to the first protocol, each run consists of 10 placements with a 15 minute time limit that are recorded for evaluation and weight is added to the open top box after each run as selected by the user.

3) Protocol 3: Predefined Combined Placement: The third protocol tests the ability for the robot to perform whole-body box manipulations on the shelves in front and to the side of the robot. This protocol involves picking an object from a shelf in front or to the side of the robot and placing it on another shelf in front or to the side of the robot of varying height in a predefined location for several runs. Again, similar to the first protocol, each run consists of 10 placements with a 15 minute time limit that are recorded for evaluation with weight being added to the open top box after each run according to the user.

4) Protocol 4: Variable Combined Placement: The fourth protocol tests the ability for the robot to perform whole-body box manipulations on the shelves in front and to the side of the robot in a variable manner. This protocol requires the robot to pick the object to be moved and place the object in



(a) Rear View



(b) Side View

Fig. 6: Manipulation Test Bed Model

the correct location according to visual indications from the LED lights on the front of the shelves where different colours indicate different pick and place locations. The positions in a single run vary as well as the weights of the boxes used as objects. However, the maximum weight will be only as high as what the robot can overcome. A fixed time of 20 minutes is given for this protocol and the test is complete once the 20 minutes is passed or the robot can no longer manipulate any more objects.

5) Performing a Protocol: With each protocol being composed of several runs, the rules for performing the protocols are all very similar and can be described by explaining how to perform a single run. A run is performed by placing the object at the pick location containing the desired weight. The robot can be placed anywhere inside the $1m^2$ workspace as long as all segments of the robot are within this workspace and not within the shelves. When the robot starts its motion, the time and joint data from the robot is recorded until the robot completes the box placements or time runs out. After each box placement the robot must retreat to the initial position it started from. The recorded data is then put into the desired CSV file format for performance processing. Further details, including pick and place orders, are provided with the complete benchmark description.

C. Key Performance Indicators

For each protocol performed with the robot a number of KPIs will be measured to determine the performance of the robot for the given task. These KPIs are generated as averages for all the box placements performed in a protocol. Table II contains the KPIs used for benchmarking.

It is worth noting that this benchmark is created mainly for humanoid robots. This benchmarking setup can be used for non-humanoid robots, though these robots would suffer in the human-likeness KPI and benchmarking these robots is not the main goal. Also, another separate benchmark for wearable devices and exoskeletons is provided in the EUROBENCH project for manipulation related tasks. In the EUROBENCH framework, to calculate KPIs for the robot's performance a set of CSV files from the experiments recording the joint data from the robot is provided to the software framework that automatically computes the KPIs. For the REEM-C robot, the joint data includes joint angle, joint velocity and joint current measures. In the case of whole-body manipulation a few additional result CSV files are provided detailing external details like the success rate, placement error and object weight. For further details on the benchmark please see the complete description.

V. Conclusion & Outlooks

An initial study of whole-body manipulation in a box manipulation scenario was provided to further investigate benchmarking whole-body manipulation of humanoid robots. This initial study used a 0.1 kg box on an IKEA IVAR shelf with the humanoid robot "Seven" for two different control methods to lift the box from the middle shelf to the top shelf.

This initial study allowed us to begin to compare two control methods and also identify other important aspects to consider for whole-body manipulation benchmarking. Then we proposed a complete benchmark for the whole-body manipulation scenario of EUROBENCH with a more complex test bed composed of three IVAR shelving units, objects of varying weight, four protocols of varying motion requirements and difficulty and a comprehensive set of key performance indicators.

For future works, the main goal is to construct the complete test bed and collect an initial data set for the whole-body manipulation benchmark. In the case of placement accuracy this could also be done with a motion capture system or install cameras on the test bed and use ArUco

TABLE II: Key Performance Indicators for the Bipedal Manipulation Test Bed

Name	**Description/Formulation**
Success Rate (%)	The ratio of objects placed on the correct shelf
Placement Position Error (m)	The average position error of the object placed on the shelf from the target location of the shelf
Placement Orientation Error (rad)	The average orientation error of the object placed on the shelf from the target location of the shelf
Time (s)	The average time for the correct placement of an object
Maximum Weight (kg)	The maximum weight of an object with which the robot can perform a successful placement
Shelf Heights Reached (m)	A list of the attainable shelf heights that the robot could successfully perform picks from and placements at
Mechanical Work (W)	The average absolute mechanical work calculated as in Equation 2
Cost of Operation ($Amps \cdot s/kg$)	The average energy consumed per placement and mass of objects calculated as in Equation 1
Power Consumed (W)	The average total amount of power consumed from the battery including all actuators, sensors and PCs
Human-likeness in Time (s)	The average difference in time of the robot trajectories to human trajectories performing the same task
Human-likeness in Path (m)	The average difference in position of the robot end effectors to human hands performing the same task

markers on the objects to determine the error. To determine the human-likeness KPIs, human motion capture will be performed to provide values for a comparison and metric calculation. Also, the generation of randomized locations triggered when the robot has finished a placement will need to be consider for the fourth protocol.

After gathering a complete data set for the benchmark, we want to employ optimal control strategies to improve the results on the benchmark and generate new control methods that may consider stabilization of the object while walking. An analysis of the bimanual workspace could certainly help find more optimal solutions and obtain a better score in the benchmark. To evaluate the fourth protocol perception and planning layers have to be included into the overall solution to the task. Furthermore, we would like to perform the benchmark with a second humanoid robot to compare the performance of robots, such as with the University of Waterloo's TALOS.

As part of the continued development of benchmarking scenarios we will also work on the pick and carry test bed.

ACKNOWLEDGMENT

We gratefully acknowledge funding from the Tri-Agency Canada Excellence Research Chair program and the University of Waterloo.

REFERENCES

[1] "Eurobench project." [Online]. Available: https://eurobench2020.eu/

[2] D. Torricelli, J. Gonzalez-Vargas, J. F. Veneman, K. Mombaur, N. Tsagarakis, A. J. del Ama, A. Gil-Agudo, J. C. Moreno, and J. L. Pons, "Benchmarking bipedal locomotion: A unified scheme for humanoids, wearable robots, and humans," *IEEE Robotics Automation Magazine*, vol. 22, no. 3, pp. 103–115, 2015.

[3] O. Stasse, K. Giraud-Esclasse, E. Brousse, M. Naveau, R. Régnier, G. Avrin, and P. Souères, "Benchmarking the hrp-2 humanoid robot during locomotion," *Frontiers in Robotics and AI*, vol. 5, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2018.00122

[4] "Koroibot project." [Online]. Available: https://orb.iwr.uni-heidelberg.de/koroibot/

[5] K. Chatzilygeroudis, B. Fichera, I. Lauzana, F. Bu, K. Yao, F. Khadivar, and A. Billard, "Benchmark for bimanual robotic manipulation of semi-deformable objects," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2443–2450, 2020.

[6] P. Sotiropoulos, M. A. Roa, M. F. Martins, W. Fried, H. Mnyusiwalla, P. Triantafyllou, and G. Deacon, "A benchmarking framework for systematic evaluation of compliant under-actuated soft end effectors in an industrial context," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 280–283.

[7] Y. Bekiroglu, N. Marturi, M. A. Roa, K. J. M. Adjigble, T. Pardi, C. Grimm, R. Balasubramanian, K. Hang, and R. Stolkin, "Benchmarking protocol for grasp planning algorithms," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 315–322, 2020.

[8] A. S. Morgan, K. Hang, W. G. Bircher, F. M. Alladkani, A. Gandhi, B. Calli, and A. M. Dollar, "Benchmarking cluttered robot pick-and-place manipulation with the box and blocks test," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 454–461, 2020.

[9] S. Cruciani, B. Sundaralingam, K. Hang, V. Kumar, T. Hermans, and D. Kragic, "Benchmarking in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 588–595, 2020.

[10] K. Harada, S. Kajita, K. Kaneko, and H. Hirukawa, "Dynamics and balance of a humanoid robot during manipulation tasks," *IEEE Transactions on Robotics*, vol. 22, no. 3, pp. 568–575, 2006.

[11] K. Shen, X. Li, H. Tian, T. Matsuno, and M. Minami, "Analyses of biped walking posture by dynamical-evaluating index," *Artificial Life and Robotics*, vol. 23, no. 2, pp. 261–270, 2018.

[12] T. Sugihara and Y. Nakamura, "Whole-body Cooperative Balancing of Humanoid Robot using COG Jacobian," *The Proceedings of JSME annual Conference on Robotics and Mechatronics (Robomec)*, vol. 2002, no. 0, p. 102, 2017.

[13] A. Di Fava, K. Bouyarmane, K. Chappellet, E. Ruffaldi, and A. Kheddar, "Multi-contact motion retargeting from human to humanoid robot," *IEEE-RAS International Conference on Humanoid Robots*, pp. 1081–1086, 2016.

[14] C. M. Dellin, K. Strabala, G. C. Haynes, D. Stager, and S. S. Srinivasa, "Guided manipulation planning at the darpa robotics challenge trials," *Springer Tracts in Advanced Robotics*, vol. 109, pp. 149–2016, 2016.

[15] S. J. Jorgensen, M. Vedantam, R. Gupta, H. Cappel, and L. Sentis, "Finding locomanipulation plans quickly in the locomotion constrained manifold," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 6611–6617.

[16] M. Murooka, I. Kumagai, M. Morisawa, F. Kanehiro, and A. Kheddar, "Humanoid loco-manipulation planning based on graph search and reachability maps," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1840–1847, 2021.

[17] P. Ferrari, M. Cognetti, and G. Oriolo, "Humanoid whole-body planning for loco-manipulation tasks," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 4741–4746.

[18] Y. Yang, W. Merkt, H. Ferrolho, V. Ivan, and S. Vijayakumar, "Efficient humanoid motion planning on uneven terrain using paired forward-inverse dynamic reachability maps," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2279–2286, 2017.

[19] S. Kajita, F. Kanehiro, K. Kaneko, K. Fujiwara, K. Harada, K. Yokoi, and H. Hirukawa, "Biped walking pattern generation by using preview control of zero-moment point," in *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, vol. 2, 2003, pp. 1620–1626 vol.2.

[20] P. Beeson and B. Ames, "Trac-ik: An open-source library for improved solving of generic inverse kinematics," in *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, 2015, pp. 928–935.

[21] A. Laurenzi, E. M. Hoffman, L. Muratore, and N. G. Tsagarakis, "Cartesi/o: A ros based real-time capable cartesian control framework," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 591–596.