TelePose: a Teleoperation Robotic System based on External Range Sensing – Application on the Centaur-like Robot CENTAURO

Emily-Jane Rolley-Parnell[1], Dimitrios Kanoulas[2], Arturo Laurenzi[2], Brian Delhaisse[2], Leonel Rozo[2], Darwin G. Caldwell[2], Nikos G. Tsagarakis[2]

***Abstract*— In this short paper, we present the TelePose system, which is a real-time, bi-manual teleoperation method to control limbed robots, through their end effectors, in the Cartesian space. This is done with an external to the robot RGB-D sensor that tracks the skeleton of a human to be imitated. The methods splits into two steps. First, an RGB image of a human is used to extract its skeleton, using the recently developed OpenPose package. Secondly, the depth of the human hand poses are extracted using the corresponding depth image. The extracted 6 Degrees-of-Freedom (DoF) hand poses are fed into a whole-body Cartesian control system of the robot's end-effectors to drive their goal pose. We demonstrate some results on the newly developed centaur-like robot CENTAURO, where a human is using TelePose to control the robot from distance and to drive some manipulation tasks.**

I. Introduction

Robotic manipulation is an important aspect for completing various tasks that can help humans in regular or extraordinary cases, such as lifting heavy items or turning valves in destroyed buildings. While the development of fully autonomous systems is the ultimate goal in robotics [1], it is still hard to apply autonomous manipulation in unstructured and uncertain environments. For this reason, teleoperation [2] gives an intermediate solution to achieve semi-autonomous performance, with the aid of a human (i.e., the teleoperator). The type of teleoperation varies and depends on the task to be completed and the used sensors. For example, there are methods that have been used in the past to provide exact human poses to the system that were mapped from the human to the robot, such as body-suits of IMU sensors [3], exoskeletons [4], or motion-capture cameras [5]. These methods are usually either expensive or require careful calibration, since they often map the full human joint space on the robot. We have used an IMU body-suit to achieve various manipulation tasks through teleoperation with the humanoid robot WALK-MAN and we faced big drifts during the process[6], [7]. An alternative way is to use GUI's [8], [9] or joystick-like devices [10] to control the robot, which requires more autonomy development than purely human mimicking teleoperation systems. In this work we aim at a cheap system (i.e. an RGB-D sensor) that will track the teleoperator's skeleton and map its end-effector poses onto the robot.

[1]University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, United Kingdom. `erolleyp@gmail.com`

[2]Humanoid and Human-Centered Mechatronics Department, Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163, Genova, Italy. {`Dimitrios.Kanoulas, Arturo.Laurenzi, Brian.Delhaisse, Leonel.Rozo, Darwin.Caldwell, Nikos.Tsagarakis`}`@iit.it`

II. System Overview

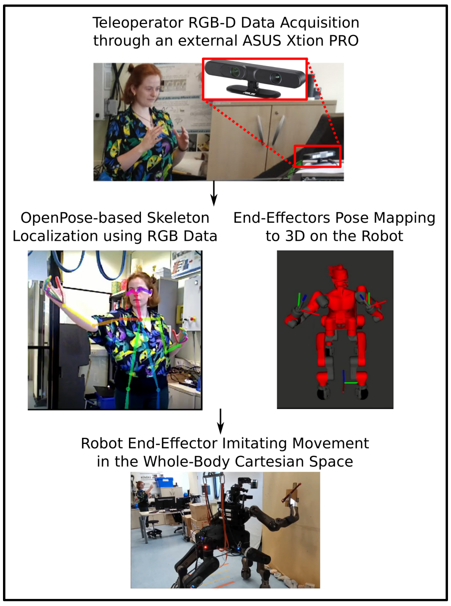

Fig. 1: System overview.

We propose a system (Fig. 1) that combines visual skeleton extraction and robot end-effector teleoperation in 3D for limbed robots. The system is divided into the following four parts:

A. RGB-D Sensing

Aiming to a low-cost solution for tracking a human body, we utilize one of the most commonly used RGB-D sensors, e.g., the ASUS Xtion PRO. The sensor works in 30 frame per second providing both RGB images and the registered depth data. The 640×480 resolution is enough for tracking human parts in $1-2$m distances. Calibration errors between the RGB and the depth sensor may cause instabilities during tracking. For this reason we always apply real-time nearest neighborhood search in the point clouds, instead of using a single point/pixel in the scene.

B. 2D Skeleton Extraction with OpenPose

OpenPose [11], is an impressive system that was recently introduced and is able to provide human skeleton poses in 2D, using a feed-forward deep neural network system. The extracted 2D poses of human joints are produced in near real-time ($8-10$Hz). The method relies on body-part detection and association, using Part Affinity Fields and Confidence Maps. The original network is trained using nearly 5.5 hours of human data from different kind of actions. We use this system as black box to extract 2D human skeleton poses, given OpenPose's robustness on occlusions. The system requires only RGB images of the teleoperator.

C. 3D Human Hand Mapping

Given the pixels that correspond to the human parts through the OpenPose, we use the corresponding depth image, acquired from the RGB-D sensor, to map the pixels to 3D points in the world frame. We face two issues during this process: 1) there are pixels that correspond to unset depth values (NaNs) or occluded body parts, 2) there are mismatched pixels-points due to calibration issues between the RGB and the depth sensor. For both cases, to avoid robot jumps during tracking, we apply an organized nearest neighborhood search in the point clouds that give better and more reliable estimations of the 3D positions. If the value remains NaN, then we handle this during the robot control.

D. Robotic Hand Control

To bi-manually control the robotic hand end-effectors, we should provide a Cartesian position of each wrist relative to the human world frame. The reference point is the initial point of the hands during tracking. Thus, the movement of the robot hands is calculated from the difference between the frames that moved in the robot frame. In this way, simplicity and effectiveness are achieved, preventing big drifts. Any frame in which the human hands are not correctly localized and tracked are removed, to prevent fast and not smooth robot moves. The robot is controlled using the hard real-time middleware XBotCore [12] in combination with the OpenSoT [13] kinematic solver for hierarchical whole-body robot control, subject to constraints, such as joint limits or joint velocities. We have developed a Cartesian Interface module (link: github.com/ADVRHumanoids/CartesianInterface) on top of the XBotCore, to control the robot's end-effectors according to the calculated and mapped 3D human hand poses in real-time.

III. Robot Demonstrations

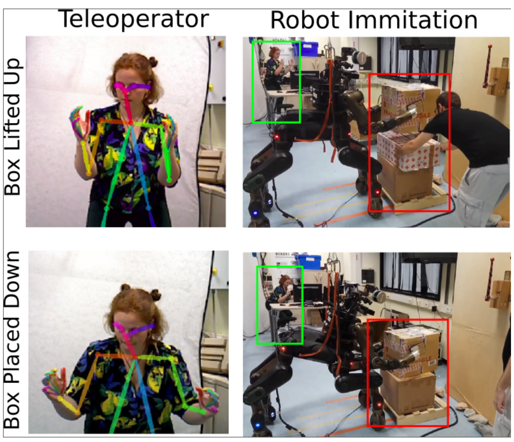

Fig. 2: Box lifting and placing task.

To demonstrate the capabilities of our system for a bi-manual teleoperated manipulation task, we use the centaur-like robot CENTAURO, which has two 7-DoF manipulator arms. The task to be completed is to have the robot lifting a box from a stack, to allow a human worker get the second box in the task. Finally, the box is placed carefully back to the stack. The demonstration is visualized in Fig. 2. Videos of more experimental results can be found under our webpage: `https://sites.google.com/view/telepose`

IV. Conclusions

This paper presents the TelePose system, for controlling a limbed-robot's end-effectors, through teleoperation. An external cheap RGB-D sensor was used to localize the teleoperator skeleton in real-time, while the robot is controlled in whole-body mode to achieve manipulation tasks, such as box lifting. We plan to develop further this package towards human-robot interaction or easily collection of training data for reinforcement learning purposes.

ACKNOWLEDGMENT

This work is supported by the CENTAURO (grant agreements no 644839) EU project. The two Titan Xp GPUs used for this research were donated by the NVIDIA Corporation. The authors would also like to thank Enrico Mingo Hoffman, Luca Muratore, and Giuseppe Rigano for their help with the OpenSoT and XBotCore packages on the robot.

References

[1] D. Kanoulas, J. Lee, D. G. Caldwell, and N. G. Tsagarakis, "Center-of-Mass-Based Grasp Pose Adaptation Using 3D Range and Force/Torque Sensing," *IJHR*, p. 1850013, 2018.

[2] M. A. Goodrich, J. W. Crandall, and E. Barakova, "Teleoperation and Beyond for Assistive Humanoid Robots," *Reviews of Human Factors and Ergonomics*, vol. 9, no. 1, pp. 175–226, 2013.

[3] N. Miller, O. C. Jenkins, M. Kallmann, and M. J. Mataric, "Motion Capture from Inertial Sensing for Untethered Humanoid Teleoperation," in *IEEE/RAS Humanoids*, vol. 2, 2004, pp. 547–565.

[4] I. Sarakoglou *et al.*, "HEXOTRAC: A Highly Under-Actuated Hand Exoskeleton for Finger Tracking and Force Feedback," in *IEEE/RSJ IROS*, 2016, pp. 1033–1040.

[5] J. Silvério, S. Calinon, L. D. Rozo, and D. G. Caldwell, "Learning competing constraints and task priorities from demonstrations of bimanual skills," *CoRR*, vol. abs/1707.06791, 2017.

[6] N. G. Tsagarakis *et al.*, "WALK-MAN:A High Performance Humanoid Platform for Realistic Environments," *JFR*, 2016.

[7] F. Negrello *et al.*, "The WALK-MAN Robot in a Postearthquake Scenario," *IEEE Robotics and Automation Magazine (RAM)*, 2018.

[8] P. Kaiser *et al.*, "An Affordance-Based Pilot Interface for High-Level Control of Humanoid Robots in Supervised Autonomy," in *IEEE-RAS Humanoids*, 2016, pp. 621–628.

[9] P. Balatti, D. Kanoulas, G. F. Rigano, L. Muratore, N. G. Tsagarakis, and A. Ajoudani, "A Self-tuning Impedance Controller for Autonomous Robotic Manipulation," in *IEEE/RSJ IROS*, 2018.

[10] T. Rodehutskors, M. Schwarz, and S. Behnke, "Intuitive Bimanual Telemanipulation under Communication Restrictions by Immersive 3D Visualization and Motion Tracking," in *IEEE-RAS Humanoids*, 2015.

[11] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in *CVPR*, 2017.

[12] L. Muratore, A. Laurenzi, E. Mingo Hoffman, A. Rocchi, D. G Caldwell, and N. Tsagarakis, "XBotCore: A Real-Time Cross-Robot Software Platform," in *IEEE IRC*, 2017, pp. 77–80.

[13] E. Mingo Hoffman, A. Rocchi, A. Laurenzi, and N. G. Tsagarakis, "Robot Control for Dummies: Insights and Examples using OpenSoT," in *17th IEEE-RAS Humanoids*, 2017, pp. 736–741.