Track Back: A Human Robot Movement Installation Utilising Unity Digital Twin and Human Bio-mimicry

John McCormick, Jagannatha Charjee Pyaraka
Swinburne University of Technology
Naarm / Melbourne Australia
jmccormick@swin.edu.au, jagannathacharjeepya@swin.edu.au

Abstract

Track Back is a human-robot art installation that investigates both live human-robot interactions as well as human movement embedded in historical video repositories as sources for the robot's movement. The artwork leverages the unity game engine as a mediating platform for the robot to access human movement and repurpose it for its own body. This allows the robot's movement to be procured from multiple sources. The unity environment can also be used to connect robots of different morphologies to the movement sources enabling flexibility in the engagement between human and robot.

Keywords

Cultural Memory, Creative Robotics, Machine learning, Human Robot Learning, Human Robot Installation, Movement Tracking, Unity, Digital Twin, Mediapipe, Pose Tracking

Introduction

What examples do humans set non-human agents in the agent's quest to better understand the world and people? How do they learn to express themselves in ways humans can comprehend as well as in their own emerging syntax? Body movement is essential to any embodied being and their understanding and engagement with the world. Embodied experience not only informs how we experience the world but social and cultural meaning is inscribed on bodies and colours their expressivity. Track Back offers a robot agent the possibility to mimic the nuances of expressive movement from direct human encounters as well as the possibilities afforded by tracking human movement back through time in archives of online videos of both mundane and seminal moments in history. The robot delves back into human cultural memory to compile its own expressive movement, seeking a unique identity of its own. Track Back is part of a larger body of work placing robots in social situations to develop better understanding of human – robot interactions in social and cultural settings.

Background

Using human movement to enhance a robot's movement capabilities has had many approaches and rationales. Teleoperation for remote and dangerous environments, teaching a robot through learning by demonstration, assigning human-like behaviour to a robot to encourage movement recognition and empathy. Stanton et al used used an xsens motion capture suit and localized joint neural networks to teleoperate a humanoid nao robot. [1] Wang et al. used a motion capture system to teleoperate a BHR-6 humanoid and Vongchumyen et al. employed a Kinect to teleoperate their humanoid robot. [2, 3] Masumori et al. used human behaviour imitation to both immediately respond to human interactors and build a store of movement sequences as its memory for later use. This used a humanoid with torso, head and arms to perform the upper body movements it learned, a process they term personogenesis. [4] Mediapipe has been used for human movement tracking for replicating robot movement as in Altayeb's operation of a robotic arm and fingers. [5] The use of motion capture and more recently single camera machine vision solutions, provides a means of quickly prototyping complex movement for robots.

Unity game engine to mediate movement

The use of digital environments to simulate robot movement is commonplace as it affords means of prototyping movement without risk to the physical robot or human interactors. There are numerous dedicated robot simulation environments including Webots, Gazebo, CoppeliaSim and Rviz. Unity itself has the Unity Robotics Hub which provides robot simulation connected to the Robot Operating System (ROS). [6] Track Back uses Unity rather than a traditional robot simulation environment for a number of reasons. Typically, simulation environments are used to test developments on a virtual robot before the programs are transferred to the physical robot for deployment. Track Back, being a live interactive installation, required a live environment connecting the human and robot movement activity in realtime. A typical robot simulation environment could potentially perform this task, however Unity provides a content driven approach with extremely flexible means of switching movement input sources.

The unity environment used a humanoid character as the mediating body for movement derived from either live human tracking input or human movement contained in videos sourced online. Having a single virtual body to mediate the human derived movement and suitably filter the movement

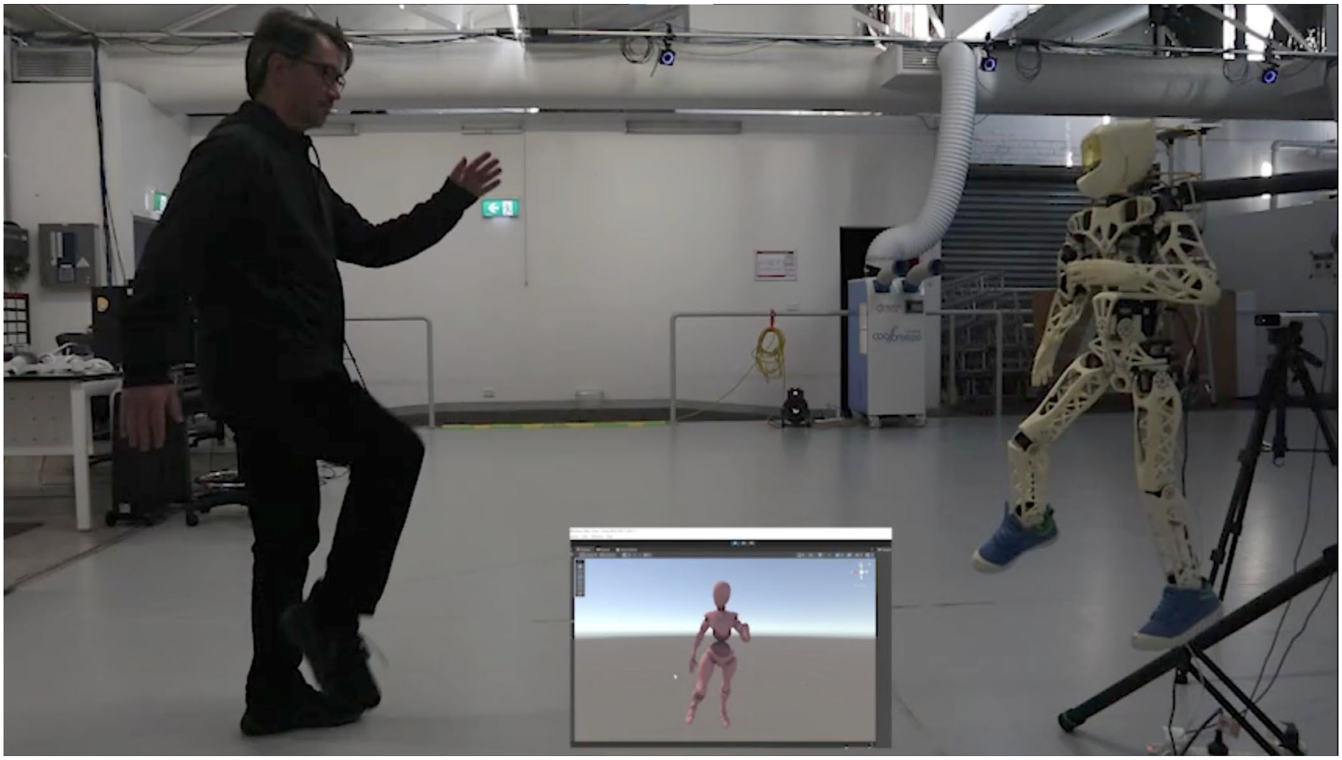

Figure 1. Using a single Kinect camera to recognize human movement and transfer the movement via a virtual humanoid in Unity to a real humanoid robot. The live Unity scene is seen in the lower centre of the image. The movement is filtered to adapt the range of motion, data sample rate and selected joints depending on the morphology of the desired robot.

for different robot morphologies created a flexible environment for rapidly testing different movement sources and robot bodies.

Considerations for Track Back as an interactive installation included ease of engagement for human participants, the ability to change from live movement tracking to recorded video streams and consistent data streams for the robot that filtered human movement to the range of the robot body. Markerless movement tracking was essential to achieve these requirements.

Live movement tracking with Azure Kinect

For the live movement tracking of human interactors, the Microsoft Azure Kinect was used using the Microsoft body tracking SDK within Unity. This method generated joint rotations based on the human movement which were applied to the virtual 3D character in Unity (Figure 1). The joint rotations of the virtual character were then sent to the robot using the UDP transport protocol. The joint rotations had a number of parameters that could be adjusted before sending to the robot. Adjusting the sample rate gave control over the frequency of joint updates sent to the robot. Joints could be turned off and on in the transmitted UDP stream for testing particular joints. Joint offsets and range limiters of the virtual characters joint rotations gave further means of adjusting the data between the original human data and that sent to the robot.

Tracking human movement in videos with Mediapipe

Analyzing movement from historical videos required a different approach. The Mediapipe Machine Learning (ML) framework was used, in particular the pose tracking ML model. Mediapipe was incorporated into Unity with the ability to use either the live camera or recorded video streams. While mediapipe could also potentially replace the Azure Kinect for the live tracking, its tracking results were found to be inferior to the Kinect while developing the artwork. More development in future will enable the entire work to use the mediapipe framework.

Mediapipe presented some major challenges to achieve tracking of human movement in videos and presenting this in a format usable by the virtual character in Unity. Mediapipe attempts to detect “landmarks” on human bodies, of interest for this work were the joint landmarks of the human torso and limbs. Mediapipe returns these landmarks as 3 dimensional positions whereas the Unity character requires 3 dimensional rotations for its joint movements. To overcome this challenge a number of approaches were tested.

Ultimately the FinalIK Inverse Kinematcs (IK) library was used to generate joint rotations for the character based on the joint positions presented by Mediapipe. This solution was quite effective though it did introduce inaccuracies in comparison to the original human movement. However, with the bodies in motion and with the filtering options available, the robot movement read alongside the source videos gave a clear reading of the intent of the robot.
The individual robot motors drive a single joint axis which matches joints such as the knees and elbows better than more complex ball and socket joints such as the shoulder and hips. Reducing these complex joints to separate single axis rotations led to further differences between the original human movement and the interpretation by the robot.

ISEA Thematic connections

Track Back connects with the ISEA themes Shifting Temporalities, Resilient Stories and Speculative Practices. Shifting Temporalities: The robot's current physical expression is directed at interactions with the humans present but can also be informed by the past in gestures and actions gleaned from online repositories. The robot looks to videos of humans in social and cultural situations as inspiration for its own expressivity. What of the past still has relevance to a robot's current interactions with humans as it draws on situated movement for inspiration? Resilient Stories: The interactive performance by the robot as it responds to humans is a non-linear encounter with the robot drawing on cultural memory contained in online videos. New realtime narratives are devised as the robot draws on movement from the past to inform its responses to human movement of the present. Speculative Practices: Track Back explores the embodied nature of non-human agents drawing on human body movement as inspiration for social and culturally meaningful expression as a physical language that suits their own body morphology.

Track Back is a human - robot performance / installation with a humanoid robot at the centre of interactive encounters with the human visitors. The robot recognises the presence and body postures of humans and attempts to mimic and communicate through its own embodied morphology. The robot can also draw on past video of humans in social and cultural situations that have been gleaned from the internet. The robot seeks to make meaning in its encounters with humans from the present and past of human movement expressions. The robot attempts to borrow social and cultural context from the past videos and refashion these expressive communications in its own image and to its own morphology.

The robot uses the azure kinect or mediapipe machine learning framework to recognize people's postures and movement and waits to interact with them. At times it will attempt to mimic the human's actions in a simple acknowledgement to encountering the human's presence. To further the nonverbal narrative it draws on videos found on the internet of both mundane social activity as well as seminal events of human physical expression. The robot tracks the movement of the humans represented in the videos and appropriates their movement for its own, mapped to its own body. The robot alternates between these mimicking and recollection performances until another human moves in front of it and proceeds to interact with them.

Track Back responds to the ISEA 2024 theme on a number of fronts. Everywhen describes a place when past, present and future are co-habiting in this space in these encounters. Track Back sees the robot engaging with the past of human movement expression in historical videos, the present of human expression with human gallery participants and the future as the robot speculates its own emerging expression of social and cultural encounters. In this the work engages directly with the sub-theme Shifting Temporalities. Track Back uses non-verbal movement traditions embedded in past videos to create non-linear interactive communication structures to further interpersonal narratives. This method engages with the sub-theme Resilient Stories.

Conclusion

Track Back comprises playful methods for robots to develop their movement in response to human visitors to the installation, using markerless motion tracking to appropriate human movement for the robot's own movement responses. The Unity game environment provides a common platform for consolidating both the human and robot movement. The same Unity environment is also able to use mediapipe's pose tracking capabilities to identify human movement in online historical videos as alternative sources for its own movement. Ongoing research is incorporating other robots into the environment and using the filtered movement as source data for further learning.

References

1. Stanton, C., A. Bogdanovych, and E. Ratanasena. *Teleoperation of a humanoid robot using full-body motion capture, example movements, and machine learning*. in *Proceedings of Australasian Conference on Robotics and Automation,* . 2012. Victoria University of Wellington, New Zealand.
2. Wang, C., et al. *Intuitive and Versatile Full-body Teleoperation of A Humanoid Robot*. 2021. IEEE.
3. Vongchumyen, C., et al. *Teleoperation of Humanoid Robot by Motion Capturing Using KINECT*. 2018. IEEE.
4. Masumori, A., N. Maruyama, and T. Ikegami, *Personogenesis Through Imitating Human Behavior in a Humanoid Robot "Alter3"*. Frontiers in Robotics and AI, 2021. **7**.
5. Altayeb, M., *Hand Gestures Replicating Robot Arm based on MediaPipe.* Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 2023. **11**: p. 727-737.
6. Unity. *Unity Robotics-Hub*. 2022; Available from: https://github.com/Unity-Technologies/Unity-Robotics-Hub.

Acknowledgements

This project is supported by the Victorian Government through Creative Victoria.