ShadowDancXR: Body Gesture Digitization for Low-cost Extended Reality (XR) Headsets

Difeng Yu
The University of Melbourne
difeng.yu@student.unimelb.edu.au

Weiwei Jiang
The University of Melbourne
weiwei.jiang@student.unimelb.edu.au

Chaofan Wang
The University of Melbourne
chaofan.wang@student.unimelb.edu.au

Tilman Dingler
The University of Melbourne
tilman.dingler@unimelb.edu.au

Eduardo Velloso
The University of Melbourne
eduardo.velloso@unimelb.edu.au

Jorge Goncalves
The University of Melbourne
jorge.goncalves@unimelb.edu.au

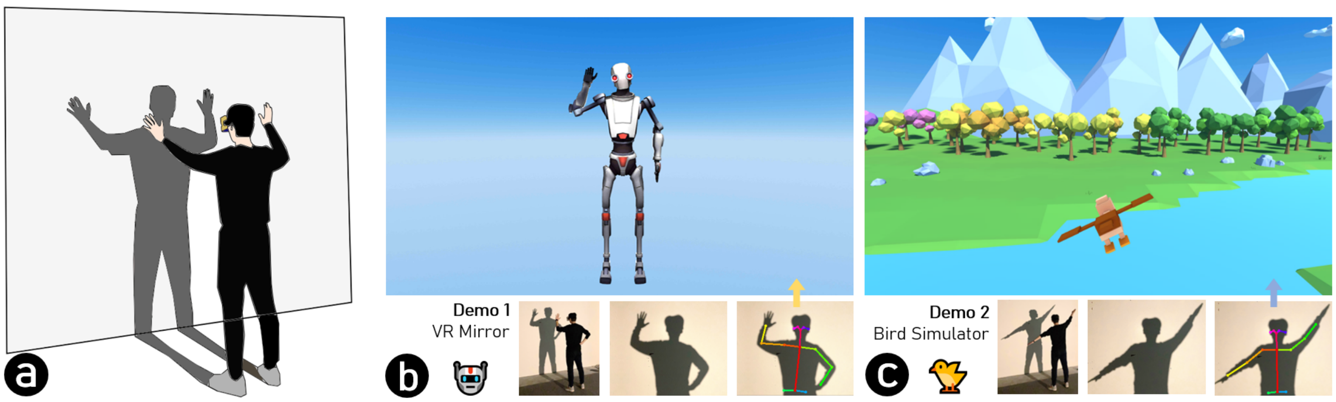

Figure 1: (a) ShadowDancXR uses the front-facing camera of the smartphone in low-cost XR headsets (e.g., Google Cardboard) to reconstruct users' body gestures through projected shadows. We designed two interactive experiences based on ShadowDancXR: (b) allows users to see themselves as robot Kyle who will mimic their body postures through VR Mirror, and (c) enables users to "fly" around a virtual world as a bird with multiple arm postures.

ABSTRACT

Low-cost, smartphone-based extended reality (XR) headsets, such as Google Cardboard, are more accessible but lack advanced features like body-tracking, which limits the expressiveness of interaction. We introduce ShadowDancXR, a technique that relies on the front-facing camera of the smartphone to reconstruct users' body gestures from projected shadows, which allows for wider range of interaction possibilities with low-cost XR. Our approach requires no embedded accessories and serves as an inexpensive, portable solution for body gesture digitization. We provide two interactive experiences to demonstrate the feasibility of our concept.

ISS '20 Companion, November 8–11, 2020, Virtual Event, Portugal

ACM ISBN 978-1-4503-7526-9/20/11…$15.00
https://doi.org/10.1145/3380867.3426222

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; *Interaction design*; • **Computing methodologies** → *Tracking*.

KEYWORDS

Body tracking; motion sensing; light; shadow; virtual reality; extended reality

ACM Reference Format:

Difeng Yu, Weiwei Jiang, Chaofan Wang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. ShadowDancXR: Body Gesture Digitization for Low-cost Extended Reality (XR) Headsets. In *Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces (ISS '20 Companion), November 8–11, 2020, Virtual Event, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3380867.3426222

1 INTRODUCTION

Low-cost extended reality (XR) headsets, including notable examples like Google Cardboard[1] and Samsung Gear VR[2], have become increasingly popular in the past few years. Such devices allow users to access spatial computing platforms by using their smartphones

[1] https://arvr.google.com/cardboard/
[2] https://www.samsung.com/global/galaxy/gear-vr/

as displays. Because they are relatively inexpensive and more convenient to carry when compared to bulkier XR equipment, they hold great potential to enable true ubiquity of XR technology. However, the functionalities enabled by these headsets are somewhat limited. For example, they cannot track the user's hand or body position, which leads to the expressiveness and usefulness of the device to be largely constrained.

There are different ways to extend the body tracking capabilities of low-cost XR headsets. Existing solutions are mostly based on the installation of new cameras, either based on retro-reflective markers [4] or computer vision techniques (e.g., Microsoft Kinect) [5, 6] to estimate human pose. Other common methods rely on body-worn suits with a set of IMU sensors to reconstruct body pose [3]. While effective, these approaches can increase the cost of the XR system rapidly. A recent prototype called Mecap [1] uses the smartphone's rear-facing camera on the low-cost XR headsets with two attached reflective mirrors to support whole-body digitization inexpensively. However, specific accessories (spherical mirrors) are required for this method, and they add weight to the headsets, which can induce increased fatigue.

We present ShadowDancXR—a cost-free solution to enable body gesture digitization for low-priced XR headsets. Our approach leverages the images captured by the smartphone's front-facing camera to reproduce body postures from projected shadows on the walls, which are made available by environmental lighting such as desk lamps (see Figure 1a). In this work, we introduce the implementation of a prototype and two applications, which allow remote participation and experiencing of our demonstration.

2 SYSTEM IMPLEMENTATION

The current implementation consists of a Google Cardboard headset, a smartphone, and a laptop PC. The front-facing camera of the smartphone is used to capture the shadow image. The image is then sent to a server over Wi-Fi. We use `tf-pose-estimation`[3], which is based on OpenPose [2], to approximate the 2D body key points from the shadow image, where the `mobilenet_thin` model can be used for fast prediction, and the `cmu` model can be applied to improve recognition accuracy. The key points data are then sent back to the server and retrieved by the smartphone for user pose reconstruction. The camera-captured image is re-scaled to the resolution of 102×92 for Internet transmission and up-sampled to 432×368 (the size of which the model was trained on) for pose estimation. This is to minimize the latency caused by the data transmission via the Internet while maintaining an accurate estimation. In our prototype, the round trip takes around 1600 ms for `mobilenet_thin` and 2400 ms for `cmu`, largely due to the Internet transmission (1400 ms). We have also tried algorithms to recover 3D body pose from the 2D shadow image, but the results are currently unusable, and thus the function is not included in the final mock-up. Our PC-end implementation is based on an Intel Core i7-9750H CPU @ 2.60GHz, rather than a dedicated GPU, to increase accessibility. The smartphone we use in our prototype is Google Pixel 2.

[3]https://github.com/ildoonet/tf-pose-estimation

3 EXPERIENCE DESIGN

We developed two applications based on ShadowDancXR to support the remote experiencing of our prototype. We have established a GitHub page containing the software and detailing the set up process, which is available on: https://github.com/Davin-Yu/ShadowDancXR-ISS.

3.1 VR Mirror

In this demonstration, users will be fully immersed inside a virtual environment with Google Cardboard VR and will face towards the projected shadow of their body (see Figure 1b bottom). When they move their upper limbs, the robot Kyle will mimic their body posture, as if the users are looking through a mirror and the robot is themselves (see Figure 1b up). We applied animation smoothening to compensate for latency.

3.2 Bird Simulator

In this experience, users can fly over a virtual landscape from a bird's view (see Figure 1c up). Similar to the previous demo, the users will be standing in front of their projected shadow with a VR headset. The bird moves forward and downward at a constant speed. To make the virtual bird fly higher, users have to wave up and down with their arms. When the upper limbs are tilting towards a particular direction, like in Figure 1c bottom, the bird will adjust its wings and turn to the indicated orientation.

4 CONCLUSION AND FUTURE WORK

We proposed ShadowDancXR, a technique to reconstruct body gestures from the projected shadows captured by smartphones, to enable body-tracking ability for low-cost extended reality (XR) headsets. We presented two demonstrations, which allow remote/online participation and experience. Future work can try to perform pose estimation on smartphones locally (e.g., using PoseNet[4]) to shorten the latency and improve the usability of this approach.

REFERENCES

[1] Karan Ahuja, Chris Harrison, Mayank Goel, and Robert Xiao. 2019. MeCap: Whole-Body Digitization for Low-Cost VR/AR Headsets. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 453–462. https://doi.org/10.1145/3332165.3347889

[2] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[3] Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J. Black, Otmar Hilliges, and Gerard Pons-Moll. 2018. Deep Inertial Poser: Learning to Reconstruct Human Pose from Sparse Inertial Measurements in Real Time. *ACM Trans. Graph.* 37, 6, Article 185 (Dec. 2018), 15 pages. https://doi.org/10.1145/3272127.3275108

[4] Matthew Loper, Naureen Mahmood, and Michael J. Black. 2014. MoSh: Motion and Shape Capture from Sparse Markers. *ACM Trans. Graph.* 33, 6, Article 220 (Nov. 2014), 13 pages. https://doi.org/10.1145/2661229.2661273

[5] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 127–136. https://doi.org/10.1109/ISMAR.2011.6092378

[6] Helge Rhodin, Christian Richardt, Dan Casas, Eldar Insafutdinov, Mohammad Shafiei, Hans-Peter Seidel, Bernt Schiele, and Christian Theobalt. 2016. EgoCap: Egocentric Marker-Less Motion Capture with Two Fisheye Cameras. *ACM Trans. Graph.* 35, 6, Article 162 (Nov. 2016), 11 pages. https://doi.org/10.1145/2980179.2980235

[4]https://www.tensorflow.org/lite/models/pose_estimation/overview