

Contact4D: A Video Dataset for Whole-Body Human Motion and Finger Contact in Dexterous Operations

– Supplementary Materials –

Anonymous 3DV submission

Paper ID 348

1. Contact4D

1.1. Dataset Statistics

Contact4D is a 100-minute video dataset collected using 18 equidistant, synchronized GoPro cameras, one pair of Aria glasses [1], and our customized finger contact sensor. It consists of 2M images, divided into 1.6M images for training and 405K images for testing, with ground-truth finger contact annotations. We manually clipped the videos into 350 sequences, ensuring each sequence is at least 15 seconds (300 frames) long for temporal continuity. Figure 1 shows images captured by our multi-view system.

1.2. Contents

We release camera intrinsics and extrinsics for all views estimated via structure-from-motion (SfM) [8]. To express poses in metric scale, we provide global scaling and rotation matrices that align the SfM frame to the Aria VIO [3] frame using a Procrustes alignment between the COLMAP Aria trajectory and the Aria VIO trajectory. For each image, we provide 2D human/hand bounding boxes, 2D/3D whole-body keypoints (COCO [4] format), and human meshes in SMPL [5] and MANO [7]. The 2D boxes are obtained by projecting SMPL/MANO vertices, and the 2D/3D poses are derived via multi-view triangulation. Mesh annotations include pose, shape, and camera-translation parameters together with the corresponding vertices. In addition, we provide per-finger contact annotations collected with our custom contact sensor. Figure 2 illustrates the annotations included in Contact4D.

1.3. Data Processing Details

Our pipeline is highly efficient: all stages run in parallel except the temporal smoothing module. This design enables processing one video sequence ($\sim 5,700$ images) processes in about 30 minutes on a single RTX 2080 GPU. We then perform a frame-by-frame quality check. Because the system is multi-view, it suffices to verify the 3D annotations in

a single reference camera per sequence, which greatly reduces manual effort. When errors are found, we correct the corresponding 2D keypoints and re-triangulate to recover accurate 3D keypoints.

2. Contact Sensor

2.1. Device Specifications

Our finger contact sensor integrates an Arduino MKR Zero [2] mounted on a protoboard and multiple force sensing resistors (FSRs) [10]. Each FSR is connected to a JST connector using fully insulated enameled wires, ensuring that electrical shorts are avoided even when the wires are in close proximity. We further put a small amount of silicon adhesive near the pins of the FSRs to stabilize the data recording. In the circuit design, each FSR operates as part of a voltage divider paired with a 1M ohm resistor. When pressure is applied, the FSR’s resistance drops, causing a change in the divided voltage. The Arduino measures this voltage change and interprets it as the contact signal. Figure 3 shows a sample image of the contact sensor.

2.2. Attachment Method

Each FSR is attached to one of the subject’s fingertips. First, we affix the FSR onto a thin film of thermoplastic polyurethane (TPU) [9] using regular glue. Then, we apply Pros-Aide Adhesive [6], a commercially available water-based adhesive, to stick the assembly (TPU +FSR) to the finger. Please refer to Figure 3 in the main Paper.

2.3. Wearable Sensor

The FSR sensors are affixed to the individual’s fingers using the method described above. To prevent wire entanglement, the Arduinos are secured to the subject’s upper arms with loop straps. Additionally, a sponge is affixed under each Arduino to avoid direct skin contact. Figure 4 shows a subject wearing our contact sensors. During capture, the sensor will be covered by the subject’s sleeve and remain invisible.



Figure 1. **Multi-View Capture System.** From a single view, finger keypoints are extremely challenging during object manipulation due to severe occlusion. Our multi-view system makes joints blocked from one viewpoint can be seen from another, and therefore is able to provide accurate joint annotations.

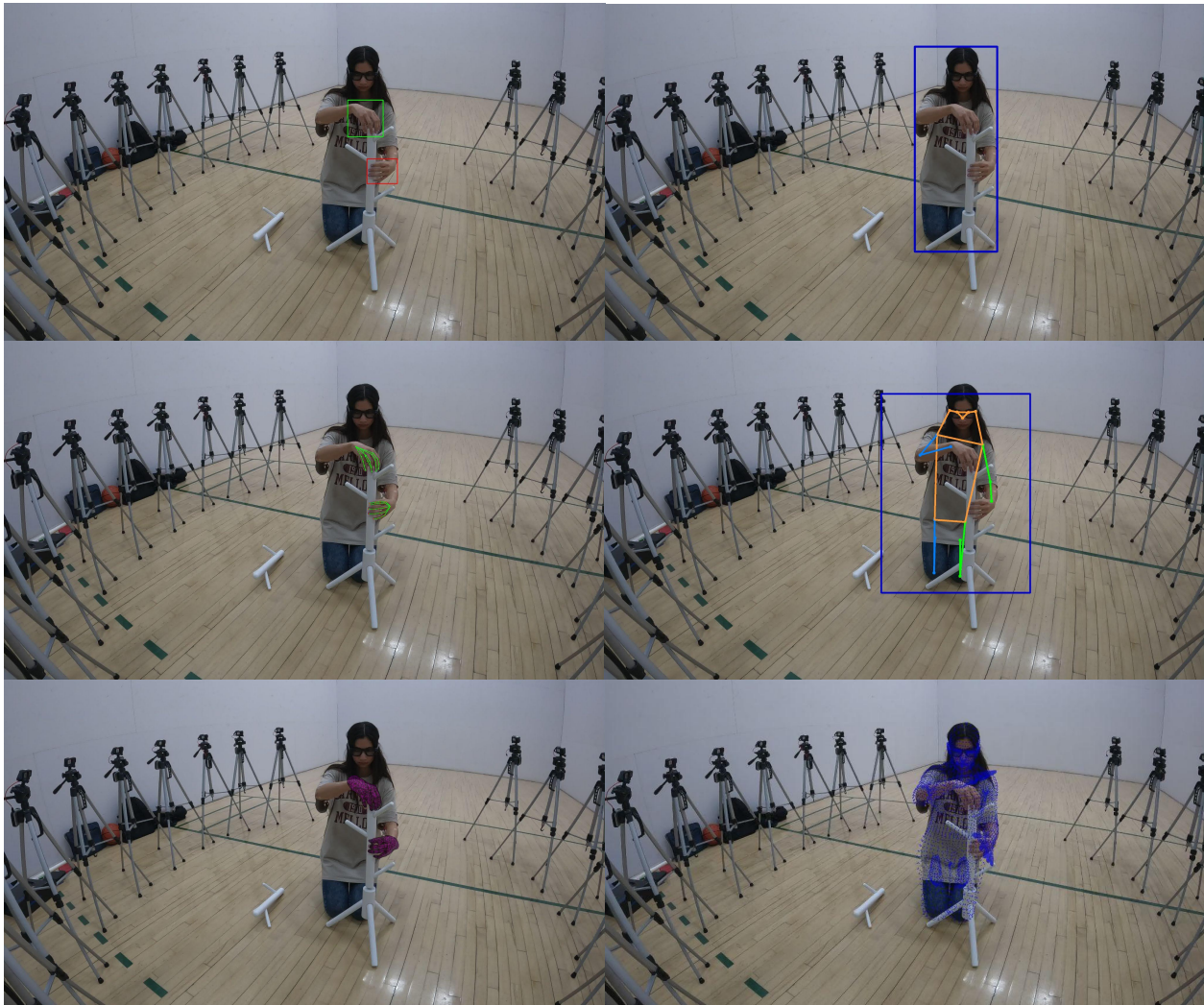


Figure 2. **Contact4D Annotations.** We provide bounding boxes, 2D/3D keypoints, and SMPL/MANO parameters for the human body and hands. Note that SMPL does not include finger annotations. Contact4D also provides ground-truth contact signals. Please refer to the accompanying videos in the supplementary materials.

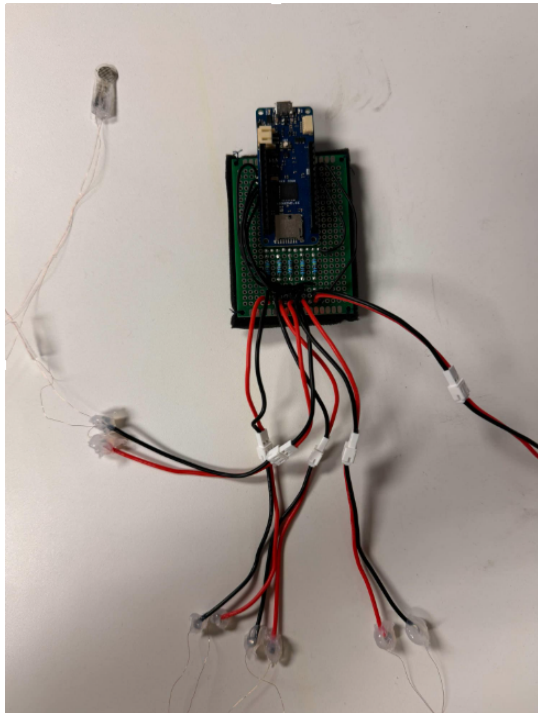


Figure 3. Our contact sensor is composed of an arduino and 5 FSRs (Each for one finger). The FSR is connected to the proto-board using fully insulated enameled wires, ensuring electrical shorts are avoided. Small amounts of silicon adhesive are applied on pins to further protect the sensors and stabilize the signals.



Figure 4. The Arduinos are secured on the individual's upper arms using loop straps, and sponges are affixed beneath each Arduino to serve as a medium. The sensor will be covered by the sleeves and remain mostly invisible during the capture.

References

- [1] Sally A Applin and Catherine Flick. Facebook’s project aria indicates problems for responsible innovation when broadly deploying ar and other pervasive technology in the commons. *Journal of Responsible Technology*, 5:100010, 2021. 1
- [2] Yusuf Abdullahi Badamasi. The working principle of an arduino. In *2014 11th international conference on electronics, computer and computation (ICECCO)*, pages 1–4. IEEE, 2014. 1
- [3] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. On-manifold preintegration for real-time visual–inertial odometry. *IEEE Transactions on Robotics*, 33(1):1–21, 2016. 1
- [4] Sheng Jin, Lumin Xu, Jin Xu, Can Wang, Wentao Liu, Chen Qian, Wanli Ouyang, and Ping Luo. Whole-body human pose estimation in the wild. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1
- [5] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 851–866. 2023. 1
- [6] Gregory L Polyzois, Gudbrand Øilo, and Jon E Dahl. Tensile bond strength of maxillofacial adhesives. *The Journal of Prosthetic Dentistry*, 69(4):374–377, 1993. 1
- [7] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), 2017. 1
- [8] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 1
- [9] Tao Xu, Wei Shen, Xiaoshan Lin, and Yi Min Xie. Mechanical properties of additively manufactured thermoplastic polyurethane (tpu) material affected by various processing parameters. *Polymers*, 12(12):3010, 2020. 1
- [10] SI Yaniger. Force sensing resistors: A review of the technology. *Electro International*, 1991, pages 666–668, 1991. 1