

Interacting with Virtual Environment through Hand Pose Estimation

Team 3

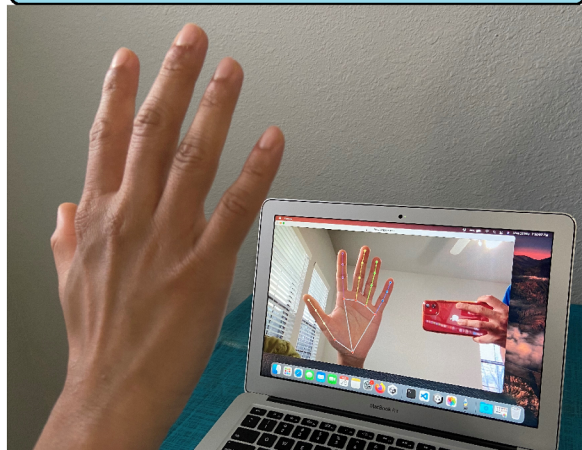
Rishik Dev

Ritwik Dev

Overview

Backend:

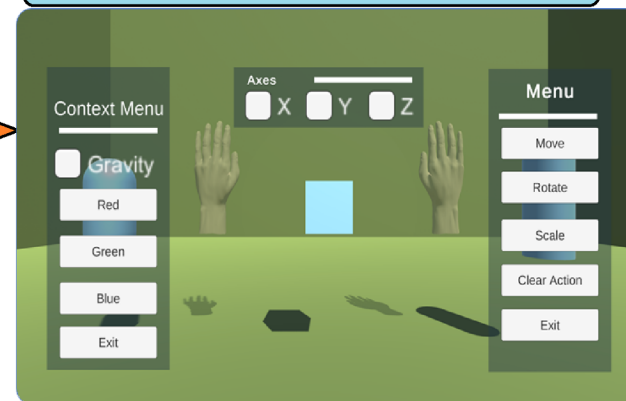
- MediaPipe for hand pose estimation
- OpenCV for MediaPipe
- Detecting the user's hand pose and gestures



ZeroMQ

Frontend:

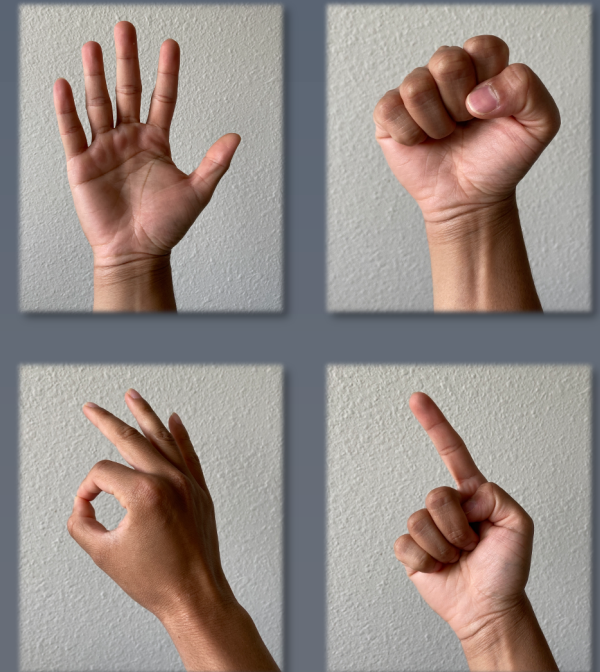
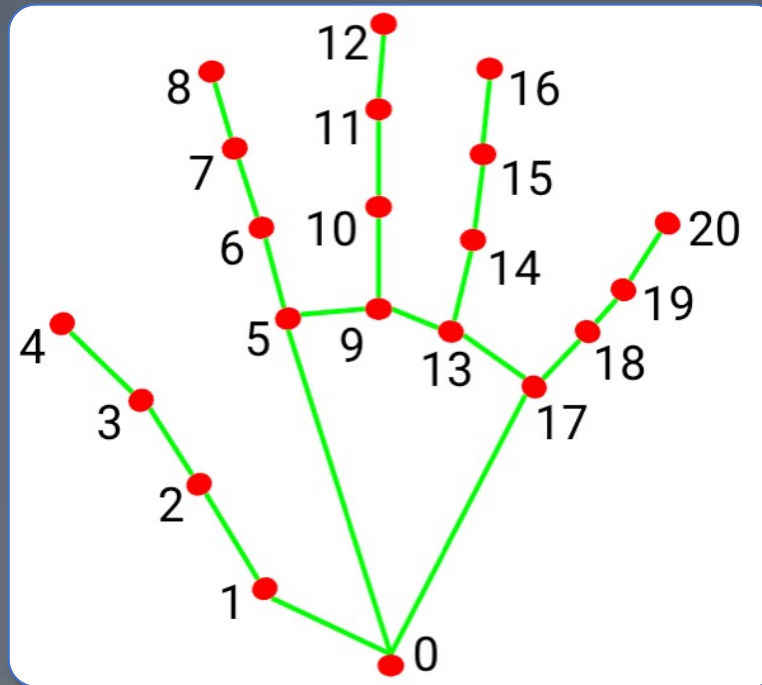
- Unity Game Engine for building the virtual environment and interactable objects
- Virtual hands to show the position and orientation of the user's hands



Multitier Architecture

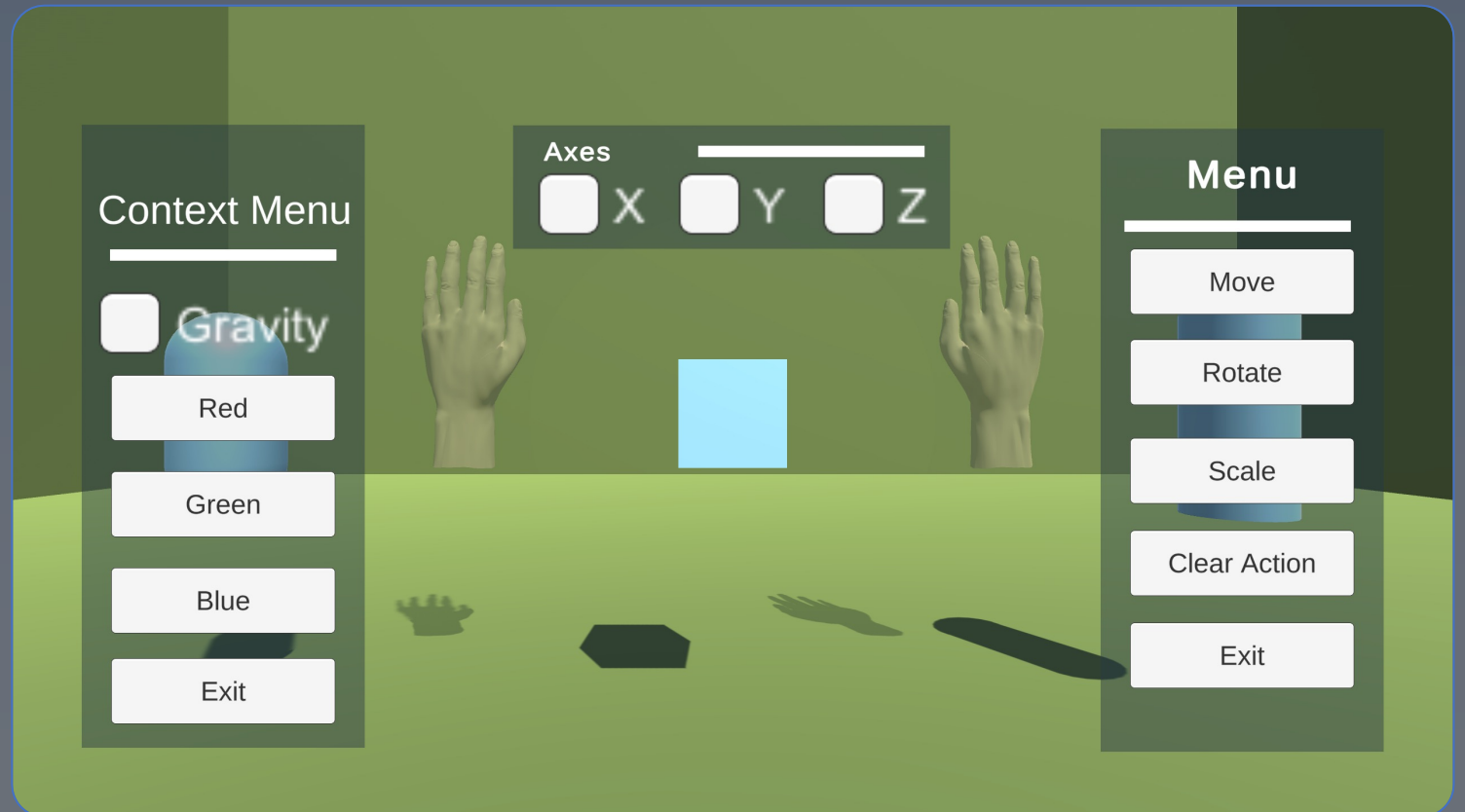
Backend

- Python script
 - MediaPipe Hands
 - OpenCV
- Tracks 2 hands
- Achieves 30+ FPS on CPU
- MediaPipe Hands solution tracks 21 3D-landmarks on each hand
- Wrist is the origin
- Landmarks normalized between 0 and 1
- Detects hand pose using the relative distances among the landmarks
- Detects 4 poses:
 - Open hand, Fist, OK, Pointing



Frontend

- Unity
- 3 objects to interact with
- Interactions:
 - Move, Rotate, Scale
 - Other properties
- Shows two hands that correspond to the user's hands
- Converts normalized landmarks to screen coordinates (pixels) – independent of the screen resolution



ZeroMQ

- High performance networking library
- Connects the backend and the frontend
- Publish – Subscribe model
 - Backend publishes the data using ZeroMQ
 - Frontend subscribes to the topic using NetMQ – a port of ZMQ for C#

More Information

For more information on how the project works, and how to run it, visit the [GitHub](#) page or click this link:

<https://github.com/rishikdev/The-University-of-Texas-at-Dallas/tree/main/CS%206384%20-%20Computer%20Vision/Project>