Human Pose Detection in 3D with Deep Learning and RGB-D Cameras

Participating researchers: Anthony Tzes, Merritt Moore
Vinicius Goncalves, Halil Utku, Nikolaos Giakoumidis, Aris Panousopoulos

Enabling Human-Robot Interaction in Dancing

- **Overview:** Exploring the integration of MediaPipe and Intel RealSense for advanced 3D human pose detection. Focused application in human-robot interaction for dancing.
- **MediaPipe Framework:** https://github.com/google/mediapipe Utilizes machine learning to facilitate real-time hand and pose tracking. Essential for interpreting complex human movements in 3D space.
- **Intel RealSense Technology:** Provides depth sensing and 3D vision capabilities. Enhances accuracy in detecting and mapping human movements.
- **Application in Dancing:** Enables robots to perceive and mimic human dance movements, creating interactive and responsive dance partners.

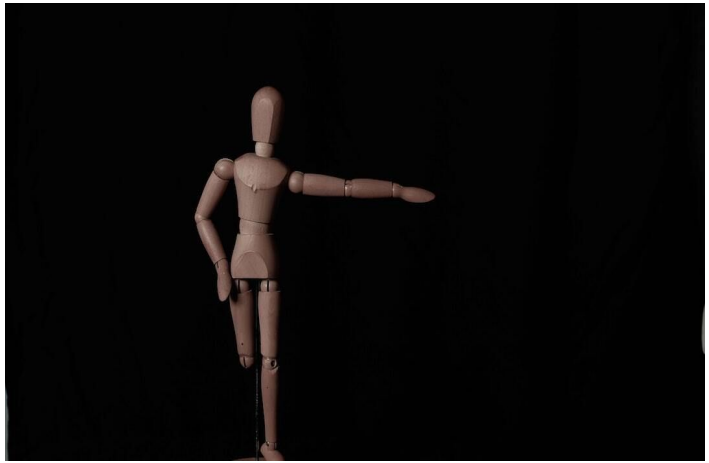


Photo by James Coleman on Unsplash

Technical Details of MediaPipe and Intel RealSense Integration

Enhancing 3D Human Pose Detection

- **MediaPipe Mechanics:** Employs a complex pipeline of machine learning models for pose estimation. Processes visual data from multiple sources, including Intel RealSense cameras.
- **Intel RealSense's Role:** Offers depth data and infrared imaging. Crucial for enhancing the 3D spatial perception capabilities of MediaPipe.
- **Real-Time Processing:** Integration allows for real-time interpretation of complex dance movements. Key for dynamic and responsive human-robot dancing interaction.
- **Accuracy and Precision:** Combination ensures high accuracy and precision in pose detection. Essential for nuanced dance movements and synchronization with a robotic partner.

Code Implementation and Live Demo

Human-Robot Dance Interaction

- **Code Overview:** Python script integrates MediaPipe and Intel RealSense for 3D pose detection. Configures streams, initiates MediaPipe Pose, and uses RealSense depth data.
- **Real-Time Landmark Detection:** Script captures and processes video frames in real-time. Extracts pose landmarks and overlays them on RGB images for visual verification.
- **3D Coordinates Extraction:** Combines color and depth information to calculate precise 3D coordinates for each landmark. Displays these coordinates on the live feed.
- **Demo Application:** Demonstrates the practical use of the script with a live dance interaction scenario, showcasing human-robot synchronization.

Challenges in Choreography and Robot-Human Synchronization

Innovating Dance Partnerships

- **Choreography Design:** Creating dance sequences that can be interpreted and performed by robots in sync with humans. Requires understanding of motion dynamics and robot capabilities.
- **Synchronization Issues:** Addressing latency and precision in robot movements to match human dancers. Essential for seamless and coordinated dance performances.
- **Utilizing UNITREE Robots:** Leveraging the capabilities of UNITREE robots (https://dev-z1.unitree.com/ and https://github.com/unitreerobotics/unitree_ros/tree/master/robots) known for their agility and human-like movement patterns.
- **Real-World Application:** Applying these technologies in live performances showcases the potential of advanced robotics and AI in entertainment and beyond.

Choreography Innovation for Limited Spaces

Maximizing Visibility for Pose Detection

- **Space-Conscious Choreography:** Designing dance moves that fit within confined areas while ensuring the dancer's full body is in the camera's view for accurate pose detection.
- **Creative Movement Solutions:** Utilizing vertical space and layered positioning to maximize the use of limited floor space. Incorporating poses that maintain visibility of all limbs.
- **Visibility and Pose Detection:** Choreography must facilitate unobstructed visibility of key movement points to ensure accurate pose detection and analysis by the system.
- **Interaction Dynamics:** Crafting choreography that allows for interactive elements with the robot, considering its range of motion and response time.

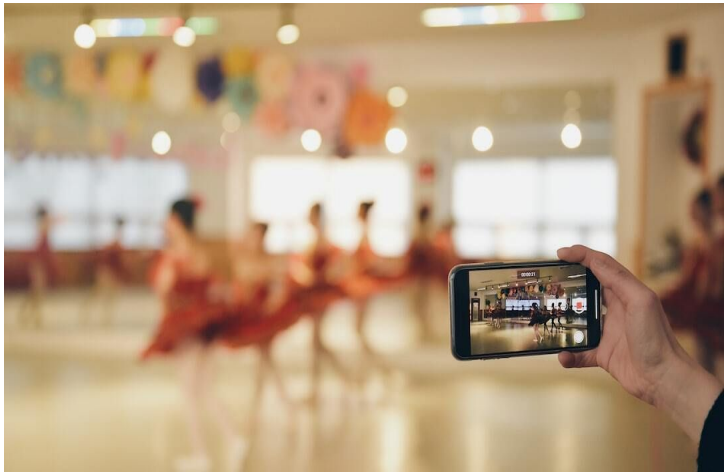
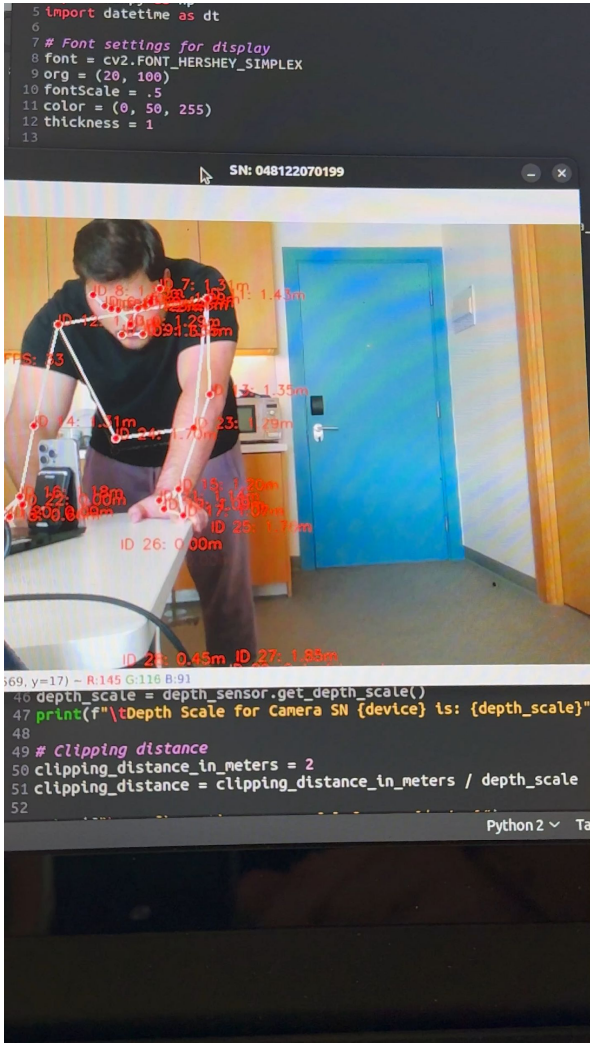


Photo by Suhyeon Choi on Unsplash

3D Human Pose Demo

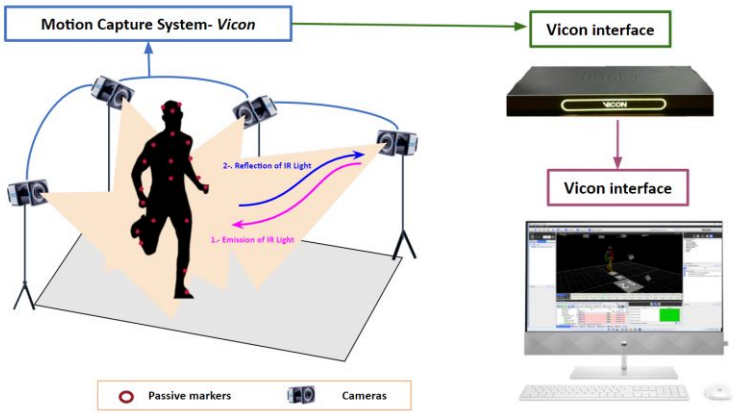
Use https://www.intelrealsense.com/depth-camera-d455/
RGB-D camera

Features	**Use environment:** Indoor/Outdoor	**Ideal Range:** .6 m to 6 m
	Image sensor technology: Global Shutter	**Inertial measurement unit:** Bosch BMI055
Depth	**Depth technology:** Stereoscopic	**Depth Field of View (FOV):** 87° × 58°
	Minimum Depth Distance (Min-Z) at Max Resolution: ~52 cm	**Depth output resolution:** Up to 1280 × 720
	Depth Accuracy: <2% at 4 m[1]	**Depth frame rate:** Up to 90 fps
RGB	**RGB frame resolution:** Up to 1280 × 800	**RGB sensor FOV (H × V):** 90 × 65°
	RGB frame rate: 30 fps	**RGB sensor resolution:** 1 MP
	RGB sensor technology: Global Shutter	

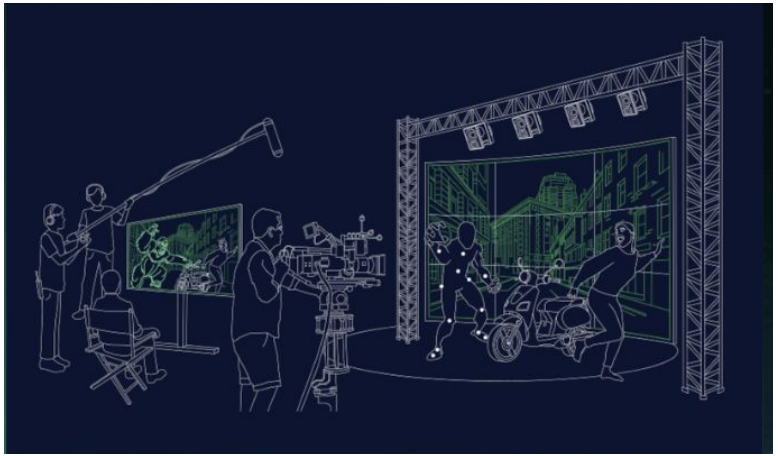


3D Human Pose Alternatives

Use Nexus for Vicon Motion Capture System
https://www.vicon.com/software/nexus/



Use Shogun for Vicon Motion Capture System https://www.vicon.com/software/shogun/



Objective

1. Submit an article to ArtsIT Conference https://artsit.eai-conferences.org/2024/ [Deadline June 15]
2. Present this article in November 2024

Next Steps

1. Decide on RGB-D or use of Motion Capture System (today)
2. Record dance by human-leader (by April 9)
3. Record 3D-pose of human-leader (by April 11)
4. Suggest robot motion based on Haussdorf distance metric (by April 15)
5. Robot Choreography, others, music…
6. Multiple cameras