UNITY-BASED REAL-TIME 3D HUMAN POSE ESTIMATION AND SHAPE MODELING

Stephanie Fortiz De Ita, Sadia Mubashshira, Ayda Eghbalian, Alejandro González and Kevin Desai.
The University of Texas at San Antonio, San Antonio TX.
Instituto Tecnológico y de Estudios Superiores de Monterrey, Master in Applied Artificial Intelligence.

ABSTRACT

This project aims to integrate pose data generated by Google MediaPipe Pose Estimation binding with a SMPL humanoid model in Unity. This multi-threaded full-body tracking system operates entirely in Python, sending 3D pose results to Unity for shape modeling and visualization. The model, webcam input, and game run on separate threads.

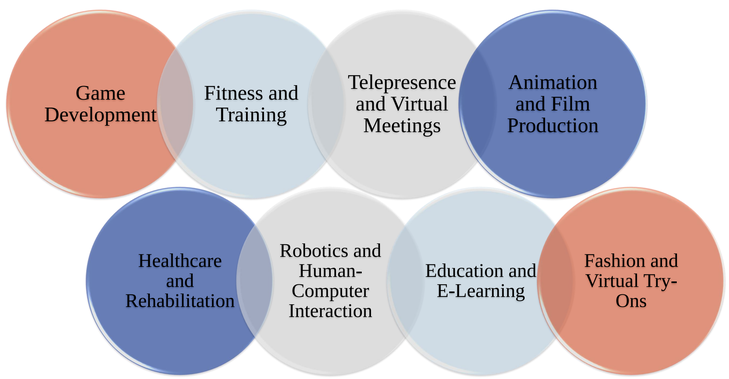
Fig. 1. Example of landmarks in MediaPipe (left) and SMPL (right).

INTRODUCTION

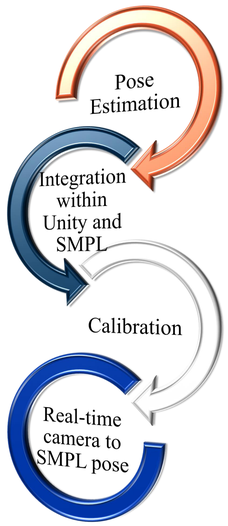
MediaPipe Pose detects key human body locations in images or videos using machine learning models in both image and 3D world coordinates.

SMPL is a realistic 3D human body model created from thousands of 3D body scans using skinning and blend shapes.

Using MediaPipe in Unity to move SMPL models can have several applications.



METHODS



Python allows the use of MediaPipe to capture the key body points and send them to Unity.

Unity stores these points and the SMPL model.

Calibration needs to be done so the MediaPipe key body points can be assigned to the SMPL model articulation points.

The joints adjust to each new position detected by the pose estimator.

RESULTS

Results on 3D pose estimation skeleton. The picture on the left shows the data captured by MediaPipe and Python, while the one on the right is the first attempt of a model in unity.

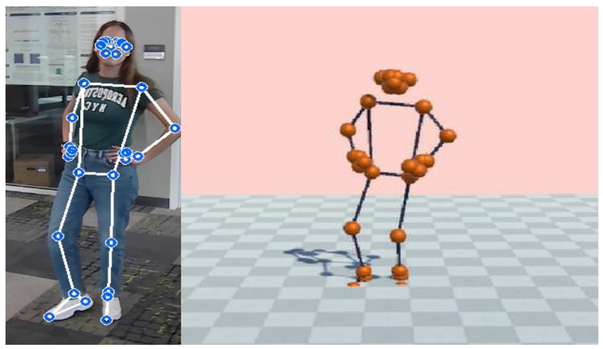


Fig. 2. Results of MediaPipe landmarks (left) and Unity landmarks (right).

Once this attempt was successful, it was possible to work on a calibration and binding with the SMPL model ahead. Sixty frames per seconds were achieved in this project.

Fig. 3. Results of MediaPipe landmarks (left) and SMPL model working with the received landmarks from MediaPipe (right).

CONCLUSIONS

- This project exemplifies the powerful synergy between advanced machine learning techniques and 3D modeling, paving the way for innovative solutions.
- The multi-threaded approach ensures efficient performance, enabling real-time applications that are both dynamic and responsive.
- The robust pose estimation provided by MediaPipe, can create highly immersive and interactive experiences.
- Working on the accuracy of details can provide even better results for applications.

FUTURE WORK

- The integration of SMPL-X models instead of SMPL can provide higher accuracy levels in the animation.
- Integration of various 3D pose estimation models.
- Multiple body estimation and modeling.

REFERENCES

Ganeshsar. (n.d.). UnityPythonMediaPipeAvatar: Creating a multi-threaded full body tracking solution supporting arbitrary humanoid avatars for Unity using Google Mediapipe Pose Python bindings. GitHub: https://github.com/ganeshsar/UnityPythonMediaPipeAvatar?tab=readme-ov-file
Google AI for developers. (n.d.). Google for Developers. https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=es-419
Williams, J. (n.d.). SMPL. https://smpl.is.tue.mpg.de/
FIGURE 10: An example of a SMPL-X [23] mesh regression approach based. (n.d.). ResearchGate. https://www.researchgate.net/figure/An-example-of-a-SMPL-X-23-mesh-regression-approach-based-on-2D-keypoint-estimation_fig4_351207865

The University of Texas at San Antonio
UTSA School of Data Science

Tecnológico de Monterrey