Optimizing Digital Human Movement: Waist-Centric Rotation and 3D Pose Estimation

Myeongsoep Kim†
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, Korea
myeongseopkim@keti.re.kr

Kyung-Taek Lee
Contents Convergence Research Center
Korea Electronics Technology Institute
Seoul, Korea
ktechlee@keti.re.kr

***Abstract*— Synchronizing real-time human movements with the digital world is crucial for enhancing realism and interaction with humans in virtual environments. However, measuring invisible joint rotations through a mono camera poses significant challenges, making it difficult to model the natural movements of a digital human. To address this issue, we calculated the waist rotation using the bilateral hip coordinates extracted from the 3D Pose estimation and applied it to the digital human. This approach is based on the fact that the waist forms the central axis of the human body and plays a pivotal role in various activities such as walking, running, and sitting. Using the waist as a reference point, we rotate the digital human and model the movement of the remaining joints using inverse kinematics. Through this method, we have implemented a digital human with more natural and realistic movements.**

Keywords—digital human, virtual reality, 3D pose estimation, human generation, digital human interaction

I. Introduction

With the rapid progression of virtual reality technologies, there is a burgeoning interest in research initiatives focused on recreating real-world human representations within digital twin platforms. To effectively synchronize human movements from real world to digital, a digital human model, complete with a skeletal structure, is indispensable. Recent academic discourse highlights the Skinned Multi-Person Linear (SMPL) model as a premier choice for rendering realistic human postures[1]. This model is extensively employed for 3D pose estimation, drawing from both static images and dynamic videos[2]. The inherent standardized skeletal structure of the SMPL model facilitates the acquisition of real-time skeletal data from users through a Pose detector. This, in turn, allows for the seamless synchronization of digital human movements with real-world actions, leveraging techniques such as Forward or Inverse Kinematics.

In their pursuit of accurate human action replication in digital models, the Simplify-x research group improved the alignment between the digitally recreated model and its real-world counterpart. They achieved this by computing a 3D model that encapsulates body pose, hand gestures, and facial expressions from a singular monocular image[3]. Concurrently, the MEVA (Variational Autoencoder) research collective introduced a methodology to derive refined 3D human postures and movements from RGB video sequences[4]. These advancements underscore the ongoing efforts in the academic community to emulate real human movements in digital realms. Particularly in studies that prioritize skeletal data over mere visual content for digital human reconstructions, there's a pressing need for advanced motion correction methodologies. The SMPL model, while robust, faces challenges when relying solely on 3D coordinates from the Pose detector for movement synchronization. Addressing this limitation, our research has harnessed bilateral hip coordinates from 3D pose data to determine the waist's rotation angle, refining the digital human model for enhanced, lifelike movements.

II. Optimizing Digital Huamn Movement

Fig 1 illustrates the 3D skeleton extracted from real-time video, its synchronization with the skeleton-based SMPL, and the digital human model incorporating our proposed waist-centric rotation. As demonstrated in the figure, our approach results in digital human movements that closely resemble the original image, surpassing the fidelity achieved with the 3D Skeleton alone.

3D pose estimation | SMPL model based on 3D pose estimation | SMPL model based on wrist-centric rotation and 3D pose estimation

This work was supported by the Technology Innovation Program (20018295, Meta-human: a virtual cooperation platform for a specialized industrial services) funded By the Ministry of Trade, Industry Energy (MOTIE, Korea).

A. *Apparatus and Environments*

For this study, we employed the Logitech C930 camera to capture input images essential for the 3D pose estimation model. Our system operates within an environment powered by the Unity 3D engine, tailored for digital human implementation. The entire system was built on Unity version 2021.2.3f1.

B. *Waist-Centric Rotation and 3D Pose Estimation*

With the advancements in computer vision technology, markerless gait analysis based on video footage has become feasible using pose estimation models such as OpenPose [5], BlazePose [6], and MMPose [7]. These pose estimation techniques employ computer vision and machine learning algorithms to extract human poses from videos and track the movement of joints and limbs in 3D space. Notably, Mediapipe Pose, an open-source tool developed by Google based on the BlazePose model, offers faster inference speeds and reliable results compared to other models [8]. Consequently, we utilize Mediapipe Pose to extract skeletons in real-time and rig each joint on the SMPL human model using inverse kinematics.

When movements are synchronized to the SMPL model using three-dimensional joints, the digital human's body shape can become misaligned with its real-world counterpart, as illustrated in the center image of Fig 1. This discrepancy arises because the data extracted from the pose detector is solely composed of 3D coordinates, neglecting the avatar's orientation. Consequently, when the model is not viewed directly from the front, its shape appears distorted. To rectify this, we factor in the anterior side of the torso when adjusting the SMPL model's movements, aiming for a more accurate digital human motion representation. Our approach calculates a perpendicular vector to the XZ plane using the left and right hip coordinates, which helps determine the direction corresponding to the front of the human body. The center point of the body is then adjusted based on this vector. As a result, the body's movement isn't just a function of the 3D coordinates of each joint; it also considers the orientation of the torso, offering a straightforward way to enhance digital human movement within the SMPL model framework.

III. RESULTS AND DISCUSSION

The primary objective of our research was to enhance the realism and accuracy of digital human movements in virtual environments. Our approach, as detailed in the preceding sections, revolved around the utilization of the waist as a pivotal reference point for movement synchronization. This section delves into the results obtained from our methodology and discusses its implications.

Upon implementing our waist-centric rotation approach, we observed a marked improvement in the alignment of the digital human's body shape with its real-world counterpart. The misalignment issues, which were previously evident when using only three-dimensional joints for synchronization, were substantially mitigated. The digital human movements, as a result, appeared more natural and closely resembled real-world human movements.

A significant challenge we encountered was the distortion of the SMPL model's shape when not viewed directly from the front. Our solution, which involved considering the anterior side of the torso and adjusting the center point of the body based on a calculated vector, effectively addressed this issue. This enhancement ensured that the digital human's movement wasn't solely dictated by the 3D coordinates of each joint but also factored in the orientation of the torso. The resultant digital human movement was more fluid, realistic, and in harmony with the real-world human movement.

The results of our research underscore the potential of waist-centric rotation in enhancing the realism of digital human movements. By addressing the inherent limitations of existing pose estimation models and introducing an innovative approach to movement synchronization, we have paved the way for more immersive and interactive virtual environments. Future research could explore further refinements to our methodology, potentially integrating more advanced pose estimation models or exploring other pivotal reference points for movement synchronization.

IV. CONCLUSION

In summary, our research has made significant strides in addressing the challenges of synchronizing real-time human movements with digital avatars. By focusing on the waist as a central axis of the human body, we have developed a methodology that not only enhances the realism of digital human movements but also mitigates the limitations inherent in existing 3D pose estimation models. The waist-centric rotation approach has proven effective in aligning the digital human's body shape more closely with its real-world counterpart, thereby achieving a more natural and realistic digital human interaction in virtual environments.

The implications of our work extend beyond the academic sphere, offering practical solutions for industries that rely on virtual reality technologies. Our findings pave the way for more immersive and interactive virtual experiences, opening up new avenues for future research. This could include the integration of more advanced pose estimation models or the exploration of other pivotal reference points for movement synchronization. Overall, our research contributes to the ongoing efforts to bridge the gap between the real and digital worlds, enhancing the quality of human interaction in virtual spaces.

ACKNOWLEDGMENT

This work was supported by the Technology Innovation Program (20018295, Meta-human: a virtual cooperation platform for a specialized industrial services) funded By the Ministry of Trade, Industry Energy (MOTIE, Korea).

REFERENCES

[1] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, ”SMPL: A skinned multi-person linear model,” ACM transactions on graphics (TOG), 34(6), 1-16, 2015.

[2] Y. Sun, Q. Bao, W. Liu, Y. Fu, M. J. Black, and T. Mei, ”Monocular, one-stage, regression of multiple 3d people,” In Proceedings of the IEEE/CVF international conference on computer vision, pp. 11179- 11188, 2021.

[3] Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10975-10985).

[4] Luo, Z., Golestaneh, S. A., & Kitani, K. M. (2020). 3d human motion estimation via motion compression and refinement. In Proceedings of the Asian Conference on Computer Vision.

[5] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, ”Realtime multi-person 2d pose estimation using part affinity fields,” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7291-7299, 2017.

[6] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, ”Blazepose: On-device real-time body pose tracking”, arXiv preprint arXiv:2006.10204, 2020.

[7] Sengupta, A., Jin, F., Zhang, R., & Cao, S. (2020). mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs. IEEE Sensors Journal, 20(17), 10032-10044.

[8] C. S. T. Hii, K. B. Gan, N. Zainal, N. M. Ibrahim, S. A. M. Rani, and N. Abd Shattar, ”Marker free gait analysis using pose estimation model,” In 2022 IEEE 20th Student Conference on Research and Development (SCOReD), pp. 109-113, IEEE, November 2022.