Ground-Relative Positioning in 3D Pose Estimation: A Novel Approach for Real-World Alignment of Skeleton Data

Myeongseop Kim[1*], Taehyeon Kim[1], Jean Oh[2], Kyuin Lee[3], and Kyung-Taek Lee[1†]

Abstract— **This paper presents a novel Dynamic Ground Alignment (DGA) algorithm, revolutionizing 3D pose estimation for enhanced ground-relative positioning in both real-world and digital twin environments. Our approach, integrating with the Mediapipe Pose framework, focuses on a body-centered coordinate system and employs 3D pose landmarks to accurately align digital characters with the ground plane in virtual settings. This method addresses the common issue of unrealistic character grounding, often found in traditional pose estimation techniques, by dynamically recalibrating the character's Y-position to maintain realistic grounding and interaction within the virtual environment. The application of DGA in various fields, including animation, virtual reality, sports science, physical therapy, and ergonomic studies, significantly improves the fidelity of digital avatars and models, offering a more integrated and realistic interpretation of human movement. While our study primarily demonstrates the efficacy of DGA through visual comparisons, it lays the groundwork for future research in quantitative validation and broader applications in human-computer interaction and digital human modeling.**

I. INTRODUCTION

The evolution of 3D pose estimation has been pivotal in diverse domains, including animation, sports science, and physical therapy. This technology is crucial for interpreting and replicating human movement in digital formats, offering vast potential across various applications [1]. Central to many 3D pose estimation methodologies is the assumption that the coordinate system is centered at the body's core. While this approach is standard, it poses significant challenges in accurately representing a subject's ground-relative position. Traditional pose estimation methods often center the coordinate system around the body or the camera, leading to scenarios where figures appear floating or misaligned with the actual ground level. This oversight compromises the realism and practical applicability of these models in real-world scenarios.

Simultaneously, the integration of 3D pose estimation in digital twin environments has revolutionized the control and animation of digital human avatars. This integration is vital for creating seamless and lifelike digital replicas of real-world actions [3]. Digital twins, replicating physical entities or environments virtually, heavily depend on the precision and realism of pose estimation, especially when adopting a body-centric coordinate system [2]. This approach ensures authentic interactions within virtual spaces. However, the accurate reflection of human posture, in relation to the ground plane in digital twin environments, adds complexity. This aspect is critical for ensuring the fidelity of digital human avatars, necessitating that their virtual interactions mimic real-world dynamics with high accuracy.

In response, this paper introduces a novel approach that enhances the alignment of 3D skeletal data with the ground plane, focusing on body-centered coordinate systems. Our method significantly improves the practical applicability of pose estimation in both real-world and digital twin scenarios. By accurately grounding both physical and virtual representations of human figures, our approach substantially enhances the fidelity of avatars in mirroring real-world interactions. This advancement contributes significantly to fields such as animation, virtual reality, sports science, physical therapy, and ergonomic studies, where precise pose replication is essential. We aim to bridge the gap between the digital and physical realms, offering an integrated and more realistic interpretation of human movement and posture, grounded in a body-centered coordinate system approach.

II. RELATED WORKS

The advancement of 3D pose estimation has been a key focus in computer vision, particularly in the context of replicating human movement in digital environments. Traditional methods have faced challenges in accurately placing avatars or models in relation to the ground plane, often resulting in unrealistic scenarios where figures appear floating or misaligned.

Tripathi, S. et al delves into enhancing pose estimation by incorporating intuitive physics. This study addresses the physical plausibility of poses, using inferred pressure heatmaps and the Center of Pressure (CoP) and Center of Mass (CoM) to improve ground contact realism. The methodology emphasizes the stability of poses in relation to the ground plane, a critical aspect often overlooked in previous works [3].

Parameswaran, V. and Chellappa, R. contributes to the field by addressing the challenge of achieving accurate pose estimations from uncalibrated camera views. This study highlights the impact of camera perspective on the interpretation

Dr. Kyung-Taek Lee is a corresponding author

[1]Myeongseop Kim, Kyung-Taek Lee, and Taehyeon Kim are with the Contents Convergence Research Center of Korea Electronics Technology Institute, 7500 AE Enschede, South Korea myeongseopkim@keti.re.kr, ktechlee@keti.re.kr, taehyeon.kim@keti.re.kr

[2]Jean Oh is with the Robotics Institute of Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA hyaejino@andrew.cmu.edu

[3]Kyuin Lee is with Computer Information Systems and Cybersecurity Department of Information Science Technology, University of Houston, 13850 University Blvd, COT, Room 314, Sugar Land, TX 77019, USA klee48@cougarnet.uh.edu

Input frame | Mediapipe Pose | Our approach

Fig. 1. DGA Algorithm applied to 3D pose estimation results

of human body postures and the importance of considering this perspective in pose estimation algorithms [4].

While these studies mark significant progress, there remains a gap in fully integrating environmental context and physical constraints into 3D pose estimation, particularly in digital twin environments. This paper aims to address these challenges by proposing a novel method that not only accounts for body-centric coordinates but also enhances the interaction of avatars with their environment, bridging the gap between digital and physical realms.

III. PROPOSED METHOD

A. Algorithmic Implementation of Ground Plane Alignment

Our approach, Dynamic Ground Alignment(DGA) for 3D Pose Estimation, enhances the accuracy of digital character alignment with the ground plane in virtual settings. With inputs comprising 3D pose landmarks and parameters for positional adjustments, the method begins by identifying the lowest Y-coordinate from the landmarks, establishing a baseline for realistic positioning. As the character moves, this baseline is dynamically updated, and the character's Y position is recalibrated to maintain this realism, as depicted in Fig 2, which illustrates the Dynamic Ground Alignment(DGA).

Our algorithm dynamically maintains the character's Y-position within realistic limits throughout its movement. When the calculated positional adjustment goes beyond the established ground limit, it is rectified to ensure proper grounding, adhering to the minYGroundLimit. This recalibration, highlighted in Algorithm 1, ensures authentic interaction with the virtual environment. Initially, the method identifies the lowest Y-coordinate from the pose landmarks, establishing a baseline, termed lowestLandmarkY. This baseline is updated continuously to keep the character's positioning grounded and realistic, thus enhancing the fidelity of interaction within the virtual space.

B. Integration of Mediapipe Pose with the Own Algorithm

In our research, we utilized the Mediapipe Pose framework to perform 3D pose estimation from RGB images [5]. This allowed us to extract 33 body keypoints, which were then visualized within the Unity 3D environment to observe the results. Mediapipe Pose inherently uses a coordinate system centered around the waist, providing a consistent reference point for our algorithm. We meticulously compared the visual representation of poses before and after applying our algorithm, using this waist-centered coordinate system as a baseline. Furthermore, the calculated skeleton information was rigged to the SMPL human model, commonly utilized in virtual environments, to represent the digital human accurately [6].

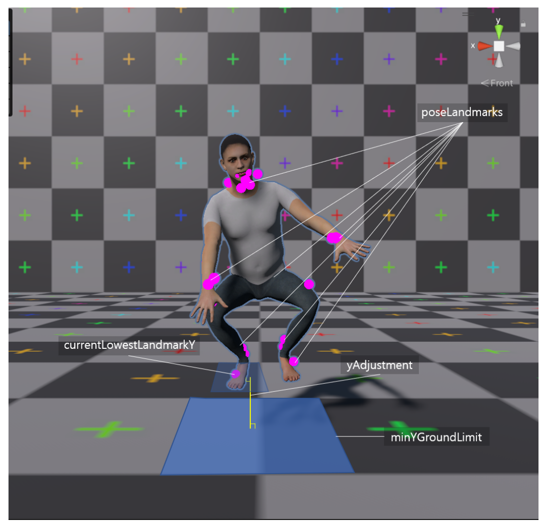
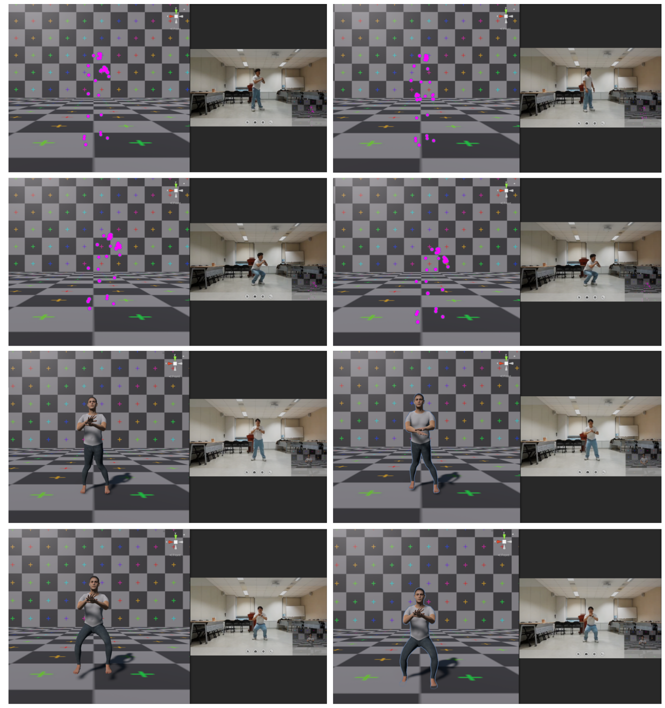


Fig. 2. Dynamic Ground Alignment for 3D Pose Estimation

Visual comparisons in Fig 1 and Fig 3 underscore the effectiveness of our method. These figures show the application of our algorithm to the pose estimation results, highlighting the enhanced ground alignment. Fig 3, in particular, presents a comparison that demonstrates the substantial improvements our method brings to the field of 3D pose estimation. While we could not confirm the performance of our algorithm with quantitative metrics, an intuitive comparison of visual

```
Algorithm 1: Dynamic Ground Alignment
Input: poseLandmarks, avatarTransform,
       desiredPosYOffset, minYGroundLimit
Output: Adjusted avatarTransform position
lowestLandmarkY ←
  GetLowestLandmarkY(poseLandmarks)
Function Main:
    avatarCurrentY ← avatarTransform.position.y
    avatarTargetX ← avatarTransform.localPosition.x
    avatarTargetZ ← avatarTransform.localPosition.z
    currentLowestLandmarkY ←
      GetLowestLandmarkY(poseLandmarks)
    yAdjustment ← max(-currentLowestLandmarkY
      + desiredPosYOffset, -minYGroundLimit)
    avatarTransform.position ←
      NewVector3(avatarTargetX, -yAdjustment,
      avatarTargetZ)
    lowestLandmarkY ← currentLowestLandmarkY
Function GetLowestLandmarkY(poseLandmarks):
    lowestY ← float.MinValue
    foreach landmark in poseLandmarks do
        if landmark.localPosition.y > lowestY then
            lowestY ← landmark.localPosition.y
    return lowestY
```



(a) Mediapipe Pose (b) Mediapipe Pose with DGA algorithm

Fig. 3. Comparison of 3D pose estimation based on Mediapipe Pose

representations before and after its application revealed a clear enhancement. This observation suggests our ground alignment technique effectively mitigates issues present in the existing models.

IV. CONCLUSION

In this paper, we introduced the Dynamic Ground Alignment (DGA) algorithm, marking a significant advancement in 3D pose estimation. This groundbreaking method seamlessly integrates with the Mediapipe Pose framework, accurately aligning digital characters with the ground plane in virtual environments. Focusing on a body-centered coordinate system and using 3D pose landmarks, the DGA method effectively addresses the long-standing challenge of ground-relative positioning, crucial for creating realistic digital avatars and models.

The dynamic recalibration of the character's Y-position within the algorithm ensures that digital characters maintain a realistic and grounded presence in virtual environments. This enhancement in visual fidelity and authenticity of interactions within the virtual space is evidenced through intuitive visual comparisons. These comparisons, especially within the Unity 3D environment using the waist-centered coordinate system and the SMPL human model, demonstrate substantial improvements over traditional pose estimation techniques.

Our approach extends its impact beyond creating digital avatars, having significant implications for fields such as animation, virtual reality, sports science, physical therapy, and ergonomic studies. The accurate replication of human posture and movement provided by our method bridges the gap between digital and physical realms, offering a more cohesive and realistic interpretation of human movement.

In conclusion, the Dynamic Ground Alignment algorithm is a pivotal step toward achieving more accurate and realistic 3D pose estimation. Although this study lacks quantitative validation, the visual improvements are clear and pave the way for future research. This method is a valuable tool for researchers and practitioners across multiple domains, setting the stage for more realistic and immersive virtual experiences. Future work should focus on refining this approach further and exploring its broader applications in areas like robotics and motion capture, maximizing its potential in enhancing human-computer interaction and high-fidelity digital human modeling.

ACKNOWLEDGMENT

This research 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] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, and L. Shao, "Deep 3D human pose estimation: A review," Computer Vision and Image Understanding, 210, 103225, 2021.

[2] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," Ieee Access, 8, 21980-22012, 2020.

[3] S. Tripathi, L. Müller, C. H. P. Huang, O. Taheri, M. J. Black, and D. Tzionas, "3D human pose estimation via intuitive physics," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4713-4725, 2023.

[4] V. Parameswaran and R. Chellappa, ”View independent human body pose estimation from a single perspective image,” in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, Vol. 2, pp. II-II, 2004.

[5] 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.

[6] 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.