UNIVERSITY of WASHINGTON

Human Pose Estimation Using Ultra-Wideband Radar Technology

Krish Jain, Jason Miller

Problem: Our project explores the fusion of ultra-wideband (UWB) radar and inertial measurement units (IMUs) to enable more accurate and immersive VR avatars. Current IMU-based tracking methods lack absolute position data, making full-body pose estimation challenging.

By integrating UWB's precise distance measurements with IMU motion data, we aim to create a low-cost, real-time motion capture system that enhances VR experiences. We will develop an embedded system, train a machine learning model to infer joint angles, and render a dynamic, responsive VR avatar using AFrame. This system has the potential to improve realism in VR gaming, social interactions, and virtual training, bringing more lifelike movement and precise tracking to immersive environments.

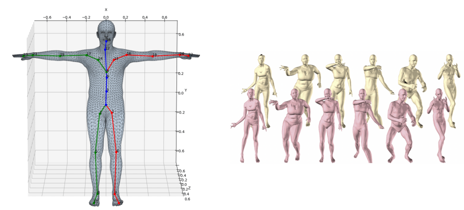

Figure 1: On the left is a representation of the 23 joint angles needed by SMPL and how they represent movement for an avatar, and the right image depicts the stylistic differences present from different poses output by the model.

Related Work

Current pose estimation efforts rely on a range of sensor mediums and modeling techniques to produce accurate 3D human representations.

- Professional systems such as Optitrack or Xsens deploy multiple cameras and full-body suits to capture motion, but these setups are costly and cumbersome for general users.
- Recent research focuses on sparse-sensor approaches, such as IMUPoser —which uses IMU sensors in phones and earbuds—and QuestSIM, which employs IMUs in a VR headset and controllers.
- Traditional inverse kinematics can fill in missing joint information, but it often leads to unrealistic movements

Figure 2: Depicts the Xsens suit, and the full body sensor needed for accurate pose estimation, shown from three different angles: front, back, and side view.

Method

- We introduced an approach combining UWB sensors with lower-body data to achieve full-body reconstruction. We generated a large synthetic dataset using AMASS, artificially producing IMU and UWB sensor readings around the ankles from motion capture data. Enabled us to train on over 20 hours of diverse movements.
- We employed a bidirectional LSTM to learn the mapping between the artificial sensor inputs and each sensor's local rotation. Leveraging the SMPL model, we transferred these learned rotations to a full-body representation.
- We used the CMU dataset, BMLMovi, ACCAD, BMLhandball, DFaust and DanceDB datasets from AMASS to synthesize our data.
- Our exact architecture was (linear, relu, bidirectional lstm layer, linear, dropout). We evaluated our loss following the method done in IMUPoser, also using the Adagrad optimizer.
- We recreated frames of SMPL using the Afram library, which converted the 3D PKL meshes to tangible VR objects.

Approach

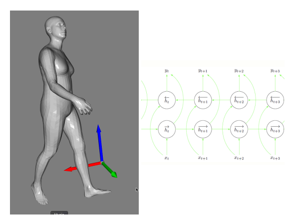

Figure 3: The right shows the LSTM architecture that we used, and the left depicts the result of the VR walking experience from an offline PKL file

Results

Our final test accuracy is 0.275, which aligns closely with the final validation loss of 0.25. These results demonstrate the potential for accurate full-body estimation using only lower-body sensors. Moreover, the training loss graph clearly shows that the model continues to learn meaningful features, with a steady decrease in loss over time. Finally, qualitative visualizations indicate that the predicted poses align closely with ground-truth poses, especially in sequences involving substantial lower-body motion (e.g., walking and jumping).

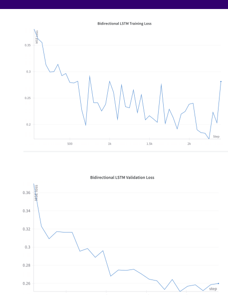

Figure 4: The top graph shows the training loss of the LSTM and then the bottom graph shows the validation loss, over 2k steps plotted against loss

Conclusion

Contributions: We introduce an innovative sensor fusion methodology that leverages both UWB radar and IMU data to overcome the limitations of traditional full-body tracking systems. Our approach utilizes the SMPL-X model and a bidirectional LSTM to accurately synthesize joint angles, demonstrating that high-fidelity motion capture is achievable with sparse sensors. Additionally, we lay the groundwork for future enhancements in real-time data streaming and adaptive noise correction, offering significant potential for advancing VR applications in gaming, social interaction, and rehabilitation.

Key Takeaway: The integration of the SMPL-X model with a bidirectional LSTM, trained on the AMASS dataset, enables the generation of realistic 3D avatars using minimal sensor input. This method not only reduces hardware complexity and cost but also enhances user immersion through natural and responsive avatar representations.

Future Work

Future work will focus on two main aspects:

1. Developing a real-time streaming pipeline for processing UWB and IMU data, which is essential for reducing latency and providing immediate feedback in interactive VR applications.

2. Comprehensive user studies to gather both qualitative and quantitative feedback on avatar realism, sensor placement, and user interface design, guiding iterative improvements to create a more engaging and intuitive VR experiences.

References

Andreas Aristidou, Joan Lasenby, Yiorgos Chrysanthou, and Ariel Shamir. 2018. Inverse kinematics techniques in computer graphics: A survey. *Computer Graphics Forum* 37 (2018), 35–58. https://doi.org/10.1111/cgf.13310

Rayan Armani, Changlin Qian, Jiaxi Jiang, and Christian Holz. 2024. Ultra Inertial Poser: Scalable Motion Capture and Tracking from Sparse Inertial Sensors and Ultra-Wideband Ranging. arXiv:2404.19541 [cs.CV]. https://arxiv.org/abs/2404.19541

Mohammad Cheraghinia, Awais Shahid, Samuel Luchie, G.-J. Gordeobeke, Olivier Caytan, Jens Fontaine, Bram Van Herbruggen, Sam Lemey, and Eli De Poorter. 2024. A comprehensive overview on UWB radar: Applications, standards, signal processing techniques, datasets, radio chips, trends, and future research directions. *IEEE Communications Surveys & Tutorials* (2024). https://doi.org/10.1109/comst.2024.3488173

Yujun Huang, Manuel Kaufmann, Emre Aksan, Michael J. Black, Otmar Hilliges, and Gerard Pons-Moll. 2018. Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics* 37, 6 (2018), 1–15. https://doi.org/10.1145/3272127.3275108

Matthew M. Loper, Naureen Mahmood, and Michael J. Black. 2014. MoSh: Motion and Shape Capture from Sparse Markers. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 33, 6 (Nov. 2014), 220:1–220:13. https://doi.org/10.1145/2661229.2661273

Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. 2019. AMASS: Archive of Motion Capture as Surface Shapes. In *International Conference on Computer Vision*, 5442–5451.

Vignesh Mollyn, Ryo Arakawa, Manas Goel, Chris Harrison, and Karan Ahuja. 2023. IMUPoser: Full-body pose estimation using IMUs in phones, watches, and earbuds. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, 1–12. https://doi.org/10.1145/3544548.3581392

NaturalPoint Inc. 2020. OptiTrack. Available at http://optitrack.com

Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. 2019. Expressive body capture: 3D hands, face, and body from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 10975–10985.

Qorvo. [n.d.]. DW1000: IEEE820.15.4-2011 UWB compliant wireless transceiver IC. Retrieved January 27, 2025, from https://www.qorvo.com/products/p/DW1000

Venkatesh Ramani, Hamed Khayami, Yang Bai, Nishit Garg, and Niloy Roy. 2024. IMUOptimize: A Data-Driven Approach to Optimal IMU Placement for Human Pose Estimation with Transformer Architecture. arXiv preprint arXiv:2402.08923 (2024). https://arxiv.org/abs/2402.08923

Alexander Winkler, Jungdam Won, and Yuting Ye. 2022. QuestSim: Human motion tracking from sparse sensors with simulated avatars. In *SIGGRAPH Asia 2022 Conference Papers*. ACM, 1–8. https://doi.org/10.1145/3550469.3551301

Xsens. 2020. Xsens Motion Capture. Available at https://www.xsens.com/motion-capture

Xinyu Yi, Yili Zhou, and Feng Xu. 2021. TransPose: Real-time 3D human translation and pose estimation with six inertial sensors. *ACM Transactions on Graphics* 40, 4 (2021), 86:1–86:13. https://doi.org/10.1145/3450626.3459786