IMU (Inertial Measurement Unit)-Based Real-Time Workout Monitoring Digital Twin Avatar System

S M Saiful Islam Badhon
Department of Information Science
Univeristy of North Texas
Denton, TX
smsaifulislambadhon@my.unt.edu

***Abstract*—Personal trainers, physical therapists, and fitness experts have historically found it difficult to track exercise technique in real time. The most popular approach, visual inspection by a human observer, is error-prone, subjective, and restricted to a single viewing angle. This is particularly troublesome in virtual contexts, when the therapist or trainer cannot physically view the user from every angle, and in gym settings, where one trainer may oversee numerous clients at once. There is no objective way for a trainer seeing a client remotely or standing across the room to measure joint-angle variations or consistently assess the quality of each repeat. By creating an IMU (Inertial Measurement Unit)-based real-time exercise monitoring system with a 3D digital twin avatar generated in Unity 3D, this project fills that gap. A Unity client receives motion data at 25 Hz via a UDP socket from ten IMU sensors that are affixed to key anatomical locations, such as the ankle, knee, hip, shoulder, and forearm. Without using pre-recorded animation keyframes, the Unity program uses the user's movements to animate a Mixamo humanoid avatar (Remy/Jody) in real time. Trainers can watch the avatar in any 360-degree view, freeze it, and replay it for the last repetition. A computer vision module that was constructed concurrently with the IMUs to compare using MediaPipe. This work is significant because it substitutes quantifiable, objective data for a very subjective process. Sports coaches examining athletic movement, researchers gathering biomechanical data, personal trainers in gyms, and physical therapists tracking healing in clinics or remotely all find it useful. Additionally, the system is scalable: as long as appropriate IMU location and threshold joint angles are established, the same architecture used to track squats may be expanded to any exercise.**

***Index Terms*—Digital Twin, IMU, Unity 3D**

I. Introduction

Precise assessment of human movement is essential for fitness training, physical therapy, and sports performance. Nevertheless, the vast majority of current methods primarily rely on human observation, which has clear disadvantages. Trainers and therapists often use a single viewpoint to evaluate movement and provide verbal commentary based on their views. This method is inherently subjective since different experts may assess the same movement in different ways. Even experienced observers may miss subtle but serious issues like mild joint misalignment, imbalance, or poor posture. These limitations become much more apparent in remote settings where trainers are forced to rely on video conferences that restrict viewing angles and lack depth information.

At the same time, there has been a noticeable increase in demand for telehealth physical therapy, remote fitness coaching, and at-home exercise programs. These techniques improve accessibility, but they make it more difficult for professionals to effectively oversee and control movement. In both in-person and remote settings, there is a lack of objective, quantitative data to evaluate performance and track progress over time. Conventional high-end solutions, such as optical motion capture systems, can provide detailed biomechanical data, but they are expensive, require particular settings, and are not practical for routine use in clinics, gyms, or homes.

Our project addresses these problems by developing a real-time motion tracking and visualization system that combines wearable sensing technology with an interactive 3D environment. By affixing multiple Inertial Measurement Unit (IMU) sensors to significant body parts, the system gathers full-body motion data. These sensors send data to a Unity 3D application continuously and steadily via a wireless connection. The application allows for real-time, fully three-dimensional visualization of the user's movements by mapping this data onto a humanoid avatar.

One of the key benefits of this approach is that it does away with the shortcomings of single-angle observation. Users can examine actions in great detail and view the animated avatar from any angle thanks to an interactive orbit camera. Additionally, the system provides quantitative feedback by calculating joint angles and displaying them within the interface.

The virtual environment is meant to resemble a typical home setting rather than a clinical laboratory (see Figure 1). This choice improves the system's usability and makes it more suitable for real-world applications, especially when remote training and rehabilitation are involved. The system supports common exercises like squats, biceps curls, and shoulder presses that are often used in both fitness and rehabilitation contexts.

This can help a number of user groups. Personal trainers can provide consistent, data-driven feedback and more accurate client tracking thanks to the technology. By employing precise metrics to track patient recovery, physical therapists can reduce their dependence on subjective assessment. Sports coaches have the ability to assess athletic performance and identify areas that require improvement. Researchers can also use the technology to collect biomechanical data outside of lab settings.

This project's main objective is to bridge the gap between

Fig. 1. **Unity-based home simulation environment for real-time motion analysis.** A living room scene from the Unity Asset Store is used to replicate practical home conditions for remote exercise monitoring.

basic visual observation and complex motion capture systems. By combining wearable sensors, real-time data processing, and interactive visualization, it provides a useful and effective movement monitoring solution. The system improves accuracy, reduces subjectivity, and allows for ongoing feedback, making it a valuable tool for modern fitness, rehabilitation, and research applications.

II. Related Work

This work builds on three main research areas: IMU-based human motion analysis, digital twin and virtual avatar systems, and automated exercise form detection. Each area has contributed important ideas that inform the design of the proposed system.

A. IMU-Based Human Motion Analysis

Human motion is frequently recorded in clinical and athletic settings using Inertial Measurement Units (IMUs). IMU-based rehabilitation systems have rapidly expanded, especially in the areas of gait and stroke analysis, according to a systematic study by Gu et al. [1]. However, issues like short battery life and sensor drift continue to be common. The importance of sensor location, density, and sampling rate for accurate motion capture is highlighted by recent studies. Phan et al. [2] show that increasing the number of sensors improves classification accuracy, but user comfort may be compromised. According to Sara et al. [3], contemporary IMU systems can estimate joint angles with errors of 3 to 8 degrees, which is a reasonable benchmark for acceptable precision. Real-time feedback technologies have also been the subject of several studies. These results support the notion that real-time, detailed body movement recording is possible with multi-sensor IMU systems.

B. Joint Kinematics and IMU Validity

Multiple studies have shown that IMU-based measurements are just as reliable as optical motion capture methods. Yin et al. [4] demonstrate that calibration is essential for enhancing accuracy, especially for hip and ankle joints. After being corrected correctly, IMU measurements match up well with optical systems. Cho et al. [5] further show that IMUs can accurately measure joint kinematics in a variety of clinical settings, indicating that they can be a useful replacement for camera-based systems. A meta-analysis conducted by Zeng et al. [6] validates the efficacy of IMUs in assessing spatiotemporal parameters, but caution is advised in the interpretation of joint-level kinematics. These results necessitate the incorporation of calibration and validation procedures in the proposed system to guarantee consistent and dependable joint angle estimation.

C. Digital Twin and Virtual Avatar Systems

Real-time tracking of human movement on virtual avatars is made possible by digital twin technology. A Unity-based system developed by Ha et al. [7] uses IMU sensors to move full-body avatars for exercise tracking and remote collaboration. Their research demonstrates that wearable sensors and real-time 3D visualization can coexist. Additional studies examine how users can connect and communicate with one another on various platforms. While Kocur et al. [8] investigate the influence of avatar representation on user behavior during exercise, Park et al. [9] stress the need to accommodate both VR and non-VR environments. IMU-driven digital twins have been applied to fitness applications as well as ergonomics and safety evaluations. Industrial applications can be facilitated by integrating IMUs with physics-based simulation, as shown by Weistroffer et al. [10]. Together, these pieces demonstrate how excellent digital avatars are for viewing motion in real time.

D. Summary

Overall, current research indicates that wearable-based joint analysis is effective, digital twin systems may be utilized for real-time visualization, and IMUs are useful for motion capture. However, the majority of the work completed thus far has focused on individual components rather than entire systems. The proposed system includes interpretable form analysis, real-time 3D avatar visualization, and multi-sensor IMU tracking. The system is highly beneficial for remote monitoring, fitness, and rehabilitation thanks to this integration and a virtual world that can be accessed from home.

III. Implementation

The proposed system is implemented through a three-phase pipeline: (i) sensor configuration and data acquisition, (ii) data transmission, and (iii) real-time avatar animation and analytics within Unity 3D. Each phase is designed to ensure low-latency, accurate, and interpretable motion tracking.

A. Sensor Configuration and Data Acquisition

Ten Inertial Measurement Unit (IMU) sensors, each with a three-axis accelerometer and three-axis gyroscope, are used in the system (see Figure 2). In order to record full-body motion pertinent to exercises like squats, biceps curls, and shoulder presses, these sensors are positioned on important body segments, such as the ankles, knees, hips, shoulders, and forearms.

Each sensor performs onboard sensor fusion to determine its orientation in quaternion form using techniques such as

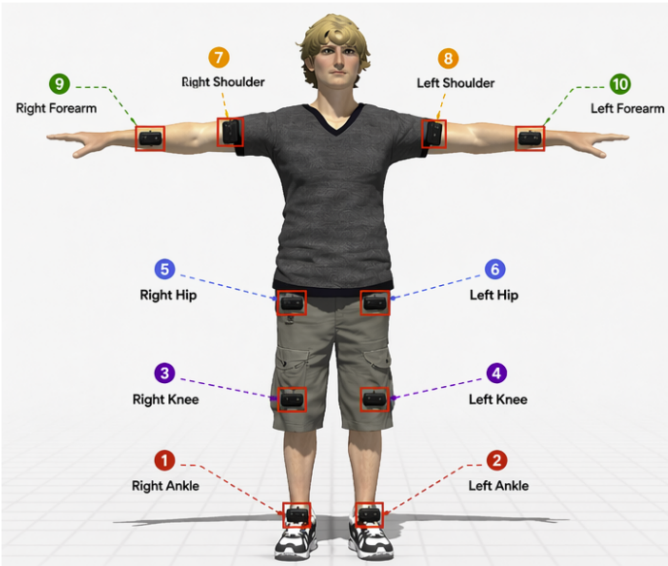

Fig. 2. **Placement of the ten IMU sensors on the human body.** The sensors are positioned on key body segments to capture joint-level motion for upper and lower extremities.

Madgwick or complementary filtering. The user establishes a reference frame by performing a T-pose calibration at system startup. By doing this, it is ensured that all joint angles are measured with respect to a neutral anatomical position.

The sensors are able to monitor controlled resistance activities at a rate of 25 Hz (one frame every 40 ms). Time resolution and system efficiency are well-balanced at this sample rate.

B. Data Transmission

Sensor data is wirelessly transmitted to a host PC via a local Wi-Fi network using UDP sockets (port 5005). Because of its low latency, which is crucial for real-time viewing, UDP was selected. Since reliable transmission systems like TCP may cause delays, it's acceptable in this situation if some packets are occasionally lost.

Each packet contains quaternion data from all ten sensors in a small-space format. By transmitting data at a steady rate of 25 Hz, the system maintains a continuous data stream. On the receiving end, a Unity-based client application processes incoming data using a C# UDP listener that runs in the background.

C. Avatar Animation and Real-Time Analytics

The main visualization platform is the Unity 3D engine. The user's motion is represented by a humanoid avatar (Remy and Jody from Mixamo) (see Figure 4). Because the avatar is set up with Unity's Humanoid rig, sensor data and skeletal joints can be directly mapped.

The corresponding bones of the avatar receive quaternion data from the sensors during each update cycle. This eliminates the need for pre-recorded animations and enables the avatar to mimic the user's movements in real time.

Video input from a camera is processed by a parallel computer vision pipeline that uses MediaPipe Pose and OpenCV

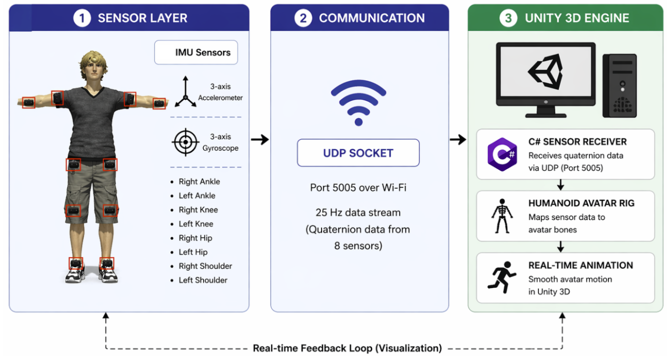

Fig. 3. **System architecture and data flow of the proposed IMU-based motion tracking system.** The sensor layer captures motion using IMU sensors placed on key body segments, and transmits quaternion data via Wi-Fi using a UDP protocol. The Unity 3D engine receives and processes this data through a C-based receiver, maps it to a humanoid avatar rig, and enables real-time animation and visualization.

Fig. 4. **Humanoid avatar options used for motion visualization.** The system supports multiple avatar types, allowing users to switch between male and female characters for personalized representation.

to increase reliability. This module independently calculates joint angles and estimates body posture. The outcomes are contrasted with values derived from the IMU.

D. System Architecture

The overall system integrates hardware sensing, wireless communication, and real-time visualization into a unified framework (see Figure 3). The software and hardware components are summarized in Table I.

Overall, the implementation provides a low-latency, interpretable, and scalable solution for real-time motion tracking and analysis.

IV. Functionality

This section describes the main functional components of the proposed system, supported by visual examples from the implementation.

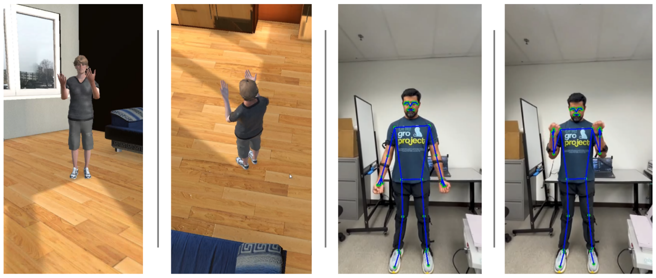

Fig. 5. **Comparison between camera-based pose estimation and 3D avatar-based motion visualization for biceps curl tracking.** The figure illustrates the limitation of single-view camera-based pose estimation, which provides joint angles from a fixed 2D perspective, compared to the proposed 3D virtual environment where the avatar can be observed from multiple viewpoints. This enables more comprehensive analysis of arm movement and joint angles during biceps curl exercises.

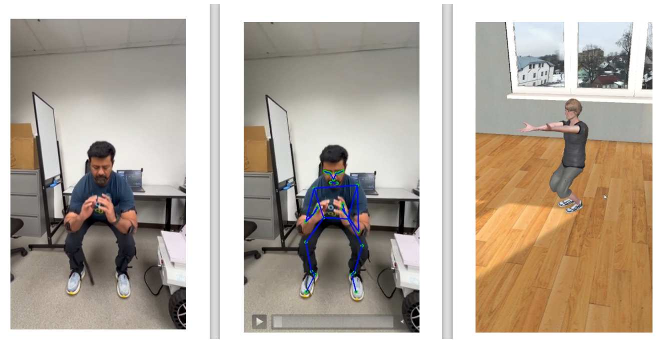

Fig. 6. **Comparison between camera-based pose estimation and 3D avatar-based visualization for squat analysis.** The left and middle images show the real user and the corresponding MediaPipe Pose skeleton from a front-view camera, where the movement appears acceptable. The right image shows the same motion in the 3D virtual environment from a different viewpoint, revealing improper posture that is not clearly visible in the 2D camera view.

TABLE I
SOFTWARE AND HARDWARE STACK FOR THE PROPOSED SYSTEM.

Category	Tool / Technology	Role
Game Engine	Unity 3D	Real-time rendering and animation
Programming	C#, Python, C++	System logic and sensor processing
3D Modeling	Blender + Mixamo	Environment and avatar setup
Sensors	10× IMU (Accel + Gyro)	Motion capture (25 Hz)
Communication	UDP (Port 5005)	Data transmission
AI / CV	MediaPipe Pose + OpenCV	Form detection and validation
Hardware	Windows PC with GPU	Processing and visualization

A. Live Avatar(Real-Time Motion Mirroring)

The live avatar is the most important part of the system because it shows the user's movements in real time. The humanoid model (Remy/Jody) is controlled directly by quaternion data from the IMU sensors, not by pre-recorded animations. The Unity C# receiver uses the sensor orientations that come in at each frame to move the bones of the humanoid rig. This makes the motion replication continuous and accurate.

Keyframe-based animation shows the ideal movement, whereas the live avatar shows the user's actual movement, even if it isn't perfect. This mirroring in real time gives quick visual feedback, which makes the device useful for training and recovery.

B. Multi-View Visualization vs. Camera-Based Form Detection

Using computer vision to analyze forms from cameras is a quick approach to figure out the positions and angles of joints from video data. This study utilizes MediaPipe Pose to derive a 2D skeletal representation from a singular camera perspective and to calculate joint angles for workout evaluation. This method makes it possible to automatically find movement patterns and do a rudimentary form evaluation.

But these kinds of solutions are limited by the camera's point of view. Because the analysis is done on only one 2D projection, some changes in posture may not be fully visible. For instance, a front-view camera may show that the joints are in the right position during a squat, but it may not catch key mistakes like not bending the hips enough or leaning forward with the trunk.

To get around this problem, the suggested system uses a digital twin of the user in a 3D virtual world. You may see the avatar from any angle with an interactive orbit camera, and it moves based on IMU sensor data. This lets you look at motion from the front, side, and oblique angles without needing more cameras.

The use of AI to estimate poses and digital twins to show them off is a big benefit of the proposed approach. The computer vision module does single-view analysis automatically, but the 3D avatar lets you see the same movement from other angles, which gives you more options (see Figure 5). This makes it easier to understand how people work together and find postural mistakes more accurately.

Figure 6 shows the difference between estimating a stance with a camera and the new 3D avatar-based visualization during a squat workout. The left and middle pictures show the real user and the MediaPipe Pose skeleton from the front-view camera. From this point of view, the movement seems fine because the detected joint alignment doesn't show any big changes.

The right image, on the other hand, shows the identical movement from a different angle utilizing the 3D avatar in the virtual world. From this point of view, it's clear that the user isn't standing correctly because the squat depth and body alignment are off. This difference shows a major flaw in single-view computer vision methods, where critical spatial information might be lost when it is projected into 2D space.

C. Trainer Controls

The system provides a set of user controls within the Unity interface to support interaction and analysis:

- **Freeze Frame:** Pauses the animation at a specific moment, allowing detailed inspection of posture.

- **Replay Last Rep:** Replays the most recent movement sequence for review.
- **Exercise Selection:** Allows switching between supported exercises (squat, biceps curl, shoulder press), updating system behavior accordingly.
- **Avatar Selection:** Enables switching between male and female avatars without interrupting the system.

These controls allow trainers to analyze movements efficiently and interact with the system in real time.

D. Supported Exercises

The system supports three primary exercises:

- **Squat:** Focuses on lower-body motion, including knee flexion, hip movement, and ankle positioning.
- **Biceps Curl:** Tracks elbow flexion and detects compensatory shoulder movement.
- **Shoulder Press:** Monitors shoulder elevation and upper-body alignment.

These exercises demonstrate the system's ability to capture and analyze both upper- and lower-body movements in real time.

V. Conclusion

This work addresses a significant issue in physical therapy and fitness tracking: the absence of objective, multi-angle analysis of human mobility, particularly when people are far away. To provide a complete 3D view of a user's movements, the suggested system combines IMU-based motion capture, real-time data transmission, and a Unity-based digital twin.

Unlike conventional camera-based techniques, which only display one view, the proposed technology enables thorough spatial investigation of motion using an interactive avatar. This improves the accuracy of form assessment by allowing trainers and practitioners to view posture from various perspectives. By automatically estimating poses for validation, a computer vision module improves the system even further.

Overall, the system demonstrates how integrating wearable sensors with virtual environments can improve the flexibility, comprehensibility, and utility of exercise monitoring for both in-person and remote settings.

During the system's development, a number of problems were found:

- Sensor Sensitivity and Drift: IMU sensors are very sensitive to noise and changes in the environment, which causes their orientation to drift over time. Calibration helps with this problem, although keeping precision over time is still hard.
- Latency and Data Stability: UDP wireless transmission can cause packet loss and jitter in some network situations, which might make avatar movement less smooth.
- Camera-Based Limitations: The MediaPipe Pose module needs the right lighting, camera placement, and user look. Because of this, posture estimation may not be as accurate in difficult situations.
- Limited Exercise Coverage: The current implementation only supports a few exercises, and adding more to the system involves more setup and testing.

A. Future Work

Future work will concentrate on enhancing both precision and system functionalities. One significant goal is to make artificial intelligence work better with virtual reality. The technology can do a better job of motion mirroring and shape detection by merging AI-based pose estimation with the digital twin representation.

Also, improved sensor fusion methods can be used to make IMU data more stable and less likely to drift. Adding more exercises that can be done, making it possible to recognize exercises automatically, and making it possible to use the cloud are all important topics for future development. These changes will make the system even more reliable and scalable for use in the real world.

Acknowledgements

The author thanks the following resources for help:

Adobe offers a free character library called Mixamo, which is where the humanoid avatar models (Remy and Jody) came from. The Unity Asset Store provided the assets for the virtual environment, while Blender was used to model it.

Google built the open-source MediaPipe Pose framework that the pose estimation part uses. OpenCV is utilized for the image processing that happens behind the scenes. The Unity 3D engine was used to make the system work in real time.

This project did not get any money from outside sources. The University of North Texas offered the DTSC 5777 course in the spring of 2026, and this study was done as part of it.

References

[1] C. Gu, W. Lin, X. He, L. Zhang, and M. Zhang, "Imu-based motion capture system for rehabilitation applications: A systematic review," *Biomimetic Intelligence and Robotics*, vol. 3, no. 2, p. 100097, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667379723000116

[2] V. Phan, K. Song, R. S. Silva, K. G. Silbernagel, J. R. Baxter, and E. Halilaj, "Seven things to know about exercise classification with inertial sensing wearables," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 6, pp. 3411–3421, 2024.

[3] S. García-de Villa, D. Casillas-Pérez, A. Jiménez-Martín, and J. J. García-Domínguez, "Inertial sensors for human motion analysis: A comprehensive review," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–39, 2023.

[4] L. Yin, P. Chen, J. Xu, Y. Gong, Y. Zhuang, Y. Chen, and L. Wang, "Validity and reliability of inertial measurement units for measuring gait kinematics in older adults across varying fall risk levels and walking speeds," *BMC Geriatrics*, vol. 25, no. 1, p. 336, 2025.

[5] Y. S. Cho, S. H. Jang, J. S. Cho, M. J. Kim, H. D. Lee, S. Y. Lee, and S. B. Moon, "Evaluation of validity and reliability of inertial measurement unit-based gait analysis systems," *Annals of Rehabilitation Medicine*, vol. 42, no. 6, pp. 872–883, 2018.

[6] Z. Zeng, Y. Liu, X. Hu, M. Tang, and L. Wang, "Validity and reliability of inertial measurement units on lower extremity kinematics during running: A systematic review and meta-analysis," *Sports Medicine Open*, vol. 8, no. 1, p. 86, 2022.

[7] E. Ha, G. Byeon, and S. Yu, "Full-body motion capture-based virtual reality multi-remote collaboration system," *Applied Sciences*, vol. 12, no. 12, p. 5862, 2022.

[8] M. Kocur, M. Kloss, C. Schaufler, V. Schwind, and N. Henze, "Investigating the impact of customized avatars during exercise in vr," in *CHI 2025*, 2025.

[9] M. Park, Y. Cho, G. Na, and J. Kim, "Application of virtual avatar using motion capture in immersive virtual environment," *International Journal of Human–Computer Interaction*, vol. 40, no. 20, 2023.

[10] V. Weistroffer, F. Keith, A. Bisiaux, C. Andriot, and A. Lasnier, "Using physics-based digital twins and extended reality for the safety and ergonomics evaluation of cobotic workstations," *Frontiers in Virtual Reality*, vol. Volume 3 - 2022, 2022. [Online]. Available: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2022.781830