Pose-Based Motion Analysis for Physical Exercise Quality Assessment

Wojciech Jakieła[1], Bartosz Chaber[2]

[1]Faculty Of Electrical Engineering, Warsaw University Of Technology, Warsaw, Poland, [0009-0008-2311-3776]
[2]Faculty Of Electrical Engineering, Warsaw University Of Technology, Warsaw, Poland, [0000-0002-0917-2162]

Faculty of Electrical Engineering
WARSAW UNIVERSITY OF TECHNOLOGY

Introduction

Correct exercise technique is key for safe and effective training, but proper feedback usually requires a personal trainer, which can be costly and inconvenient. With the rise of home workouts and fitness apps, there's a growing need for automatic systems that provide objective, real-time feedback. Existing solutions often rely on wearables or RGB-D cameras, which limit accessibility.

In this work, we propose a method that uses only a single RGB camera (e.g., a smartphone) to assess the quality of exercise execution in 3D. We present a full processing pipeline, highlighting key design choices and directions for future improvement. We focus on one exercise type — the overhead deep squat.

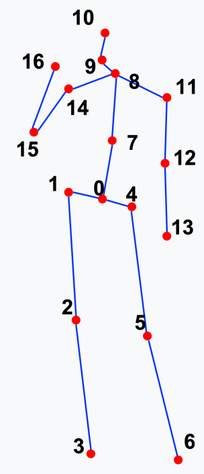
Datasets

Human3.6M (H3.6M) is a widely used dataset for 3D human pose estimation. It provides motion capture data and a standardized skeletal joint model. This dataset was used to train the pose estimation component.

Custom ODS Dataset was created to evaluate the movement assessment pipeline. It contains 15 recordings of Overhead Deep Squats (ODS) and 3 vertical jump recordings used as negative examples. Each ODS clip includes 3 repetitions, annotated by experts with quality scores:

- **Perfect:** 1.0 – 0.9
- **Good:** 0.9 – 0.7
- **Average:** 0.7 – 0.3
- **Poor:** below 0.3

A single recording rated as "perfect" was selected as the reference. A frame-accurate execution fragment was extracted for comparison using the described method.



Spatial model based on H3.6M

Methodology: Processing Pipeline

1 — 3D Pose Estimation

The system begins with real-time localization and tracking using pretrained **YOLOv4 for human detection** and **SORT for target tracking.** Then the pose estimationperformed using a **MotionAGFormer model, which lifts 2D joint predictions to 3D coordinates.** This hybrid architecture combines global Transformer-based reasoning with local GCN-based joint interaction, producing frame-wise 3D skeletal poses.

2 — Skeletal Model

A simplified skeletal model is built using the estimated 3D joint coordinates. For each joint connected to two or more segments, unit vectors are calculated to represent the direction of connected limbs. Then the segment lengths are normalized eliminating individual anthropometric differences. The model is centered at the pelvis, placing it at the origin and expressing all other joints relative to it.

3 — Feature Vector

To represent joint rotations, the system uses quaternions instead of Euler angles to avoid issues like gimbal lock and ambiguity. For each joint, the relative rotation between two connected bones is calculated using quaternion algebra based on connections beetwen them. The result is a angular feature vector computed for each frame, with importance weights assigned per joint to reflect exercise-specific relevance.

4 — Exercise Fragment Detection

To isolate valid exercise repetitions, the system performs action detection using simple thresholding method. A smoothed signal derived from joint rotations is thresholded to detect motion segments. The method should favor sensitivity, allowing some false positives in this step to ensure no valid repetitions are missed. The output is a set of filtered feature vectors for each identified exercise instance.

5 — Temporal Alignment

To compare user performance with a reference despite speed differences, the system uses **Dynamic Time Warping (DTW)**. DTW aligns two sequences of quaternion-based features by minimizing their cumulative distance, allowing comparison across variable-length motions. The algorithm outputs both the total alignment cost and a mapping between time steps, enabling temporal synchronization of the input and reference sequences.

6 — Quaternion Distance Metric

Since quaternions represent rotations on a hypersphere, the system uses **geodesic distance** instead of Euclidean for comparing feature vectors. This ensures rotational consistency and correct alignment, especially when quaternions differ only in sign.

7 — Quality Score Function

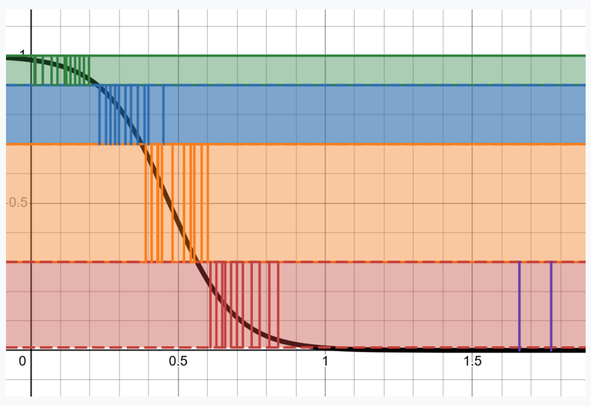
The total DTW cost reflects the mismatch between the user's movement and the reference, but is not user-friendly. To make results interpretable, the system uses a quality score function that transforms the DTW cost into an intuitive measure.

Experiments

To map DTW alignment cost to a normalized quality score, an inverted sigmoid function was experimentally fitted to using dataset labels. Sigmoid was modified using additional parameters, tuned during experiments, to reflect data trend.

$$y = \frac{1}{1+\exp(k(x-x_0))}$$

This approach enables clear feedback by transforming motion similarity into a normalized 0–1 quality scale and it becomes a basic classifier.



The distribution of exercises data with their DTW cost (after aligning with reference); marked annotated quality zones and fitted classifier function.

Conclusion

The fitting result of the score function is satisfactory and provides a solid foundation for further research. The selected function fits very well in the Perfect and Poor zones. Instances of different movements (jumps) are correctly separated. However, there are issues in the Good and Average zones, where some points do not align well. Mixing of these zones is also visible, preventing ideal separation. The DTW cost space alone is insufficient to capture subtle differences in these regions.

Future Work

An important contribution to this area of research would be to build a dataset with more recordings, different types of exercises, and most importantly expert annotations of common exercise execution errors. Future research directions include:

1. Creating a classification method for specific execution faults, such as:
   - knee valgus
   - excessive forward lean
   - incorrect tempo
   - limited range of motion
   - heel lift
2. Providing more detailed and personalized feedback, as suggested in related literature.
3. Switching to a skeletal model based on 3DPW, which better reflects natural movement of body segments.
4. Testing and analyzing the effect of DTW constraints, such as Sakoe-Chiba and Itakura windows.
5. Exploring the role of additional descriptors, including angular or linear velocities, accelerations, and curvatures.

Key References

- Huang et al. (2023) – Wearable sensors for motion monitoring.
- Jiang et al. (2020) – Real-time 3D pose from RGB-D.
- Ionescu et al. (2014) – Human3.6M dataset.
- Bochkovskiy et al. (2020) – YOLOv4 object detection.
- Mehraban et al. (2024) – Transformer-based pose estimation.
- Tharatipyakul & Pongnumkul (2023) – Feedback in movement analysis.
- Sakoe & Chiba (1978) – DTW alignment techniques.
- Shoemake et al. (1985+) – Quaternion rotation methods.
- Świtoński et al. (2019) – Motion segmentation with DTW.
- von Marcard et al. (2018) – 3DPW dataset with IMUs.

**Full reference list available in the article.*