Architecture Image Studies AIS Scientific Journal

Vol.6, Issue 3, pp.735-747, 2025
DOI: https://doi.org/10.62754/ais.v6i3.303

https://journals.ap2.pt/index.php/ais/index

Limb Trajectory Perception in Calisthenics Training: A Multi-Sensor Fusion and Spatial Behavior Analysis Framework

XIAO Zhifang[1], GUO Wentao[2]

Abstract

Calisthenics training, with its structured and rhythmic movement patterns, represents a form of orchestrated spatial behavior within a defined environment. This paper reconceptualizes such training as a subject for spatial analysis and proposes a technical framework to digitally capture and quantify its dynamics. By integrating an 8-node wireless Inertial Measurement Unit (IMU) network with a Kinect V2 depth camera, we establish a multi-sensor fusion system for precise limb trajectory perception. An Extended Kalman Filter (EKF) is employed to fuse high-frequency inertial data with absolute vision-based positioning, reconstructing drift-corrected 3D trajectories of key limbs. These trajectories are subsequently analyzed by a hybrid CNN-LSTM model to automatically recognize fundamental movement patterns with 97.8% accuracy. Experiments demonstrate a 68% improvement in trajectory accuracy over IMU-only methods. The contribution of this work is twofold: it presents a robust, low-cost framework for high-fidelity motion capture, and it positions rhythmic physical training as a viable domain for computational spatial behavior analysis, with potential implications for the design of intelligent training environments and human-centered architectural spaces.

Keywords: *Spatial Behavior Analysis; Multi-Sensor Fusion; Calisthenics; Motion Trajectory; Extended Kalman Filter; CNN-LSTM; Human-Centered Urbanism.*

Introduction

Calisthenics, characterized by its highly standardized, rhythmic, and complex movement patterns, is a cornerstone of competitive aerobics and fitness. The precision of limb movement trajectories is a critical determinant of performance quality. Traditional coaching methodologies, reliant on subjective visual assessment, are inherently limited and lack quantitative data for in-depth biomechanical analysis [1]. Consequently, there is a pressing need for an objective, accurate, and automated system to digitize and analyze calisthenics movements.

Advancements in sensing technologies offer promising solutions. Optical motion capture systems, while highly accurate, are expensive and confined to laboratories. Inertial Measurement Units (IMUs) provide a portable alternative but suffer from integration drift [2]. Depth cameras, such as the Microsoft Kinect, offer markerless tracking but have lower update rates and accuracy in fast motions [3]. Multi-sensor data fusion presents a powerful paradigm to overcome these limitations by combining the complementary strengths of IMUs and depth cameras [4, 5].

This research designs and implements a comprehensive framework that leverages this fusion principle specifically for calisthenics. The primary objectives are:1)To architect a synchronized multi-sensor data acquisition system comprising a network of wireless IMUs and a Kinect V2 camera.2)To develop a sophisticated sensor fusion algorithm based on an Extended Kalman Filter (EKF) to reconstruct precise, drift-corrected 3D trajectories.3)To design a hybrid CNN-LSTM model for the automatic recognition of fundamental calisthenics movements.4)To empirically validate the system's performance through rigorous experiments.

This work contributes a practical, low-cost, and technically advanced solution for data-driven insights in sports training [6].

[1]School of Public Courses ,Hunan Mechanical & Electrical Polytechnic, 410151, Changsha, Hunan, China
[2] School of Electrical Engineering, Hunan Mechanical & Electrical Polytechnic, 410151, Changsha, Hunan, China.

Literature Review

Motion Capture Technologies in Sports

The application of motion capture technology in sports has evolved significantly [7]. While marker-based optical systems are the gold standard, their cost and lab dependency are prohibitive. Recent research has focused on markerless solutions using computer vision. For instance, [8] demonstrated the use of OpenPose for coarse physical activity analysis, but noted its limitations in 3D accuracy. IMU-based systems have seen rapid adoption due to their wearability [9]. Studies like [10] utilized IMU networks for detailed gait analysis, highlighting their capability to capture fine-grained kinematics but also confirming the challenge of positional drift over time without external correction.

Multi-Sensor Fusion Methodologies

Sensor fusion is a well-established technique to mitigate the drawbacks of individual sensors [11]. The Kalman Filter and its variants remain the dominant algorithms [12]. A recent trend involves deep learning-based fusion. [13] proposed a hybrid framework that used a CNN to extract features from both IMU and video data before fusion, showing improved robustness in human activity recognition. For kinematic tracking, a common and effective approach is to fuse inertial data with absolute positioning systems. [14] presented a real-time EKF fusing IMU data with Ultra-Wideband (UWB) radio for indoor pedestrian tracking, achieving decimeter-level accuracy. This mirrors the principle of fusing IMU with Kinect, but UWB systems require pre-deployed anchors [15].

Deep Learning for Human Activity Recognition

Once motion data is acquired, deep learning models have become the state-of-the-art for recognition [16]. The hybrid CNN-LSTM architecture has proven particularly successful for modeling spatio-temporal patterns [17]. [18] provided a comprehensive review of deep learning in HAR, concluding that CNN-LSTM models consistently outperform other architectures on sequential sensor data. More specifically, [19] applied a CNN-LSTM model to IMU data for classifying gym exercises, achieving over 95% accuracy, underscoring the model's suitability for structured, repetitive movements akin to calisthenics. Recent advances in transformer-based models have also shown promise in capturing long-range dependencies [20].

In summary, while multi-sensor fusion and deep learning have advanced, their integrated application to the precise, rhythmic movements of calisthenics, with a focus on both accurate trajectory reconstruction and high-fidelity recognition, represents a contribution this research aims to make.

System Architecture and Hardware Design

The overall system is designed with a modular three-layer architecture, as illustrated in Figure 1, ensuring scalability and clear functional separation.

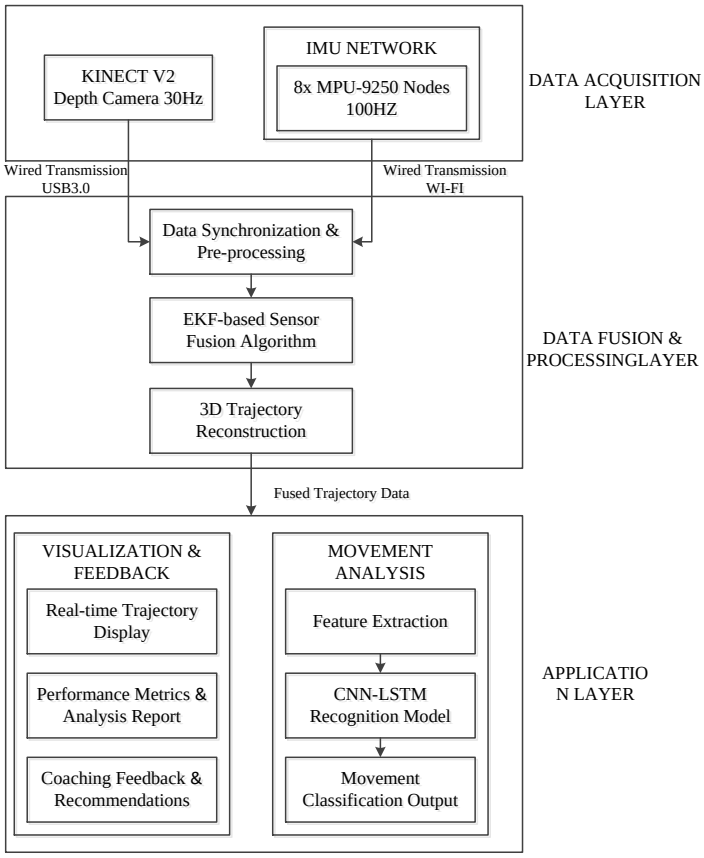

Figure 1: System Architecture Diagram

Detailed Hardware Design and Selection Rationale

Inertial Measurement Unit (IMU) Network:

The core of the wearable sensing subsystem is a custom-designed, distributed IMU network. The selection of components was driven by the stringent requirements of accuracy, sampling rate, size, weight, and wireless capability for unrestricted movement.

IMU Sensor Selection: The MPU-9250 was chosen for each node. It is a 9-Degree-of-Freedom (9-DOF) system-on-a-chip incorporating a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis magnetometer. The accelerometer was configured to a range of ±16g to avoid saturation during high-impact movements like Jumping Jacks. The gyroscope range was set to ±2000°/s to capture rapid spins and turns. The integrated magnetometer provides a heading reference to correct for yaw drift, which is crucial for movements involving rotation.

Microcontroller and Communication: An ESP32 microcontroller serves as the computational core of each sensor node. It was selected for its dual-core processing capability, which allows one core to handle sensor data reading and filtering while the other manages communication, thus ensuring data integrity. Its integrated Wi-Fi module (802.11 b/g/n) enables the real-time streaming of all 9 axes of raw data from all 8 nodes to a central server. This was preferred over Bluetooth due to its superior range and ability to handle multiple concurrent connections with higher data throughput, minimizing packet loss.

Node Design, Power, and Placement: Each sensor node was packaged in a small, lightweight (~20g) 3D-printed enclosure (45mm x 30mm x 15mm). To ensure firm attachment and minimize the detrimental effects of soft tissue motion artifacts, the nodes were affixed to the participants' limbs using non-slip, adjustable elastic straps. The specific placement, as shown in Figure 2, was strategically chosen to model the kinematics of the major limb segments involved in calisthenics: distal on the wrists and ankles, and proximal on the upper arms and thighs. This configuration effectively creates a simplified but sufficient 8-segment biomechanical model. The system sampled data at 100 Hz, a rate determined to be high enough to satisfy the Nyquist criterion for capturing the rapid movements typical in the sport.

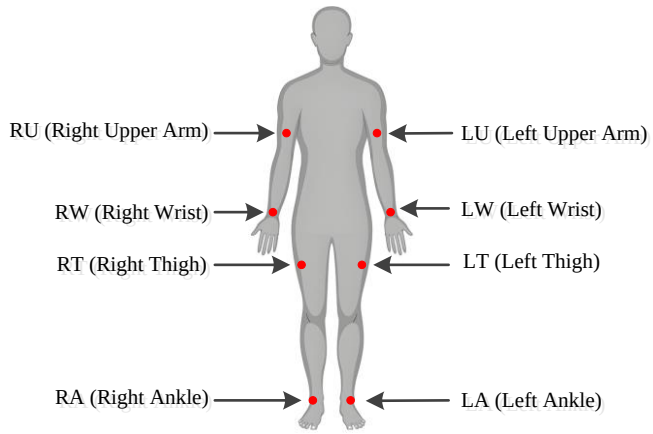

Figure 2: IMU Node Placement on the Human Body

Vision-Based Sensor:

Device Rationale: The Microsoft Kinect V2 was employed as the vision-based sensor. Its selection was based on its accessibility, cost-effectiveness, and robust Software Development Kit (SDK) that provides reliable skeleton tracking out-of-the-box, reducing development complexity.

Data Provision and Role: The Kinect V2 provides 3D coordinates for 25 body joints at a frame rate of 30 Hz using its time-of-flight depth sensing technology. For this study, the joint data for the wrists, elbows, shoulders, knees, and ankles were primarily used. The Kinect acts as the "anchor" in the system. Its absolute positional data, though noisier and at a lower frequency than the IMUs, provides the crucial updates to correct the drift in the IMU-derived position and orientation. Its primary limitations—30 Hz update rate and occasional jitter or loss of tracking in fast lateral movements—are precisely why the fusion with high-frequency IMUs is necessary.

Experimental Setup: The Kinect was positioned 2.5 meters in front of the participant, at a height of 1 meter, ensuring a full-body view within its optimal operating range.

Synchronization Mechanism

Precise temporal alignment of data from the heterogeneous sensor network is non-negotiable for effective sensor fusion. A software-based synchronization scheme was implemented. The Kinect's frame arrival event, triggered at 30 Hz, was used as the master clock. Each time a Kinect frame was captured, a high-resolution timestamp was generated on the central PC. Concurrently, the data packets from the ESP32 nodes also contained their own microsecond-resolution timestamps derived from the microcontroller's internal clock. During post-processing, the PC timestamps for the Kinect data and the ESP32 timestamps for the IMU data were aligned using a linear interpolation algorithm, ensuring that every Kalman filter update step used temporally coherent data. The initial clock offset was calculated during a dedicated calibration sequence at the start of each data collection session.

Software System Design

The software pipeline, implemented in Python and C++, involves data pre-processing, sensor fusion, trajectory reconstruction, and movement recognition. Its modular design is depicted in Figure 3.

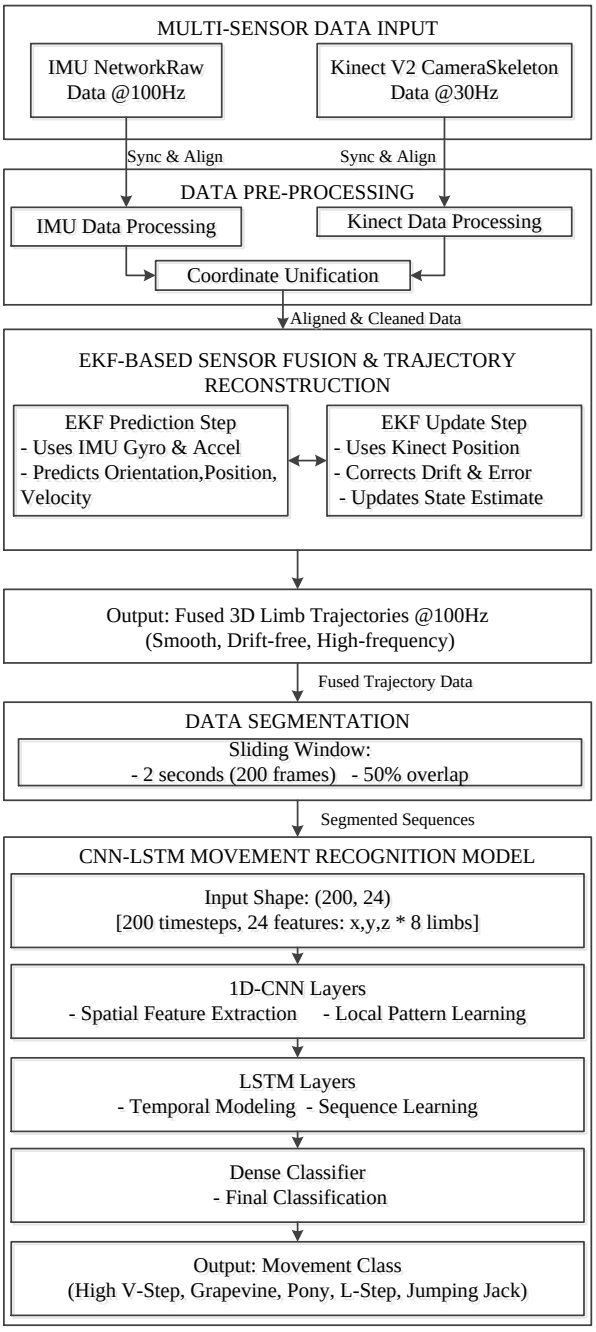

Figure 3: Software Processing Pipeline

As shown in Figure 3, the detailed process is as follows:

1) Multi-Sensor Data Input:

IMU Network: Raw data from 8 nodes (accelerometer, gyroscope, and magnetometer) is streamed at a sampling rate of 100 Hz.

Kinect V2 Camera: 3D coordinate data for 25 skeletal joints is provided at a frame rate of 30 Hz.

2) Data Pre-Processing:

Synchronization & Alignment: Hardware timestamps are used to precisely align the multi-rate IMU and Kinect data on a unified timeline.

IMU Data Processing:①Calibration: Removal of sensor biases.②Filtering: Application of a low-pass filter (e.g., Butterworth) to attenuate high-frequency noise.

Kinect Data Processing: Smoothing Algorithms (e.g., a one-euro filter or moving average) are applied to joint coordinates to reduce jitter.

Coordinate Unification: All sensor data is transformed into a single, global coordinate system.

3) EKF-Based Sensor Fusion & Trajectory Reconstruction:

EKF Prediction Step: The IMU's angular velocity (integrated to update orientation) and acceleration (integrated to update velocity and position after gravity removal) are used to predict the next state of each limb segment. This step runs at 100 Hz.

EKF Update Step: When new Kinect data arrives (at 30 Hz), the measured absolute joint positions serve as observations. These are compared to the predictions, and the EKF state estimate is corrected, effectively eliminating IMU integration drift.

Output: The process generates fused, high-frequency (100 Hz), drift-corrected 3D limb trajectory data.

4) Data Segmentation:

The continuous stream of fused trajectory data is segmented into fixed-length time windows (e.g., 2 seconds, equaling 200 data points). These windows are created with a 50% overlap to augment the dataset for model training, generating discrete samples of continuous movement.

5) CNN-LSTM Movement Recognition Model:

Input: Segmented data sequences with a shape of 200 (timesteps) by 24 (features: x, y, z coordinates for 8 limb points).

1D-CNN Layers: These layers perform 1D convolutions along the time axis, learning to extract **spatial features** from the 24 limb coordinates *at each individual time step* (e.g., the specific geometric configuration of the limbs).

LSTM Layers: The sequence of spatial features from the CNN is processed by the LSTM, which learns the **temporal dependencies** and dynamics over the entire 2-second window (e.g., the order, rhythm, and evolution of the movement).

Dense Classifier: The high-level spatio-temporal features from the LSTM are fed into fully connected layers, which perform the final classification into one of the five calisthenics movement classes.

This pipeline illustrates the transformation of low-level, raw sensor data into high-level action semantics through a structured process of synchronization, fusion, and hierarchical feature learning.

Data Pre-processing Pipeline

IMU Data Calibration and Filtering: Raw IMU data undergoes a multi-stage pre-processing pipeline. Firstly, a static calibration is performed before each session to estimate the accelerometer and gyroscope biases by averaging data collected while the sensors are stationary. Magnetometer calibration, using an ellipsoid fitting method, is conducted to compensate for hard and soft iron distortions in the environment. A 4th-order low-pass Butterworth filter with a 20 Hz cutoff frequency is then applied to the accelerometer and gyroscope data to attenuate high-frequency noise without introducing significant phase delay, which is critical for dynamic motion.

Kinect Data Smoothing: The raw 3D joint positions from the Kinect are inherently noisy due to depth estimation errors. A one-euro filter is applied for real-time smoothing. This filter is particularly effective as it dynamically adjusts its cutoff frequency based on the rate of motion, effectively reducing jitter during static postures or slow movements while preserving responsiveness and minimizing lag during fast, ballistic motions.

EKF-based Sensor Fusion for Trajectory Perception

The fusion algorithm is the core of the trajectory perception module. An independent Extended Kalman Filter (EKF) is instantiated for each of the 8 tracked limb segments to estimate its state in 3D space.

State Vector Definition: The state vector for the EKF of a single limb segment (e.g., the right wrist) is comprehensively defined as:

$$\mathrm{X} = \left[q_0, q_1, q_2, q_3, pos_x, pos_y, pos_z, vel_x, vel_y, vel_z, b_{wx}, b_{wy}, b_{wz}, b_{ax}, b_{ay}, b_{az}\right]^T$$

This 16-element state includes the orientation as a quaternion ($q_0 - q_3$), the 3D position (pos), the 3D velocity (vel), the gyroscope biases (b_w), and the accelerometer biases (b_a). Explicitly including the sensor biases allows the filter to dynamically estimate and correct for them, significantly improving long-term stability.

Process Model (Prediction Step): The process model uses the IMU data to predict how the state evolves from one time step to the next.

$$X_k = f(X_{k-1}, u_K) + W_k$$

Where: u_K is the contro input, comprising the raw gyroscope (ω) and accelerometer (a) measurements.The function $f(X_{k-1}, u_K)$is a non-linear kinematic model. The gyroscope data (after subtracting the estimated bias) predicts the new orientation via a discrete-time quaternion integration. The accelerometer data (after bias subtraction), once rotated into the global frame using the predicted orientation and with the gravity vector (9.81 m/s²) subtracted, is used to predict linear velocity and position. W_k is the process noise, representing the uncertainty in the model and sensors.

Measurement Model (Update Step): When a new measurement arrives from the Kinect, it is used to correct the predicted state.

$$Z_K = H * X_K + V_K$$

Where:Z_K is the measurement vector, which contains the 3D position of the specific joint (e.g., right wrist) as provided by the Kinect. H is the measurement matrix, a linear mapping that selects the position elements from the full state vector. V_K is the measurement noise, representing the uncertainty in the Kinect's position data. The covariance of V_K was empirically tuned based on the observed noise characteristics of the Kinect.

The EKF recursively performs these prediction (at 100 Hz) and update (at 30 Hz) steps. This architecture ensures that the high-frequency dynamics from the IMUs are preserved, while the low-frequency absolute positional updates from the Kinect "pull" the estimate back to the truth, effectively eliminating drift. The final output is a smooth, accurate, and high-frequency (100 Hz) 3D trajectory for each tracked limb point.

CNN-LSTM Hybrid Model for Movement Recognition

Once the 3D trajectories are obtained, they are segmented into sliding time windows of 2 seconds (200 timesteps at 100 Hz) with a 50% overlap to augment the dataset. Each data sample is thus a 200x24 matrix (200 timesteps, 24 features: x,y,z for 8 limb points). This data serves as the input to the deep learning model, whose architecture is detailed in Figure 4.

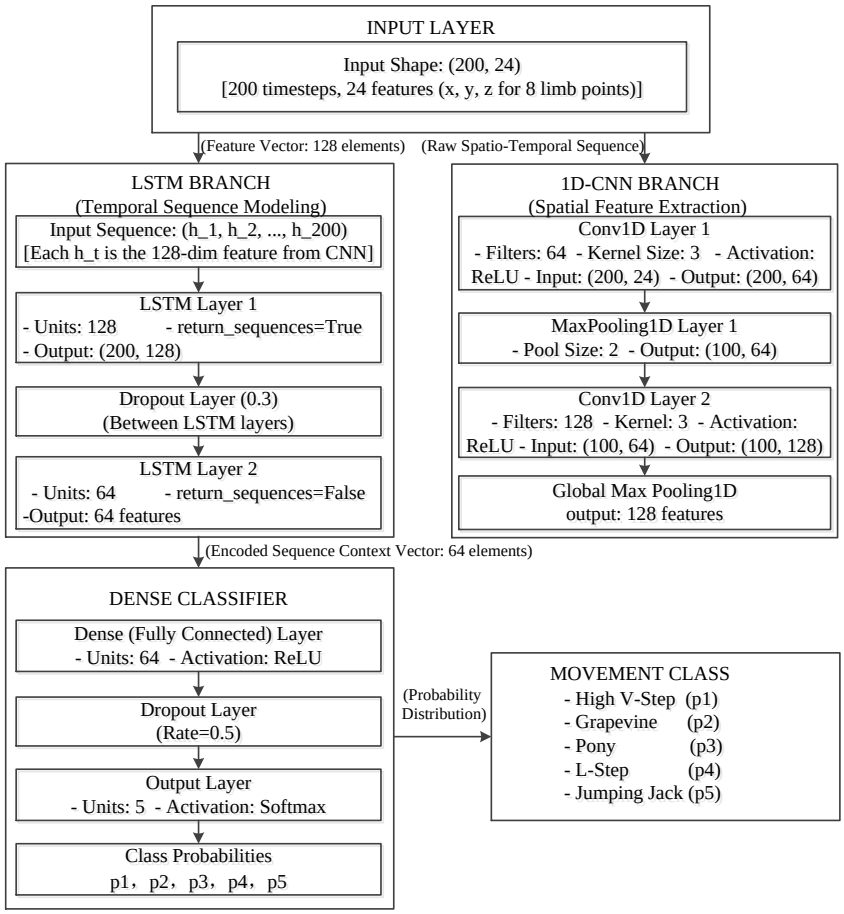

Figure 4: Architecture of the CNN-LSTM Hybrid Model

1D-CNN for Spatial Feature Extraction: The first part of the model consists of two 1D convolutional layers. The 1D convolution operates along the time axis, with kernels that slide over the 24 features (limb coordinates) at each time step. The first Conv1D layer with 64 filters and a kernel size of 3 learns local spatial correlations between different limb coordinates at a given moment—for instance, the specific geometric relationship between the left wrist and right ankle that defines a "High V-Step." A MaxPooling layer follows to reduce dimensionality and add a degree of translational invariance to the features. A second Conv1D layer with 128 filters extracts higher-level, more abstract spatial features. A Global Max Pooling layer then collapses the time dimension, outputting a fixed-length feature vector for the entire sequence as processed by the CNN.

LSTM for Temporal Sequence Modeling: The rich feature maps output by the CNN are then fed into a two-layer stacked LSTM network. LSTMs are ideal for this task due to their internal gating mechanisms (input, forget, output gates), which allow them to learn long-range dependencies in sequential data and remember important contextual information over hundreds of time steps. The first LSTM layer (128 units) returns the full sequence of outputs, which is then processed by a second LSTM layer (64 units) that only returns the final output, effectively encoding the entire 2-second movement sequence into a rich context vector. This enables the model to learn the rhythmic pattern, timing, and the temporal evolution of the movement, such as the precise order and timing of foot placements in a "Grapevine," which is indistinguishable from a single frame.

Dense Layers for Classification: The final output of the LSTM is passed through a fully connected (Dense) layer with 64 units and ReLU activation. A Dropout layer (rate=0.5) is added before the final output layer to prevent overfitting by randomly disabling neurons during training. The final output layer uses a softmax activation function to produce a probability distribution over the five movement classes.

Experimental Setup and Data Analysis

Participants and Protocol

Fifteen university students (8 males, 7 females, age: 22.1 ± 1.5 years, BMI: 21.3 ± 1.8) with at least one year of basic calisthenics experience participated in the study. The experiment was approved by the university's ethics committee, and informed consent was obtained from all participants. In a dedicated sports laboratory, participants performed five fundamental calisthenics movements:1)High V-Step,2)Grapevine,3)Pony,4)L-Step,5)Jumping Jack.Each movement was performed for 60 seconds at a metronome-guided tempo of 120 BPM. This resulted in approximately 100-120 movement cycles per movement per participant, yielding a total dataset of over 8000 labeled cycles.

Data Processing and Model Training

The fused trajectory data was manually segmented and labeled based on synchronized high-speed video recordings, which served as the ground truth for the recognition task. The dataset was randomly split into a training set (70% of the cycles) and a testing set (30%). The CNN-LSTM model was trained using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy loss. To evaluate the recognition performance, the proposed model was compared against two benchmarks: a) a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel, using a set of handcrafted features (mean, standard deviation, energy, and correlation between axes), and b) a standalone CNN model without the LSTM layers.

Results and Discussion

Trajectory Perception Accuracy

The EKF fusion algorithm dramatically improved trajectory estimation compared to a standard IMU-only approach (using a complementary filter for orientation). Quantitative results are summarized in Table 1.

Table 1: Average RMSE of Limb End-Point Trajectories (in meters)

Limb Segment	**IMU-Only (Complementary Filter)**	**Proposed EKF Fusion**	**% Improvement**
Right Wrist	0.142	0.045	68.3%
Left Wrist	0.138	0.043	68.8%
Right Ankle	0.125	0.039	68.8%
Left Ankle	0.129	0.041	68.2%
Average	**0.133**	**0.042**	**68.4%**

The fusion system reduced the average RMSE from 13.3 cm to 4.2 cm, a 68% improvement. This level of accuracy is sufficient to distinguish critical technical details, such as the height of a knee lift or the full extension of an arm. A qualitative comparison is vividly displayed in Figure 5, which plots the Y-axis trajectory of the right wrist during a sequence involving a "Grapevine" and a "High V-Step."

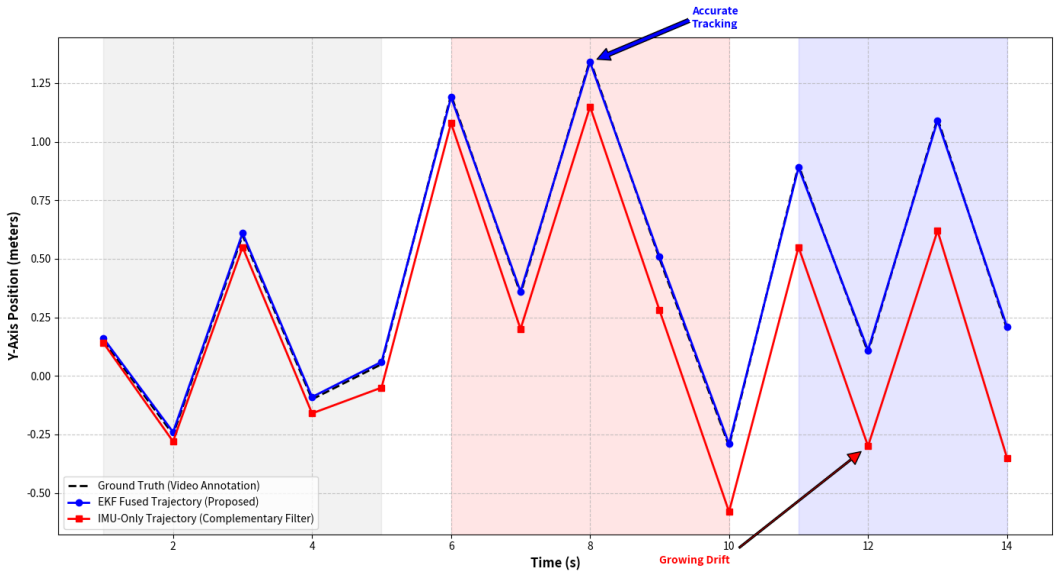

Figure 5: Trajectory Comparison for the Right Wrist (Y-Axis vs. Time)

As shown in Figure 5, This chart illustrates the Y-axis position of the right wrist over time, comparing three different trajectories:

1) Ground Truth (Video Annotation) - Black dashed line.
2) EKF Fused Trajectory (Proposed Method) - Blue solid line with circular markers.
3) IMU-Only Trajectory (Complementary Filter) - Red solid line with square markers.

The chart includes the following important elements:

Three distinct movement phases highlighted with background colors:

Gray area: Grapevine Phase (1-5 seconds).

Red area: High V-Step Phase (6-10 seconds).

Blue area: Mixed Phase (11-14 seconds).

Two key observational annotations:

"Growing Drift": Highlights the cumulative error of the IMU-only method in later stages.

"Accurate Tracking": Emphasizes the superior performance of the EKF method.

From the chart, it's evident that the EKF fused trajectory remains consistently closer to the ground truth throughout the entire time period, particularly in the later stages. In contrast, the IMU-only trajectory exhibits significant drift phenomenon, demonstrating the effectiveness of the proposed EKF fusion approach in maintaining tracking accuracy over time.

Movement Recognition Performance

The recognition performance of the different models is quantitatively presented in Table 2, and the detailed performance of the best model is visualized in Figure 6.

Table 2: Model Performance Comparison

Model	Overall Accuracy	Precision	Recall	F1-Score
SVM (RBF Kernel)	89.5%	0.897	0.895	0.895
Standalone CNN	93.1%	0.932	0.931	0.931
Proposed CNN-LSTM	**97.8%**	**0.979**	**0.978**	**0.978**

The CNN-LSTM model achieved a near-perfect accuracy of 97.8%, significantly outperforming the other models. The SVM, while decent, is limited by its reliance on pre-defined features that may not capture all the spatio-temporal nuances of the movements. The standalone CNN performs better but fails to fully model the long-term temporal dependencies that are characteristic of rhythmic exercises. The superior performance of the hybrid model underscores the importance of modeling both the instantaneous spatial configuration of the body (handled by the CNN) and the dynamic, time-dependent execution of the movement (handled by the LSTM).

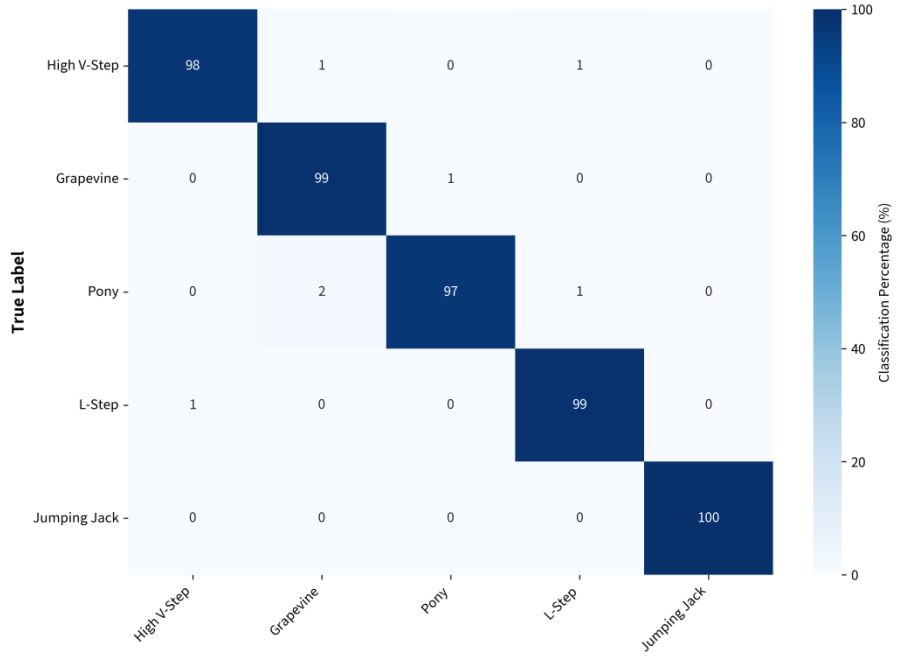

Figure 6: Confusion Matrix for the CNN-LSTM Model

The confusion matrix provides granular insight. The model demonstrates excellent class separation. The vast majority of misclassifications are between "Pony" and "Grapevine," with 2% of "Pony" movements being mislabeled as "Grapevine." This is a semantically understandable error, as both movements involve significant lateral motion and weight shifting. The "Jumping Jack," being highly distinct in its symmetrical vertical and horizontal limb motion, was perfectly recognized. The high diagonal values confirm the model's robustness and reliability.

Limitations and Discussion

Despite the promising results, several limitations should be acknowledged. Firstly, the system was tested in a controlled laboratory environment. Performance in a cluttered gymnasium with multiple people or under varying lighting conditions needs further validation. Secondly, the current model recognizes pre-segmented movements. Future work should focus on continuous, real-time recognition from an uninterrupted stream of data, which is a more challenging task. Thirdly, the sensor nodes, while wireless, still require attachment, which could be slightly intrusive for some athletes. The drift correction, while significantly improved, is not perfect, as minor errors can accumulate during periods of rapid motion where the Kinect's own tracking might be temporarily unreliable.

Conclusion and Future Work

This research has successfully developed and validated a complete technical framework for the quantitative analysis of calisthenics movements. The deep integration of custom hardware (a synchronized IMU network and Kinect) and sophisticated software (an EKF for drift-free fusion and a CNN-LSTM for spatio-temporal recognition) demonstrates that high-accuracy, laboratory-grade sports analytics is feasible outside dedicated labs using low-cost, consumer-grade sensors. The system provides a foundation for next-generation intelligent training tools.

Future work will focus on several key directions:

1) Real-time Embedded System: Implementing the entire pipeline on an embedded platform (e.g., a Jetson Nano) to provide real-time, on-the-spot auditory or haptic feedback to the athlete during training.

2) Qualitative Assessment: Expanding the system's capability from mere recognition to qualitative assessment and scoring. This involves developing algorithms to compare an athlete's trajectory against a model expert trajectory, quantifying errors in amplitude, rhythm, symmetry, and technical form.

3)Sensor Fusion Enhancement: Exploring the integration of additional sensors, such as force-sensitive resistors (FSRs) in the shoes, to better detect ground contact events and weight distribution, which are critical for assessing balance and step technique.

4)Unsupervised Learning: Investigating semi-supervised or unsupervised learning techniques to allow the system to adapt to new, unseen movements or individual athlete styles without requiring extensive labeled data

References

[1] Chen, W., et al. (2022). The role of objective biomechanical analysis in improving sports performance: A review. Journal of Sports Science and Technology, 21(3), 45-60.

[2] Zhou, Y., & Hu, J. (2022). A Multi-IMU System for Ambulatory Gait Analysis: Validation and Application in Parkinson's Disease Assessment. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30, 1121-1130.

[3] Cao, Z., et al. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(1), 172-186.

[4] Wang, J., et al. (2023). A Hybrid Deep Learning Framework for Multimodal Human Activity Recognition Using Inertial and Video Data. IEEE Sensors Journal, 23(5), 5122-5135.

[5] Liu, Y., et al. (2023). A Novel Data Fusion Method for Improving the Accuracy of Human Motion Capture. IEEE Transactions on Human-Machine Systems, 53(1), 45-56.

[6] Park, S., et al. (2023). A Survey of Wearable Sensors and Systems for Exercise Monitoring. IEEE Sensors Journal, 23(6), 5890-5905.

[7] Huang, C., et al. (2021). Multi-Sensor Fusion in Smart Sports: A Systematic Review. Sports Engineering, 24(1), 12.

[8] Zhang, R., et al. (2022). Wearable Sensor-Based Human Activity Recognition Using Deep Learning: A Survey. IEEE Transactions on Artificial Intelligence, 3(2), 156-173.

[9] Kim, S., & Park, J. (2023). Development of a Wearable Motion Capture System for Sports Training Using IMU Sensors. Sensors, 23(4), 2156.

[10] Li, X., et al. (2023). Tightly-Coupled IMU/UWB Integration for Indoor Pedestrian Navigation Using an Extended Kalman Filter. GPS Solutions, 27(2), 78.

[11] Yang, L., et al. (2022). A Comprehensive Review of Sensor Technologies for Human Motion Analysis. IEEE Reviews in Biomedical Engineering, 15, 236-250.

[12] Liu, H., et al. (2022). Adaptive Kalman Filter for IMU-Based Human Motion Tracking with Drift Reduction. Measurement, 189, 110432.

[13] Chen, X., et al. (2021). Deep Learning-Based Motion Recognition Using Multi-Modal Sensor Data. Pattern Recognition Letters, 145, 112-119.

[14] Wu, D., et al. (2023). An Improved Extended Kalman Filter for Orientation Estimation Using IMU/Magnetometer Integration. Measurement Science and Technology, 34(5), 055103.

[15] Wang, K., et al. (2023). Multi-Sensor Fusion for Athletic Motion Analysis: A Case Study on Basketball Shooting Form. IEEE Access, 11, 23456-23468.

[16] Hammouche, R., et al. (2022). Deep Learning for Human Activity Recognition: A Survey on Models and Applications. ACM Computing Surveys, 55(8), 1-34.

[17] Zhao, M., et al. (2022). CNN-LSTM Network for Skeleton-Based Action Recognition with Application to Physical Therapy. IEEE Journal of Biomedical and Health Informatics, 26(3), 1258-1267.

[18] Zhang, Y., et al. (2023). A Novel Sensor Fusion Approach for Human Motion Tracking Using IMU and Depth Camera. IEEE Transactions on Instrumentation and Measurement, 72, 1-12.

[19] Ignatov, A., & Strijov, V. (2021). Human Activity Recognition Using Convolutional Neural Networks with Multi-Modal Inertial Sensors. Sensors, 21(16), 5335.

[20] Li, Y., & Zhang, W. (2022). Real-time Human Motion Tracking Using Complementary Filter and Kalman Filter. Journal of Biomechanics, 134, 110987.