World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:18, No:7, 2024

Real-Time Fitness Monitoring with MediaPipe

Chandra Prayaga, Lakshmi Prayaga, Aaron Wade, Kyle Rank, Gopi Shankar Mallu, Sri Satya Harsha Pola

***Abstract*—**In today's tech-driven world, where connectivity shapes our daily lives, maintaining physical and emotional health is crucial. Athletic trainers play a vital role in optimizing athletes' performance and preventing injuries. However, a shortage of trainers impacts the quality of care. This study introduces a vision-based exercise monitoring system leveraging Google's MediaPipe library for precise tracking of bicep curl exercises and simultaneous posture monitoring. We propose a three-stage methodology: landmark detection, side detection, and angle computation. Our system calculates angles at the elbow, wrist, neck, and torso to assess exercise form. Experimental results demonstrate the system's effectiveness in distinguishing between good and partial repetitions and evaluating body posture during exercises, providing real-time feedback for precise fitness monitoring.

***Keywords*—**Physical health, athletic trainers, fitness monitoring, technology driven solutions, Google's MediaPipe, landmark detection, angle computation, real-time feedback.

I. Introduction

ATHLETIC trainers are indispensable in ensuring that athletes receive state-of-the-art training to improve performance while preventing injuries, a "central concern in sports medicine" [1]. a "central concern in sports medicine" [1]. However, a shortage of trainers impacts the quality of training. By providing precise techniques to prevent injuries and enhance performance, athletic trainers contribute significantly to overall well-being.

However, despite their pivotal role, a notable shortage of athletic trainers persists due to inadequate hiring or recruitment efforts [2], [3]. This shortage has significant implications for athlete care, with more than half (53%) of schools in California lacking athletic trainers, potentially leading to a decline in the quality of care provided [2]. Additionally, sporting injuries in young adolescents result in around 500,000 visits to physicians annually, with over 50% of these injuries preventable through suitable athletic training [3]. Thus, amidst disruptions to routines, such as unpredictable weather and holiday festivities, mindfulness in maintaining both physical and emotional health, supported by the guidance of athletic trainers, becomes essential.In this context, our research introduces a vision-based exercise monitoring system leveraging the MediaPipe library. This system enables precise tracking of bicep curl exercises, including good and partial range of motion, while simultaneously monitoring body posture. By tracking neck and torso inclination, our system ensures that users not only maintain proper form but also engage in a full and effective range of motion.

Lakshmi Prayaga is with the University of West Florida, USA (e-mail: lprayaga@uwf.edu).

II. Related Work

Numerous vision-based solutions have been proposed for exercise classification, each leveraging distinct methodologies and technologies to address the challenges of accurate activity recognition. Fit Coach [4], a virtual fitness coach, utilizes wearable devices to assess users' patterns and positions during workouts. Its aim is to facilitate effective workouts while preventing potential injuries. Recently, IMU (Inertial Measurement Unit) sensors have become more widely adopted for physical activity recognition. Techniques such as the Robust Step Counting from wearables are employed to track steps and detect walking, aiming to motivate users to enhance their walking-based physical activity [5]. Other methods automatically recognize various walking workouts. Exercise Trak, another approach a wearable system using a commodity smartwatch, which can continuously reconstruct the 3D posture of a single arm for 15 types of common arm exercises and provide visualization to help users adjust movements [6]. MM-Fit a different study utilizes multimodal deep learning to improve activity segmentation, exercise recognition, and repetition counting. It introduces the MM-Fit dataset, which includes inertial sensor data and synchronized RGB-D video, enhancing model robustness. The study demonstrates the efficacy of CNN-based architecture in extracting features from sensor and video data. Multimodal learning achieves high accuracy across diverse sensing modalities. The study's findings contribute to automatic recognition and segmentation of exercises, offering insights for exercise logging systems [7]. Another study VCOACH (A Virtual Coaching System Based on Visual Streaming) based on Visual Streaming explains that a critical aspect of effective exercise routines involves the integration of a virtual athletic training system, which is essential for monitoring performance and delivering timely feedback to prevent potential physical harm. In this study, innovative techniques based on learning models were developed to infer exercise recognition, predict subject biometrics, assess exercise performance, and concurrently recognize and evaluate various exercises conducted indoors, including squats, push-ups, shoulder-press, and lunges. The study explores thermal and RGB imagery using diverse descriptors like raw images, Gait Energy Image (GEI), and skeletal pose, with a thorough comparison of their performance [8]. Additionally, a study on Machine Learning-Based Exercise Posture Recognition System Using MediaPipe Pose Estimation [9], experimented to investigate three specific exercises tailored for the elderly population: 1) arm swing, 2) arm abduction and adduction, and 3) high knee movements. The extraction of key body points is facilitated through the utilization of the

MediaPipe library, which serves as a crucial tool in identifying and capturing essential anatomical landmarks, contributing to the comprehensive analysis of body movements during the specified exercises. Following the extraction of features, referred to as key points, machine learning models: Logistic regression, Ridge classifier, Random Forest classifier, and Gradient boosting classifier were used to effectively detect and classify different posture classes within each type of exercise and reported very high accuracy rates to detect individual types of postures for different exercises.

In continuation of prior research efforts, this study endeavors to broaden its scope by incorporating additional exercises, specifically focusing on bicep curls. Notably, the extension encompasses the nuanced evaluation of both Good and partial Range of Motion (ROM) variations. Simultaneously, the system is designed to adeptly detect and classify postures associated with these exercises, adding a multifaceted dimension to the overall exercise monitoring framework which provides real time feedback.

III. Dataset

Our dataset, collected specifically for this study, comprises video recordings of student participants engaging in specific exercises. These individuals were guided by a coach during their workout sessions at the University of West Florida gym. The data collection process involved recording video footage of the participants as they performed a range of exercises. The dataset encompasses a diverse set of exercises, including but not limited to bicep curls, variations in Range of Motion (ROM), and other exercises commonly performed in a gym setting. Each video was visually inspected and categorized by an exercise science expert as a good or bad workout, which comprises the target data.

IV. Methodology

A. Contribution

Our methodology integrates expert classification criteria with advanced technology to enhance exercise monitoring and form assessment. By combining insights from exercise science experts with cutting-edge vision-based monitoring systems, we aim to provide users with precise feedback for optimizing their workout routines. This section outlines our approach, from expert-defined criteria for good bicep curl form to the utilization of Google's MediaPipe framework for real-time posture detection and analysis.

B. Exercise Science Expert's Classification Methodology

Good bicep curl: Torso vertical while performing the exercises. Ears in line with shoulders to maintain good posture. Elbow fully extended at the bottom of movement with the arms perpendicular to the ground. During the concentric phase, the torso remains still while the elbow flexes and palm stays supinated. On the eccentric portion the motion returns to start and is controlled.

C. Analysis Methodology

Following inputs from the Exercise Science expert, our vision-based exercise monitoring system employs an angle-based approach to precisely track bicep curl exercises and concurrently monitor body posture, specifically focusing on the angles at the elbow, wrist, neck, and torso. This approach adds a quantitative dimension to exercise analysis, facilitating accurate form assessment and providing users with real-time feedback for optimization.

D. MediaPipe Framework

Google's MediaPipe, initially designed for real-time analysis of audio and video content on YouTube, has transformed into a versatile tool widely utilized across various domains within Google. As computer vision technology progressed, MediaPipe found applications in camera target detection, augmented reality advertising, and the integration of Google's Cloud Vision API. Since 2012, MediaPipe has evolved into a prominent open-source framework for learning and development, gaining recognition for its prowess in human posture detection. Offering a versatile and adaptive machine learning solution for live and streaming media applications. MediaPipe functions as a multimedia machine learning model pipeline, efficiently processing time series data, encompassing video and audio, with minimal resource consumption. Its cross-platform compatibility spans mobile devices, workstations, and servers, harnessing GPU acceleration across diverse devices to optimize performance; however, Mediapipe's accuracy in detecting human posture exhibits occasional inaccuracies, particularly in frames affected by changes in ambient light, leading to a decline in the recognition rate of human posture detection. To enhance the precision of 2D human posture recognition when using Mediapipe for body detection, this study gathers standard motion videos featuring professional athletes and fitness coaches engaging in various activities, serving as benchmarks for standard human posture. The methodology involves running the 12 standard human motion videos through Mediapipe, ensuring accurate human posture detection in each frame, with the 2D coordinates of each joint point sampled and recorded. Corrective measures are applied to joint point coordinates when the change rate surpasses the established threshold, mitigating instances of false detections to a certain extent [10].

In the proposed methodology, the research unfolds in three key stages. The initial stage involves landmark detection, where Google's MediaPipe is employed to accurately identify and locate key anatomical landmarks in the human body during various exercises. This landmark detection serves as the foundational step for capturing precise 2D coordinates of joint points. Following this, the second stage focuses on side detection, determining whether a person is facing right or left during the exercise. Finally, the third stage involves the computation of angles using the detected points. Leveraging the recorded coordinates, the research delves into the computation of angles, specifically targeting the angles formed at the elbow for bicep curl exercises and the angles associated with neck and torso inclination for posture assessment. These angles play a pivotal role in the detailed analysis of exercise movements and body postures, contributing to the overall accuracy and effectiveness of the proposed vision-based exercise monitoring system.

A. Flow Chart

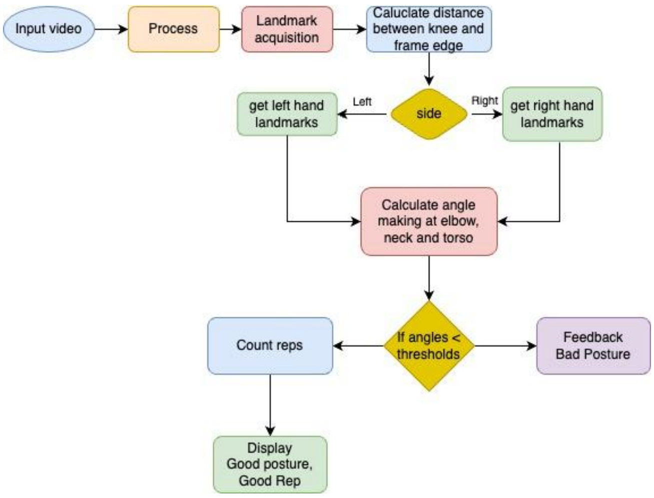

Fig. 1 Flowchart diagram of system

B. Landmarks Detection

The MediaPipe Pose Landmarker employs a series of models to anticipate pose landmarks. In the initial model, the presence of human bodies within an image frame is detected, while the subsequent model is dedicated to pinpointing landmarks on the identified bodies.

This amalgamated set of models is conveniently available as a downloadable bundle:

1. **Pose detection model**: Recognizes the existence of bodies with a selection of key pose landmarks.
2. **Pose landmarker model**: Enhances the mapping of the pose by providing an estimate of 33 3-dimensional pose landmarks.

Leveraging a convolutional neural network akin to MobileNetV2, this bundled model is specifically optimized for on-device, real-time applications in the realm of fitness. Notably, this iteration of the Blaze Pose model integrates GHUM, a 3D human shape modeling pipeline, enabling the estimation of the complete 3D body pose of an individual in both images and videos.

The pose landmarker model adeptly tracks 33 body landmark locations, providing an approximation of the spatial positioning for the body parts in Fig. 2.

C. Side Detection

During the experimentation, a limitation in Mediapipe's ability to accurately detect the left hand was noticed, particularly in scenarios where only one hand is visible in the videos. The library consistently assumed that the visible hand belonged to the right side, leading to an inability to detect the left hand accurately. To address this issue and enhance hand detection precision, a mechanism to initially identify the side of the person was devised. This was accomplished by calculating the distance between the knee and the left and right sides of the frame. If the distance between the knee point and the right side of the frame is less than the distance between the knee and the left side of the frame, it is classified as the right side, indicating the hand being used for the exercise, and vice versa. This approach aims to mitigate inaccuracies in hand detection and improve overall robustness.

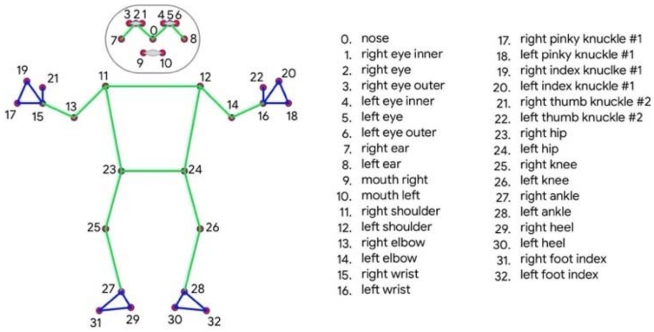

Fig. 2 MediaPipe Pose Landmarks

D. Bicep Curl Angle Calculation

Leveraging the landmarks provided by the MediaPipe Pose Landmarker model (Fig. 2), the angle at the elbow can be calculated by utilizing specific landmark points—12, 14, and 16 for the left hand, and 11, 13, and 15 for the right hand. For this paper, "Angle at the Elbow" is defined as the angle measured between the line 11 - 13 (the upper arm) and the line 13 - 15 (the forearm) for the right hand, and the line 12 - 14 (the upper arm) and the line 14-16 (the forearm). This calculated angle serves as a crucial metric for distinguishing between a proper and improper repetition of a bicep curl, allowing for the effectively identification and classifying the quality of each performed repetition.

$$\text{radians} = \text{atan2}(y3 - y2, x3 - x2) - \text{atan2}(y1 - y2, x1 - x2) \quad (1)$$

Equation (1): atan2 equation to calculate angle in radians.

$$\text{angle} = \text{degress}(radians) \quad (2)$$

Equation (2): converting radians to degrees.

$$\text{atan2}(y, x) = \begin{cases} arctan(y/x) & if\ x > 0 \\ arctan(y/x) + \pi & if\ x < 0\ and\ y \geq 0 \\ \arctan\left(\frac{y}{x}\right) - \pi & if\ x < 0\ and\ y < 0 \end{cases}$$

arctangent function atan2(y,x) for different cases of x and y.

Equation (1) calculates the angle at the elbow using the points 11, 13 and 15 from Fig. 2 as the three points x1, y1, x2, y2 and x3, y3 in the 2D plane of the video using the arctangent function. The resulting angle is then converted from radians to degrees using (2). If the angle is negative, it is adjusted to ensure it falls within the range of 0 to 360 degrees.

After computing the angle to determine the quality of the repetition, we establish a threshold, in consultation with the coach. If the elbow angle is below 75 degrees, we categorize it as a good repetition. Conversely, if the angle exceeds 75 degrees during the closing phase, it is classified as a partial repetition.

Fig 3 (a) Closing elbow angle

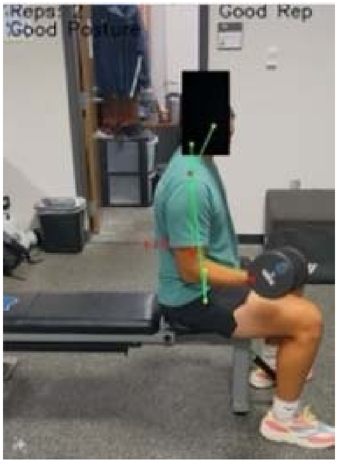

Fig 3 (b) Opening elbow angle

E. Wrist Angle Calculation

The angle at wrist can be calculated by utilizing the landmarks points 14, 16, 20 for right hand and 13, 15, 19 for left hand. So, the angle is defined as the angle measured between line 14-16 (the forearm) and 16-20 (wrist to index) for right hand and the line 13-15 (the forearm) and 15-19 (wrist to index) for left hand and to calculate the angle making at wrist, we have applied the same formula which is used to calculate the bicep curl angle.

F. Posture Angle Calculation

To precisely identify proper and improper posture during the bicep curl, we computed neck inclination and torso inclination. The neck inclination is determined by the x and y coordinates of the shoulder (11) and the x and y coordinates of the ear (7). Similarly, the torso inclination is calculated using the coordinates of the hip (23) and the coordinates of the shoulder (11). By analyzing these inclinations, we can discern correct and incorrect body postures during the exercise.

To accurately identify proper and improper body posture during bicep curls, we determined neck and torso inclination angles based on specific coordinates. The neck angle was computed using the x and y coordinates of the shoulder and ear, while the torso angle utilized the x and y coordinates of the hip and shoulder. Subsequently, we established distinct thresholds for each angle to discern correct and incorrect postures. A neck angle exceeding 40 was deemed indicative of a poor posture, whereas a torso angle surpassing 10 signified a poor posture.

Fig 4 (a) Closing wrist angle

Fig 4 (b) Opening wrist angle

V. RESULTS

The performance evaluation of our vision-based exercise monitoring system involved the calculation of the number of good and bad repetitions, as well as the identification of frames containing good and bad posture. The system exhibited remarkable accuracy, achieving a 97% success rate in accurately detecting good repetitions, good posture, and bad posture. This high level of accuracy underscores the effectiveness of the system in providing precise feedback on exercise performance and posture, contributing to its reliability as a comprehensive fitness monitoring tool.

VI. CONCLUSION

In conclusion, our methodology focused on the precise detection and classification of bicep curls, both in complete and partial Range of Motion (ROM), while concurrently identifying posture nuances such as neck and torso inclination. The system, designed for single-person detection, has demonstrated promising results in accurately capturing and analyzing the intricacies of these targeted exercises. The utilization of the MediaPipe library, combined with robust classifiers, has proven effective in achieving high-performance metrics.

Fig. 5 (a) Good Posture

Fig. 5 (b) Bad Posture

The successful implementation of our methodology not only advances the understanding of key exercises but also lays the groundwork for future developments in the realm of vision-based exercise monitoring. The insights gained from this study contribute valuable knowledge to the ongoing efforts to enhance the accuracy and versatility of exercise recognition systems.

VII. Future Work

Future work aims to broaden the exercise scope, incorporating squats, bench presses, pull-ups, and forearm exercises for a comprehensive understanding of diverse workout routines. Simultaneously, our commitment lies in improving system scalability for the simultaneous detection and classification of multiple individuals, enhancing applicability in gym or group training scenarios. This endeavor seeks to create a versatile exercise detection system adaptable to real-world scenarios, providing insights into various fitness activities while accommodating the presence of multiple participants.

Fig. 6 (a) Good Rep-Good Posture

Fig. 6 (b) Good Rep-Bad Posture

Fig. 6 (c) Good Rep-Good Posture

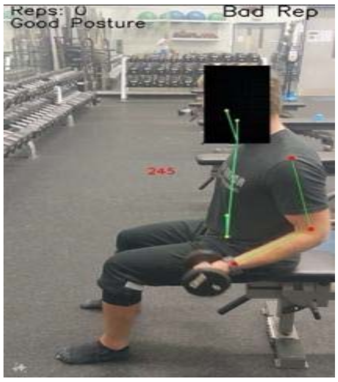

Fig. 6 (d) Bad Rep-Good posture

REFERENCES

[1] Impellizzeri, F. M., Menaspà, P., Coutts, A. J., Kalkhoven, J., & Menaspà, M. J. (2020). Training load and its role in injury prevention, part I: back to the future. *Journal of athletic training*, *55*(9), 885-892.

[2] Shanley, E., Thigpen, C. A., Chapman, C. G., Thorpe, J., Gilliland, R. G., & Sease, W. F. (2019). Athletic Trainers' Effect on Population Health: Improving Access to and Quality of Care. Journal of Athletic Training, 54(2), 124–132. https://doi.org/10.4085/1062-6050-219-17.

[3] Post, E. G., Roos, K. G., Rivas, S., Kasamatsu, T. M., & Bennett, J. (2019). Access to Athletic Trainer Services in California Secondary Schools. Journal of Athletic Training, 54(12), 1229–1236. https://doi.org/10.4085/1062-6050-268-19.

[4] *FitCoach: Virtual fitness coach empowered by wearable mobile devices | IEEE Conference Publication | IEEE Xplore*. (n.d.). Ieeexplore.ieee.org. Retrieved February 16, 2024, from https://ieeexplore.ieee.org/document/8057208

[5] *IMU-Based Walking Workouts Recognition | IEEE Conference Publication | IEEE Xplore*. (n.d.). Ieeexplore.ieee.org. https://ieeexplore.ieee.org/abstract/document/8767285?casa_token=gR9EyQ8b5a4AAAAA:5ygo3ow3pkaTY3ufvGoxKp6JBaOYjNEaWLay3lUnhtT1JkRtseFl_PIiplwle_j7c_VvSgEh-Q

[6] Wang, N. (2020). ExerciseTrak: Reconstructing Arm Posture for Upper-Body Exercises using a Wrist-Mounted Motion Sensing Device. *Ecommons.cornell.edu*. https://ecommons.cornell.edu/items/4e429547-2ee5-4f9d-af3f-9355874de6d1

[7] Strömbäck, D., Huang, S., & Radu, V. (2020). MM-Fit. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(4), 1–22. https://doi.org/10.1145/3432701

[8] Youssef, F. F., Parque, V., & Gomaa, W. (2023). VCOACH: A Virtual Coaching System Based on Visual Streaming. *Procedia Computer Science*, *222*, 207–216. https://doi.org/10.1016/j.procs.2023.08.158

[9] Lin, Y., Jiao, X., & Zhao, L. (2023). Detection of 3D Human Posture Based on Improved Mediapipe. *Journal of Computer and Communications*, *11*(2), 102–121. https://doi.org/10.4236/jcc.2023.112008

[10] *Body Posture Detection & Analysis System using MediaPipe*. (2022, March 8). https://learnopencv.com/building-a-body-posture-analysis-system-using-mediapipe/