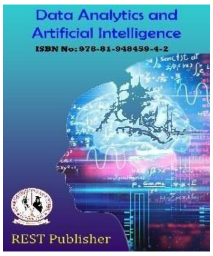
Data Analytics and Artificial Intelligence

Vol: 6(1), 2026

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/

DOI: https://doi.org/10.46632/daai/6/1/2



Smart AI Vision System for Real-Time Fitness Exercise Detection and Repetition Counting using Computer Vision and Human Pose Estimation

***Priyanshu Shaw, Sohini Chatterjee, Aditi Dey**

Narula Institute of Technology, Kolkata, India.

*Corresponding Author Email: priyanshushaw492@gmail.com

Abstract: *The advances in Artificial Intelligence (AI) and Computer Vision (CV) over the last decade have revolutionized the healthcare and fitness industry. Exercise monitoring systems in the past have been manually observed, wearable sensor based or a trainer assisted. Such systems are usually expensive, inconvenient and often rely on the human factor in terms of tedious setup and error. In this paper, the real-time exercise detection and repetition counting for multiple exercises using CV and Human Pose Estimation Tech-nuques has been proposed. The OpenCV and Media Pipe Pose based system uses human landmark detection, joint angle calculation and exercise motion analysis algorithms. Support for Push-Ups, Squats and Bicep Curls exercises with instant visualization and automated repetition counting has been proposed. An experimentation-based evaluation ensured accurate detection and repetition counting performance for normal exercise movements performed under normal indoor lighting conditions. The proposed human pose-based system is automated and provides an inexpensive solution for scalable fitness tracking without the aid of any additional wearable hardware.*

Keywords: *Artificial Intelligence, Machine Learning, Computer vision, Media Pipe, OpenCV, Human pose estimation, Exercise identification, Repetition counter.*

1. INTRODUCTION

Artificial Intelligence is one of the most powerful technologies in the field of computer science today. The application of AI in connection with Computer Vision allows machines to analyse and interpret visual information as efficiently as humans do. Increasing awareness about own physical health and fitness in recent past has significantly contributed to rising demand of intelligent fitness monitoring systems. Existing methods to track fitness performance include the use of wrist-worn devices such as a smartwatch or a fitness band. While they do offer some physiological parameters, they are not sufficient to effectively determine an exercise 's quality or the user 's form. Likewise, repetition counting via manual means (users and trainers) is often plagued with errors and inconsistencies. Most recently, the Human Pose Estimation has emerged as a promising approach to allow for contactless fitness monitoring. The pose estimation algorithms focus on detecting key points of the human body from a stream of video input and analyse the motion of those points to determine exercise repetitions and types. The now-open-source framework like Media Pipe Pose enables exercise monitoring to be done in real time on a range of computing platforms. The main aim of this work is to implement a Smart AI Vision System for Exercise Detection and Set repetition Counting automatically, using only a webcam. The system then offers a point of view while not required for the use of any wearable sensors. Enhanced healthcare and fitness training applications that are powered by combination of AI, ML and Computer Vision usage can now be designed. Activities of human body can now be monitored at every step using all these aids which offer effective Body Posture tracking with real-time detection and feedback for exercise detection. Such intelligent monitoring systems help ensure safety and maximize the efficiency of the user exercises and can also be used to promote the timely correction of the user whenever it appears that they are not performing the sports properly. In addition, booming accessibility of low-cost cameras and open-source frameworks enables vision-based fitness applications to develop faster. Frameworks such as OpenCV and Media Pipe can perform image processing, body landmark detection, as well as movement tracking efficiently. Developers of applications can now design real-time applications that are accurate and computationally inexpensive. Therefore, applications based on AI to monitor fitness have become more desirable to use in-house exercises, rehabilitation facilities, sport training facility or hospitals where Performance monitoring is necessary.

2. LITERATURE SURVEY

Recently, there has been increasing interest in developing AI algorithms and applying them to solve real-world problems in health and fitness monitoring. Researchers have investigated several approaches to automate the process of exercise monitoring, postural analysis, activity detection, and fitness evaluation, so to cost less time on manual supervision. Conventional fitness monitoring systems use wearable systems that are constituted of several types of sensors (e.g. accelerometers or gyroscopes, heart-rate sensor) as a measure of physical activity. Some of them give physiological data (e.g. heart rate, number of steps, number of calories burned). However, such systems are not free of drawbacks. For example, the user must continuously wear the system on himself, or the numerous existing systems demand a high frequency of charge. Sensor data cannot provide ergonomic information; that is to say, proper posture analysis. Thanks to the progress of Computer Vision technology, the contemporary camera-based fitness monitoring systems, which can be used to analyse the human body motion from picture and video streams, have been developed. Since the computers equipped with Computer Vision methods are able to observe and interpret the activities of human without using any sensors or devices touching the users' body, the Computer Vision has become one of the widely adopted technologies for sports, rehabilitation, surveillance, health care and gesture recognition. Human motion analysis in real time is another of its potential advantages. Human Pose Estimation has become one of the most important tools for activity recognition and exercise monitoring. Human Pose Estimation is the process of extracting body joints and landmarks from images/ videos. Human pose estimation algorithms analyse video frames to locate the positions of various body parts like shoulders, elbows, wrists, hips, knees and ankles, and then reconstruct human body pose and movement. Open Pose was one of the first frameworks to produce reliable multi-person Human Pose Estimation and has made significant advances to areas such as motion analysis systems. However, Open Pose is high in computational demand and inefficient for real-time applications on low-cost devices. More recently, two new frameworks have been available in the form of Blaze Pose and Media Pipe Pose. The latter has been used in this work. These frameworks give low computational complexity in addition to being extremely fast and accurate in landmark detection (33 body landmarks detected in real time). Compared to earlier frameworks that have used image-based pose estimation, these frameworks are faster, can successfully track body landmarks across multiple frames and are more suitable for use in fitness and rehab monitoring systems. A few recent articles proved the effectiveness of computer vision using in exercise detection and repetition counting. The pose estimation algorithms trained on these data sets can recognize the exercises like push-up, squat, lunge, jumping jacks, bicep curl, etc. Most of the systems define the stage of exercise by estimating the angles of human joints and tracking the trajectories of body movements. The results show that Computer Vision-based solutions could reach a very high accuracy of exercise recognition without requiring the introducing of costly wearable devices. However, most of the systems are still with inconveniences like low support for exercise range, light sensitivity, occlusion problem, time consuming, expensive hardware equipment needed, and no real time feedback to regard user behaviours. As a result, there is a need to develop a fast, inexpensive and friendly fitness monitoring system which can provide real-time exercise information on normal computer PC. Driven by these research gaps, this work aims to introduce a Smart Ai Vision System based on OpenCV and Media Pipe Pose Estimation. The system is capable of body landmarks recognition, exercise cataloguing, joint angles calculation and automatic counting on repetition of exercises with just a basic webcam. In practice, this offers an easier and more economical solution to fitness checking.

3. MATERIALS AND METHODS

In the following section, the hardware components, software tools and the methodology adopted for the development of Smart AI Vision System are discussed. The system under study amalgamates Artificial Intelligence, Computer Vision and Human Pose Estimation techniques to enable real-time exercise detection and counting. The Live video feed is obtained through a webcam that is streamed into the system, movements of the body are subsequently tracked by Media Pipe Pose Estimation module, joint angles are calculated, exercise stages are identified and repetitions are automatically counted. Thus, a cost-effective and sensor less body tracking system is designed to monitor work out.

Hardware Requirements: All that is needed for the Smart AI Vision System to run is core pieces of hardware, thus it is feasible to run on a common PC. The hardware I have used for this project is:

- PC or Notebook with Intel i3/i5/i7 Processor.
- Minimum 4 GB RAM
- Integrated or External USB Webcam
- Display Monitor
- Stable Power Supply

The webcam is the main input device and constantly takes live video streams of the exercising user. Frame processing, pose estimation and counting of the repetitions is carried out locally by computer.

Software Requirements: The system design can be implemented by using open source software tools and libraries which offer good support for Artificial Intelligence (AI) and Computer Vision applications.

- Python Programming Language

- OpenCV Library
- MediaPipe Pose Framework
- NumPy Library
- Flask Framework
- Tkinter GUI Framework
- Visual Studio Code / PyCharm IDE

Python as as its basic language, it is easy-to-learn and have a large set of libraries. That is compatible with Artificial Intelligence and Computer Vision. OpenCV is also used to do the image acquisition and video processing. Media Pipe Pose is adopted for real-time human pose tracking. NumPy is used to do mathematical calculations and angle calculations. Flask is used to support web-based deployment while Tkinter for desktop Graphical User Interface development.

Methodology

Intended working procedure of the Smart AI Vision System comprises a number of stages working in a sequence to fulfill exercises recognition and repetitions counting tasks. In general, the whole procedure starts with capturing the video and finishes with user feedback.

Video Acquisition: Is the initial step where videos are captured in real time through a webcam. OpenCV constantly collects video frames from the camera and feeds the stream into the network. The frame can be used as a source for pose estimation or exercise recognition. It is important to collect frames constantly to keep the system up and running smoothly.

Image Preprocessing: Preprocessing of the frames extracted is done before being fed to pose estimation by the model. As MediaPipe Pose runs on RGB images, the captured frames are converted from BGR to RGB by OpenCV. Basic preprocessing helps improve the performance efficiency of the landmark detection.

Human Pose Detection: The preprocessed frames are then fed to the MediaPipe Pose model. MediaPipe identifies 33 body landmarks, critical body joints and skeletal keypoints both in the face and in the rest of the body. Such landmarks include the shoulders, elbows, wrists, hips, knees, ankles, eyes, ears and other bodily references, used for recognizing the exercise and specific body postures.

Landmark Extraction: Upon pose detection, the individual body joint coordinates are collected from the detected land marks. Special consideration is taken to joints that are relevant for the specific exercise movement, for example the elbows for bicep curls and push-ups, and the knees for squats. The coordinate data from the land marks can then be used for the calculations.

Joint Angle Calculation: The spatially extracted landmark locations are converted into joint angles using geometric calculations. The angles between body segments are mathematically determined by applying formulas founded on trigonometric relations. Joint angles are an effective way of representing body position and motion development while performing.

Exercise Recognition: The proposed angles are used to characterize the various stages of an exercise. Thresholds are set for different angles to identify the exercise performed, the stage of the exercise and to determine if the user is executing the exercise correctly. Various angles have different patterns for different exercises. For instance, variation of elbow angles shows distinct patterns for push ups and biceps curls, while variation of knee angles differentiates between squat exercises.

Repetition Counting: Repetition counting is then done through state transition logic after exercise segments are recognized. Different types of repetition are counted when a complete motion cycle is performed by users. In bicep curl, the number of repetitions will be increased from 0 to 1 when users gone from extension to flexion and over back to the original state.

Real-Time User Feedback: The last stage of the process is the visual feedback to the user. The system overlays the identified landmarks, exercised type, the exercised stage and the number of repetitions over the video stream. This real-time feedback can motivate the user and allow the user to ensure they are executing the exercise correctly. This comprehensive approach allows the Smart AI Vision System to not only be able to function in real time but also accurately detect exercise dynamically and track repetitions in a high-quality manner. The use of Open CV and Media Pipe Pose delivers quick processing times, accurate landmark detection, and a positive fitness monitoring experience as a whole.

4. PROPOSED SYSTEM ARCHITECTURE

The architecture of the proposed Smart AI Vision System is to realize real-time exercise detection and reps counting via Computer Vision and Human Pose Estimation. As shown in Figure 1, the realtime video stream is obtained from a web-camera; the video frames are processed in OpenCV for image acquisition and pre-processing operations, then sent to the Media Pipe Pose framework. The MediaPipe Pose framework detects and tracks 33 body landmarks for identified human joints and skeletal keypoints, which capture detailed postures of the human body during movement.

The extracted landmarks are then inputs to calculate joint angles using geometric and trigonometric calculation. The angles are then used for identifying different exercise stages and types of exercises (Push-UPS, Squats, Bicep Curl). Repetition counter monitors exercise stages and increment rep counter when a full repetition cycle is passed. Finally, the results are visualized to user via desktop application and web application respectively (body landmarks, exercise type, exercise stage, repetition count).

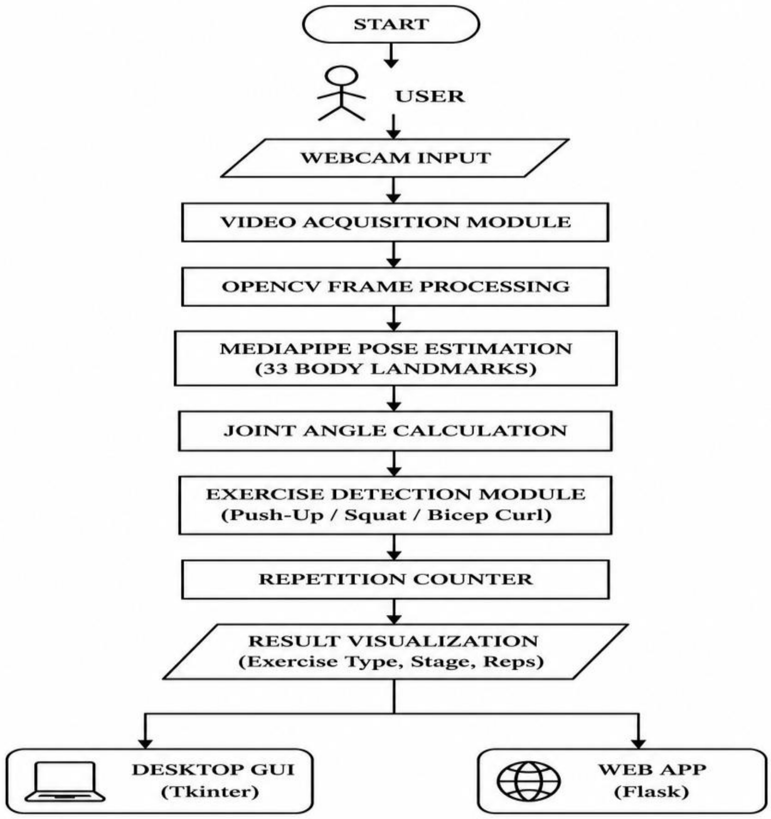


Figure 1. System Architecture of the Smart AI Vision System

FIGURE 1. System Articulture Of The Smart Al Vision System

5. RESULTS AND DISCUSSION

The system that has been proposed has been evaluated with a number of users doing Push-Ups, Squats and Bicep Curls, in the typical environment inside a room.

TABLE 1.

Detection and Counting Accuracy of Different Exercises

Exercise	Detection Accuracy (%)	Counting Accuracy (%)
Push-Up	96.5	95.8
Squat	97.2	96.6
Bicep Curl	98.1	97.4

Results from testing show great exercise recognition accuracy Bicep Curl detection has achieved the highest accuracy while Push-Up detection falls behind, as the body part is less visible due to larger amount of occlusion.

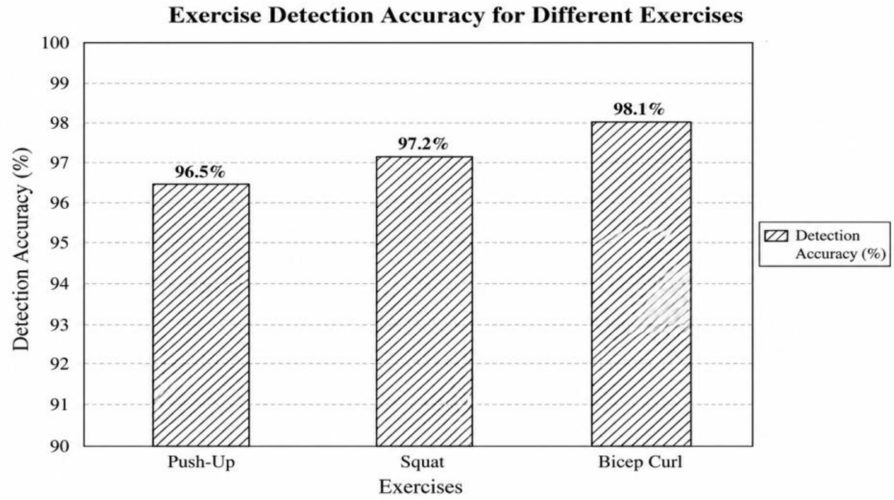


Figure 2: Detection Accuracy Graph for Different Exercises

FIGURE 2.

As shown in figure 2, exercise detection accuracy for various exercises. It can be seen that the accuracy is always above 96% for exercises we supported.

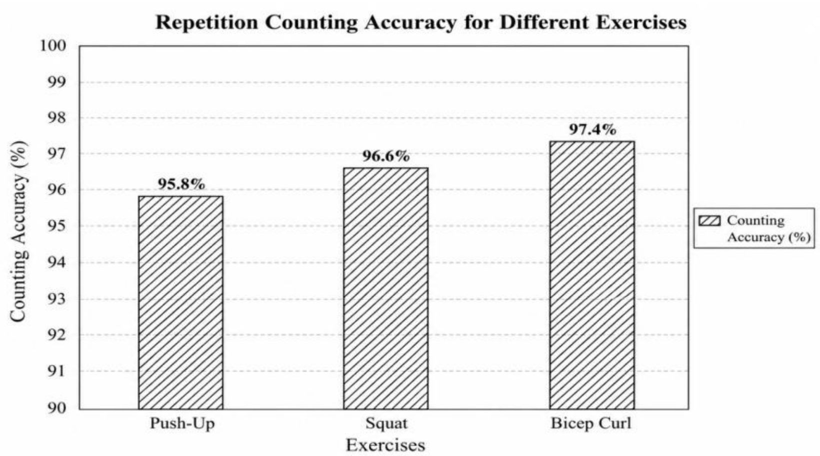


Figure 3: Counting Accuracy Graph for Different Exercises

FIGURE 3.

Figure 3 shows the repetition counting accuracy. The counting accuracy of the system remains above 95% indicating the joint-angle-based counting method is accurate.

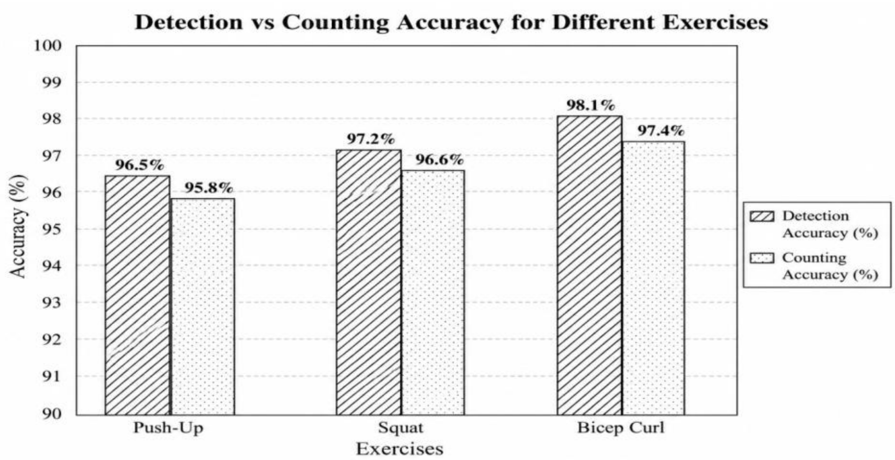


Figure 4: Detection vs Counting Accuracy Graph for Different Exercises

FIGURE 4.

Comparing the detection accuracy and the counting accuracy in Figure 4 reveals the proposed method has a noteworthily high stability.

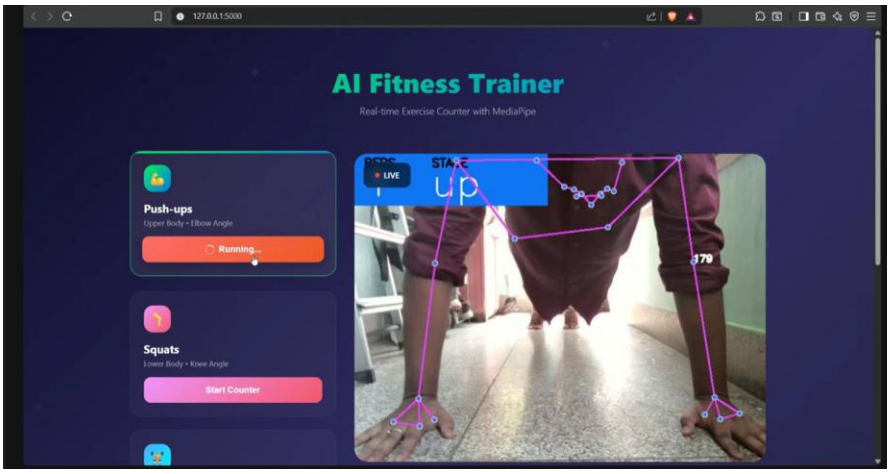


FIGURE 5. Push-Up Detection Output

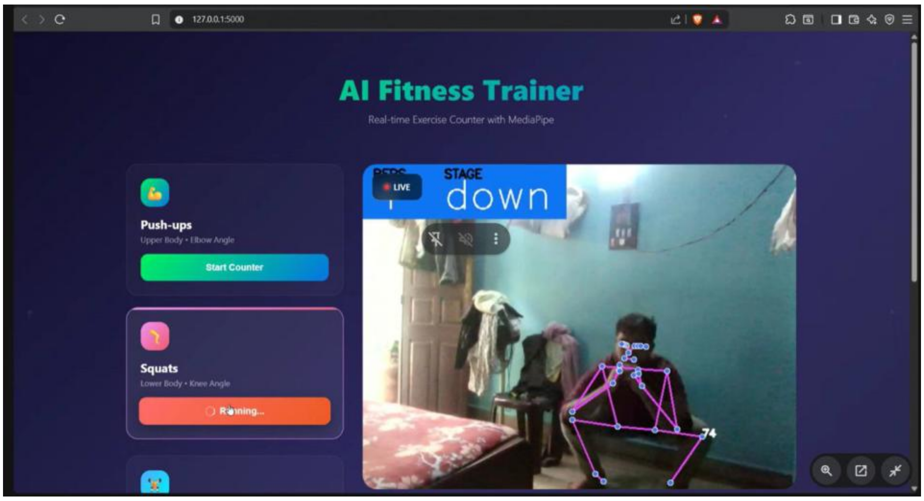


FIGURE 6. Squat Detection Output

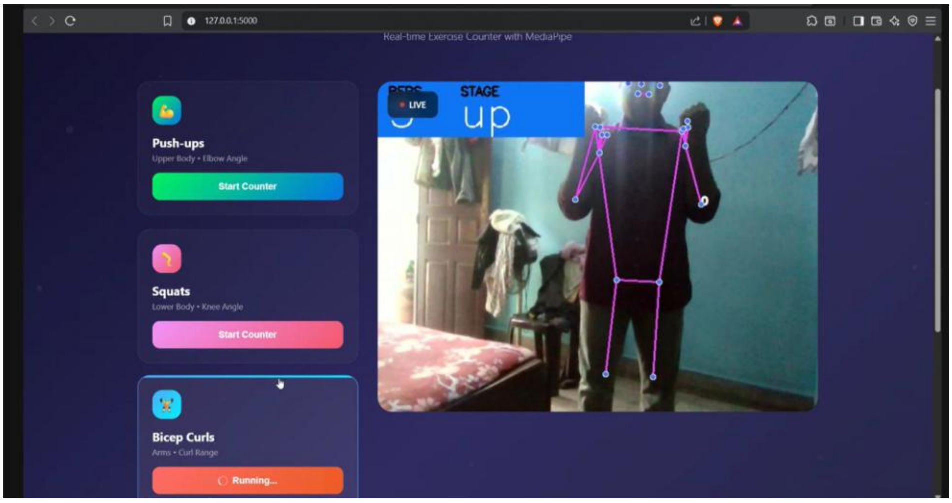


FIGURE 7. Bicep Curl Detection Output

In Figures 5.1,5.2,5.3 we can see the real exercise monitoring interface. The system quickly recognizes body landmarks and shows exercise information.

6. ADVANTAGES OF THE PROPOSED SYSTEM

It provides immediate feedback and exercise monitoring in real time. With our system every user can keep track of their workout performance with accuracy and efficiency. Use only a common webcam to detect and analyse exercises. The system does not require wearable devices or additional sensors. The system uses an open-source technologies such as OpenCV and Media Pipe Pose which makes it economical and both Hardware requirements are very less. Deep Human Pose Estimation helps the system detect most exercises and count repetitions accurately across multiple fitness exercises. Automatic reps counting method eliminates potential counting errors induced by human error and facilitate accurate workout monitoring. A simple attractive interface provides body landmarks, exercise position and repetition count in real-time. The system is always easily deployed and scaled up, with potential for increased number of exercises, correction of any possible postures, as well as mobile system presence and cloud reporting.

7. CONCLUSION

This paper proposes that a Smart Ai Vision System for real time exercise detection and repetitions counting using Ai, Computer Vision, Open CV Media Pipe Pose Estimation. The system can detect landmarks of human body, calculate joint angles, recognize exercise movements and count number of repetitions automatically. Experimental evaluation showed that detection accuracy over 96% and count accuracy over 95% for many exercises. This approach offers an inexpensive, yet scalable fitness monitoring system for current fitness and healthcare services. The results demonstrate that such a Computer Vision-based exercise monitoring system can replace manual tracking methods effectively and may end up increasing accuracy and user friendliness.

8. FUTURE SCOPE

The proposed Smart AI Vision System can be further improved by integrating a posture correction module that guides user in real-time to perform exercise with good posture and avoid injury. A voice-based virtual fitness coach can be incorporated to provide instructions on various exercises, motivational feedback and recommendations for workout suggestions through natural spoken interaction. This can be transformed into a mobile application, so the customer can access the fitness monitoring features via mobile phones and tablets. Performance Tracking using cloud-based performance analytics can be used to save workout history, track long-term Fitness progression, and produce performance reports. A calorie module can be added on to find the calories burnt by the workout which takes the exercise type and time and some parameter which it takes from us. Other future features: Multi-person gameplay support for multiple users working out or competing in the same video. Personalized workout recommendation systems based on Artificial Intelligence techniques can be created to provide workout plans tailored to a person ‘s fitness goals and current motor capabilities. In addition, the system is suitable for more advanced rehabilitation monitoring, to support physiotherapists and patients in recording specific movements of rehabilitation exercises and monitoring the rehabilitation process.

REFERENCES

[1]. Zhao, J., Wu, M., Zhou, L., Wang, X., and Jia, J., “Cognitive Psychology-Based Artificial Intelligence Review,” Frontiers in Neuroscience, vol. 16, 2022. Available: https://www.frontiersin.org/articles/10.3389/fnins.2022.1024316

[2]. Lugaresi, C. et al., “Media Pipe: A Framework for Building Perception Pipelines,” Google Research, 2019. Available: https://arxiv.org/abs/1906.08172

[3]. Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., and Sheikh, Y., “Open Pose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, 2021. Available: https://ieeexplore.ieee.org/document/8765346

[4]. Bradski, G. and Kaehler, A., Learning OpenCV: Computer Vision with the OpenCV Library, O’Reilly Media, 2008. Available: https://www.oreilly.com/library/view/learning-opencv/9780596516130

[5]. Russell, S. and Norvig, P., Artificial Intelligence: A Modern Approach, 4th Edition, Pearson, 2020. Available: https://aima.cs.berkeley.edu

[6]. Bishop, C. M., Pattern Recognition and Machine Learning, Springer, 2006. Available: https://link.springer.com/book/10.1007/978-0-387-45528-0

[7]. Szeliski, R., Computer Vision: Algorithms and Applications, 2nd Edition, Springer, 2022. Available: https://szeliski.org/Book

[8]. Google Media Pipe Documentation. Available: https://developers.google.com/mediapipe

[9]. OpenCV Documentation. Available: https://docs.opencv.org

[10]. Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning, MIT Press, 2016. Available: https://www.deeplearningbook.org

