International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET2411221

AI Trainer : Video-Based Squat Analysis

Prof. Anuja Garande, Kushank Patil, Rasika Deshmukh, Siddhi Gurav, Chaitanya Yadav

Artificial Intelligence & Data Science Engineering, Zeal College of Engineering and Research, Pune, India

ARTICLEINFO

Article History:

Accepted: 20 March 2024
Published: 06 April 2024

Publication Issue :
Volume 11, Issue 2
March-April-2024

Page Number :
172-179

ABSTRACT

This research proposes a video-based system for analyzing human squats and providing real-time feedback to improve posture. The system leverages MediaPipe, an open-source pose estimation library, to identify key body joints during squats. By calculating crucial joint angles (knee flexion, hip flexion, ankle dorsiflexion), the system assesses squat form against established biomechanical principles. Deviations from these principles trigger real-time feedback messages or visual cues to guide users towards optimal squat posture. The paper details the system architecture, with a client-side application performing pose estimation and feedback generation. The methodology outlines data collection with various squat variations, system development integrating MediaPipe, and evaluation through user testing with comparison to expert evaluations. Key features include real-time feedback and customizable thresholds for user adaptation. Potential applications encompass fitness training, physical therapy, and sports training. Finally, the paper explores future work possibilities like mobile integration, advanced feedback mechanisms, and machine learning for automatic threshold adjustments. This research offers a valuable tool for squat analysis, empowering users to achieve their fitness goals with proper form and reduced injury risk.

Keywords : Video-Based Squat Analysis, Mediapipe Pose Estimation, Real-Time Feedback, Squat Form Assessment, Biomechanics, Joint Angles (Knee Flexion, Hip Flexion, Ankle Dorsiflexion), Fitness Training, Physical Therapy, Sports Training, User Interface (UI) design.

I. INTRODUCTION

Squats are a fundamental exercise in strength training, known for building lower body strength, improving mobility, and enhancing overall fitness. However, improper squat form can lead to inefficiencies in muscle activation, increased risk of injury, and hinder progress towards fitness goals.

Problem: Many individuals, especially those new to exercise, struggle to perform squats with proper form. Traditional methods of form correction rely on visual observation by a qualified trainer, which can be expensive and not always readily available.

Challenges: Developing a system for automated squat form analysis presents several challenges:

Accuracy of Pose Estimation: Reliance on computer vision algorithms for pose estimation can be susceptible to errors due to factors like lighting conditions, clothing occlusions, or user movement outside the frame.

Real-time Processing: For effective feedback, the system needs to analyze squats in real time, requiring efficient algorithms that balance accuracy with computational speed.

Clear and Actionable Feedback: The system's feedback needs to be clear, concise, and easily understood by users of varying fitness levels. It should not only highlight form deviations but also provide actionable guidance for improvement.

This research addresses these challenges by proposing a video-based system for squat analysis utilizing MediaPipe, an open-source pose estimation library. The system provides real-time feedback on crucial joint angles (knee flexion, hip flexion, ankle dorsiflexion) to guide users towards proper squat form and enhance their overall workout experience.

II. VISION AND DUAL OBJECTIVES

Vision: Our vision is to empower individuals of all fitness levels to achieve optimal squat form through a user-friendly, accessible, and real-time feedback system. This system will contribute to safer and more effective exercise routines, maximizing workout benefits and minimizing injury risk. Dual Objectives:

1. Develop a robust and accurate video-based system for squat analysis: This objective focuses on the technical aspects of the system. We aim to leverage MediaPipe's pose estimation capabilities to create a reliable system for identifying key body joints during squats. The system should be able to calculate crucial joint angles (knee flexion, hip flexion, ankle dorsiflexion) with minimal errors under various conditions.

2. Provide real-time feedback that promotes proper squat form: This objective emphasizes the user experience and the effectiveness of the system's feedback mechanism. We aim to design clear, concise, and actionable feedback messages or visual cues that guide users towards optimal squat posture in real-time. The feedback should be adaptable to users of varying fitness levels and customizable based on individual needs.

By achieving these dual objectives, this research will contribute to a valuable tool for improving squat form and promoting safe and effective exercise practices.

III. ARCHITECTURE EXPLANATION

System Architecture for Video-Based Squat Analysis.This research proposes a client-server architecture for the video-based squat analysis system with real-time feedback. Here's a detailed breakdown of its components and functionalities:

1. Client-Side Application

Video Capture: The application captures video input from a webcam or other video source, ensuring the user's body is positioned within the frame for proper analysis.

MediaPipe Integration: The client-side application integrates MediaPipe's pose estimation library. This involves loading the chosen pose estimation model, pre-processing each video frame for compatibility with the model, and extracting the detected body keypoint coordinates (2D or 3D depending on the model) for each frame.

Joint Angle Calculation: Based on the extracted key points, the application calculates the crucial joint angles for squat analysis. This includes the knee flexion angle, hip flexion angle, and ankle dorsiflexion angle using mathematical formulas based on the 3D coordinates of relevant key points (e.g., ankle, knee, hip).

Form Assessment and Feedback Generation: The system compares the calculated joint angles in each frame with pre-defined reference ranges established based on biomechanical principles and recommendations from qualified fitness professionals. Deviations from these ranges trigger real-time feedback messages or visual cues.

Feedback Messages: Clear and concise text messages are displayed on the user interface, highlighting deviations from proper form (e.g., "Knees caving in," "Maintain a straight back").

Visual Cues: Overlays on the video feed can be used to provide visual indicators of improper form (e.g., highlighting misaligned knees with a colour overlay).

User Interface (UI): The application features a user-friendly interface with the following functionalities:

Live video feed displaying the user performing squats.

Visualization of the detected body key points overlaid on the video feed.

Real-time feedback messages or visual cues based on the squat form analysis.

Optional features like calibration options to adjust the system for individual body proportions or user profile management for personalized feedback thresholds.

2. Server-Side

While the core analysis can be performed on the client side, a server can be implemented for additional functionalities, especially when dealing with multiple users:

Centralized Data Storage: The server can store data from multiple users, including video recordings, keypoint data, and calculated joint angles. This data can be used for training machine learning models or analyzing user trends over time.

Advanced Feedback Generation: The server can host more complex feedback generation algorithms, potentially using machine learning to personalize feedback based on user data and performance history.

User Management and Access Control: The server can manage user accounts, provide access control for data storage and analysis, and potentially offer different service tiers with varying functionalities. Benefits of the Architecture:

Real-time Processing: The client-side processing enables real-time feedback without relying on a server for analysis, improving user experience.

Scalability: The architecture can be scaled by adding additional client applications for multiple users or a server for centralized data management.

Customization: The client-side application can be customized with user-specific preferences for feedback style or reference ranges for joint angles.

IV. METHODOLOGY/PLANNING OF WORK

The methodology section outlines the step-by-step process of developing and evaluating the video-based squat analysis system with MediaPipe and real-time feedback. Here's a breakdown of each stage:

1. Data Collection

Subject Recruitment: Recruit a diverse group of participants with varying levels of fitness experience.

Data Recording:

Set up a controlled environment with a high-quality camera positioned to capture the entire body during squats.

Correct Form: Participants perform squats with proper depth, back straight, knees tracking over toes, etc.

Depth Variations: Record squats with shallow depth (not reaching full knee flexion) and excessive depth (exceeding recommended knee flexion).

2. Data Labeling:

For each video frame, manually label the key body joint positions (knees, hips, ankles, etc.) using dedicated software or create a custom labelling tool.
Alternatively, explore using a more sophisticated system for automatic labelling, but be prepared to validate its accuracy against manual labelling.
Annotate the labelled data with corresponding joint angles calculated using the labelled positions.

3. System Development

Development Environment: Choose a suitable programming language and framework (e.g., Python with OpenCV or TensorFlow) based on your expertise and project requirements. MediaPipe Integration: Integrate MediaPipe's pose estimation library into your application. This involves following MediaPipe's documentation to set up the library and handle functionalities like model loading, video frame processing, and keypoint extraction. Joint Angle Calculation: Develop algorithms to calculate the crucial joint angles for squat analysis based on the detected key points from MediaPipe. This involves utilizing mathematical formulas based on the 3D coordinates of the key points. The key angles to focus on include:
Knee Flexion Angle: Angle between thigh and calf segments.
Hip Flexion Angle: Angle between torso and thigh segments.
Ankle Dorsiflexion Angle: Angle between foot and shin segments.

4..Form Assessment and Feedback Generation:

Develop algorithms to compare the calculated joint angles in each video frame with the reference ranges.
Design real-time feedback mechanisms based on the comparison results. This can involve:
Visual Cues: Overlay text messages or visual indicators on the user interface highlighting deviations from proper form (e.g., "Knees caving in," "Maintain a straight back").
Audio Cues: Generate synthesized audio messages providing corrective instructions during the exercise. (Future development).

User Interface (UI) Design: Develop a user-friendly interface with the following functionalities:
Live video feed displaying the user performing squats.
Visualization of the detected body key points overlaid on the video feed.
Real-time feedback messages or visual cues based on the squat form analysis.
Optional features like calibration options or user profile management.

4. Evaluation

Participant Selection: Recruit a new group of participants who were not involved in the data collection stage. Testing Procedure:
Participants perform squats in front of the system while following on-screen instructions.
The system provides real-time feedback on their squat form.
Simultaneously, a qualified fitness professional observes the participants and provides their assessment of their squat form.

Data Analysis:

Compare the system's feedback with the evaluations from the fitness professional.
Calculate metrics like accuracy, precision, and recall to quantify the system's effectiveness in identifying squat form deviations.
Gather user feedback on the system's usability, clarity of feedback, and overall user experience.
Refinement: Based on the evaluation results and user feedback, refine the system by:

Adjusting the reference ranges for joint angles if needed.

Improving the algorithms for joint angle calculation or feedback generation.

Enhancing the user interface for better clarity or adding new features.

Following this comprehensive methodology will ensure a well-developed and evaluated system for video-based squat analysis with real-time feedback using MediaPipe. Remember to document each step meticulously for future reference and potential improvements.

V. Algorithms

Algorithms for Squat Analysis with MediaPipe

The core algorithms in this system revolve around utilizing MediaPipe's pose estimation and applying calculations to assess squat form based on key joint angles. Here's a detailed breakdown of the algorithms involved:

1. MediaPipe Pose Estimation Integration

This initial step leverages MediaPipe's pre-trained model to detect human body key points from each video frame captured by the webcam. MediaPipe offers various pose estimation models with different trade-offs between accuracy and processing speed. Choose a model that balances these factors based on your project requirements.

The MediaPipe library provides functions to handle tasks like:

Model loading: load the pre-trained pose estimation model from MediaPipe.

Video frame processing: Preprocess each video frame for compatibility with the model's input format.

Keypoint extraction: Extract the 2D or 3D coordinates (depending on the chosen model) of the detected body key points for each frame.

2. Keypoint Filtering

MediaPipe's pose estimation might occasionally have minor inaccuracies in keypoint detection. You can implement an optional filtering algorithm to address this. Common filtering techniques include:

Kalman Filter: This can be used to smooth out noisy keypoint detections by considering the motion dynamics between frames.

Median Filter: This can be applied to a small window of consecutive frames to replace outlier keypoint positions with the median value within the window, reducing the impact of individual detection errors.

3. Joint Angle Calculation

Once you have reliable keypoint data, calculate the crucial joint angles for squat analysis. Here's how to approach this from the key angles mentioned earlier:

Knee Flexion Angle:

Identify key points for the ankle (ankle_kp), knee (knee_kp), and hip (hip_kp).

Calculate the vectors representing the thigh segment (hip_kp-kknee_kp) and calf segment (knee_kp-aankle_kp).

Use the dot product and vector magnitudes to calculate the angle between these two vectors. This represents the knee flexion angle.

Hip Flexion Angle:

Identify key points for the shoulder (shoulder_kp), hip (hip_kp), and knee (knee_kp).

Similar to the knee angle, calculate the angle between these vectors using the dot product and vector magnitudes. This represents the hip flexion angle.

Ankle Dorsiflexion Angle:

Identify key points for the ankle (ankle_kp), knee (knee_kp), and foot (foot_kp) (assuming a key point for the midfoot or toes is available).

Calculate the vectors representing the foot segment (ankle_kp-ffoot_kp) and the shin segment (ankle_kp-kknee_kp).

Apply the same principle as the previous angles to calculate the angle between these vectors. This represents the ankle dorsiflexion angle.

4. Squat Form Assessment and Feedback Generation

Establish reference ranges for each joint angle based on biomechanical principles and recommendations from qualified fitness professionals. These ranges will define the acceptable degree of flexion for proper squat form.
In each video frame, compare the calculated joint angles with the established reference ranges.
Develop algorithms to generate real-time feedback based on the comparison results:
If a joint angle falls outside the acceptable range, trigger a feedback message or visual cue. This message should clearly indicate the issue (e.g., "Knees caving in") and suggest corrective action (e.g., "Maintain knees aligned with toes").
The feedback mechanism can be visual overlays on the user interface or audio cues (future development).

VI. Applications of Video-Based Squat Analysis with MediaPipe Feedback

The video-based squat analysis system with MediaPipe and real-time feedback offers a valuable tool across various fitness and healthcare domains. Here's a detailed exploration of its potential applications:

1. Fitness Training:

Self-Guided Workouts: Individuals can utilize the system at home or in the gym to receive real-time feedback on their squat form during solo workouts. This empowers them to maintain proper posture and maximize workout effectiveness without relying on a personal trainer.
Personalized Training Programs: Fitness professionals can integrate the system into their training programs. Users can perform squats with the system's feedback, allowing trainers to remotely monitor form and provide further guidance through video conferencing or mobile apps.
Progression Tracking: Over time, the system can track user data on squat form metrics (e.g., average joint angles). This data can be used to monitor progress, identify areas for improvement, and tailor training programs for individual needs.

2. Physical Therapy:

Rehabilitation Exercises: Physical therapists can leverage the system for patients undergoing rehabilitation that incorporates squats. The real-time feedback can assist patients in maintaining proper form during exercises crucial for recovery.
Remote Monitoring: Therapists can use the system for remote monitoring of patients performing squats at home. This allows for continuous feedback and ensures patients adhere to proper form, reducing the need for frequent in-person appointments.
Data-Driven Therapy Plans: The system can collect data on a patient's squat form over time. This data can be used by therapists to tailor rehabilitation plans, track progress, and measure the effectiveness of therapy interventions.

3. Sports Training:

Improved Technique: Athletes can utilize the system to refine their squat technique, a crucial exercise for building lower body strength and power. The real-time feedback can help them achieve optimal form, enhancing performance in sports that require strong squats, such as weightlifting, jumping sports, and sprinting.
Injury Prevention: Proper squat form is essential for preventing injuries. The system's feedback can help athletes identify and correct form deviations before they lead to injuries, promoting safe training practices.
Performance Analysis: Coaches can use the system to analyze squat data from athletes, potentially

identifying weaknesses in form that can be addressed through targeted training drills.

4. Future Applications:

Mobile Integration: The system can be adapted for mobile applications, allowing users to perform squat analysis with their smartphones or tablets, further increasing accessibility and portability.

Advanced Feedback Mechanisms: With advancements in machine learning, the system could provide more personalized feedback based on user data and performance history. Additionally, audio cues could be incorporated alongside visual cues for a more comprehensive feedback experience. Gamification: The system could be integrated into fitness or rehabilitation apps with gamification elements, making squat analysis more engaging and motivating for users, especially for long-term adherence to proper form practices.

VII. CONCLUSION

This research presented a video-based system for squat analysis utilizing MediaPipe's pose estimation capabilities and providing real-time feedback for posture improvement. The system leverages MediaPipe to identify key body joints during squats and calculates crucial joint angles to assess form against biomechanical principles. Deviations from these principles trigger real-time feedback messages or visual cues to guide users towards optimal squat posture.

The research methodology outlined a comprehensive approach to data collection, system development, and evaluation. The dual objectives focused on developing a robust and accurate analysis system and providing clear, actionable feedback for users. The client-server architecture offered flexibility for real-time processing and potential scalability.

VIII. REFERENCES

[1]. Stover, Rodríguez-Moreno, I.; Martínez-Otzeta, J.M.; Sierra, B.; Rodriguez, I.; Jauregi, E. Video Activity Recognition: State-of-the-Art. Sensors 2019, 19, 3160.

[2]. Wren, C.R.; Azarbayejani, A.J.; Darrell, T.J.; Pentland, A.P. Integration Issues in Large Commercial Media Delivery Systems; SPIE: Washington, DC, USA, 1996.

[3]. Elgammal, A.; Harwood, D.; Davis, L. Non-parametric model for background subtraction. In Computer Vision—ECCV 2000; Springer: Berlin, Germany, 2000; pp. 751–767.

[4]. Barnich, O.; Van Droogenbroeck, M. ViBE: A powerful random technique to estimate the background in video sequences. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 945–948.

[5]. McFarlane, N.J.B.; Schofield, C.P. Segmentation and tracking of piglets in images. Mach. Vis. Appl. 1995, 8, 187–193.

[6]. Lucas, B.D.; Kanadee, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the Imaging Understanding Workshop, Pittsburgh, PA, USA, 24–28 August 1981; pp. 121–130.

[7]. Horn, B.K.; Schunck, B.G. Determining optical flow. In Techniques and Applications of Image Understanding; Technical Symposium East; International Society for Optics and Photonics: Washington, DC, USA, 1981; Volume 17, pp. 185–203.

[8]. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999.

[9]. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition

(CVPR'05), San Diego, CA, USA, 20–25 June 2005.

[10]. Comaniciu, D.; Meer, P. Mean shift analysis and applications. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999.

[11]. Koppula, H.S.; Gupta, R.; Saxena, A. Learning human activities and object affordances from RGB-D videos. Int. J. Robot. Res. 2013, 32, 951–970.

[12]. Ni, B.; Pei, Y.; Moulin, P.; Yan, S. Multilevel Depth and Image Fusion for Human Activity Detection. IEEE Trans. Cybern. 2013, 43, 1383–1394. [PubMed]

[13]. Wang, J.; Liu, Z.; Wu, Y.; Yuan, J. Learning Actionlet Ensemble for 3D Human Action Recognition. IEEE Trans. Pattern Anal. Machin. Intel. 2014, 36, 914–927. [PubMed]

[14]. Shan, J.; Akella, S. 3D human action segmentation and recognition using pose kinetic energy. In Proceedings of the 2014 IEEE International Workshop on Advanced Robotics and Its Social Impacts, Evanston, IL, USA, 11–13 September 2014.

[15]. Cippitelli, E.; Gasparrini, S.; Gambi, E.; Spinsante, S. A Human Activity Recognition System Using Skeleton Data from RGBD Sensors. Comput. Intel. Neurosci. 2016, 2016, 1–14.

[16]. Gao, Y.; Xiang, X.; Xiong, N.; Huang, B.; Lee, H.J.; Alrifai, R.; Jiang, X.; Fang, Z. Human Action Monitoring for Healthcare based on Deep Learning. IEEE Access 2018, 6, 52277–52285.

[17]. Adama, D.A.; Lotfi, A.; Langensiepen, C.; Lee, K.; Trindade, P. Human activity learning for assistive robotics using a classifier ensemble. Soft Comp. 2018, 22, 7027–7039.

[18]. Albu, V. Measuring Customer Behavior with Deep Convolutional Neural Networks; BRAIN. Broad Research in Artificial Intelligence and Neuroscience: Bacau, Romania, 2016; pp. 74–79.

[19]. Majd, L. Human action recognition using support vector machines and 3D convolutional neural networks. Intern. J. Adv. Intel. Inf. 2017, 3, 47–55.

[20]. Murad, A.; Pyun, J.-Y. Deep Recurrent Neural Networks for Human Activity Recognition. Sensors 2017, 17, 2556.

[21]. Qin, Z.; Zhang, Y.; Meng, S.; Qin, Z.; Choo, K.-K.R. Imaging and fusing time series for wearable sensors based human activity recognition. Inf. Fusion 2020, 53, 80–87.