Bridging the Gap: Integrating Pose Estimation And Face Detection For Enhanced Exercise Monitoring

Mohit Sharma, Sahil Choudhary, Abhishek Kumar Pandey, Shivani Chauhan, Mehak Bhatnagar

Abstract: In recent years, exercise tracking has developed as a rapidly increasing trend. This study addresses the demand for precise and comprehensive exercise monitoring by combining a posture estimation model based on MediaPipe, the face detection methodology based on the Haar Cascade method, and the OpenCV is used for computer vision tasks like video capturing. As people grow more conscious of the value of fitness and wellness, the necessity for precise and thorough activity monitoring has become critical. The primary goal of this study is to count the number of push-ups performed during exercise sessions while also detecting the individual performing the activity. So, a more comprehensive dataset for subsequent evaluation and analysis can be collected by combining pose estimation and face detection. The integrated technique is highlighted in the methodology and as a result, demonstrates its accuracy and reliability in exercise monitoring scenarios. Exercise monitoring systems can provide individuals with real-time feedback, track progress, and make personalized recommendations by utilizing this integrated approach. Furthermore, the extensive data produced as a result of this integration provides for in-depth analysis and significant insights into workout routines, allowing for the development of evidence-based fitness regimens and improving general well-being.

Keywords: MediaPipe, OpenCV, Haar Cascade

I. INTRODUCTION

Human Pose estimation produces better results and is more accurate when several view pictures are used as input. The difficulty, however, is that few people have access to specialized tools or environments capable of recording such multi-view data at home. As a result, there is an increasing demand for pose estimate programs that run on smartphones and take advantage of the capabilities of a single camera. While some posture estimate techniques, such as YOLO-posture, are intended for single-camera configurations, their shortcomings become apparent when dealing with dynamic video streams. Google MediaPipe Pose, on the other hand, provides better accuracy and processing speed, especially when working with video streams that incorporate camera inputs. [5] found that by using the power of machine learning and computer vision, Google MediaPipe beats YOLO-Pose with its advanced algorithms and methodologies. It can accurately predict human positions from video frames acquired by a single camera, making it suitable for usage on smartphones or devices with minimal hardware resources. By combining face identification with position estimation, this integrated approach allows for real-time feedback, tracking progress, and offering personalized recommendations. Furthermore, the enormous data generated as a result of this connection can be utilized for in-depth analysis and important insights into workout routines, allowing for the development of evaluation-based fitness regimens and enhancing overall well-being.

The goal of this study is to count the number of push-ups performed by individuals while also detecting their faces. The study also intends to collect pertinent information such as the individual's identity, the total number of pushups, gather photos of every single pushup, and the start and end times of the exercise session. This research aims to provide comprehensive data for exercise monitoring and can further extend to an analysis of exercise.

II. LITERATURE REVIEW

Several studies have suggested face detection using the Haar Cascade Classifier and human pose estimation with Mediapipe in recent years. However, there is still a lot of unexplored territory in the areas of face identification and position estimation for exercise tracking.

A. Human Face Detection

Paula Voila and Michael Jones [1] were the first to propose a method to combine integral images for human face detection using the Haar feature and AdaBoost based classifier.

H. Hatem [2] displayed integrated method for face detection, tracking, and head-pose estimation in his research paper. Face detection, based on Haar-like features, identifies people's faces, noses, eyes, and mouths, as well as tracking, based on affine motion model estimation, to detect profile faces. The paper includes detailed experiments on a challenging face detection and tracking dataset, covering various conditions such as illumination, scale, pose, and camera variations. As conditions change, the algorithm adapts and becomes more flexible.

Tejashree Dhawle [3] offered suggestions for how deep learning, a crucial area of computer science, might be utilized to identify faces using a variety of OpenCV modules and Python, and how this implementation can be applied to other platforms.

B. Human Pose Detection

H.-J.Park [6] paper outlines a technique for counting pushups using 2D video footage. The following section looks at significant motion characteristics related to push-up counts. 220 pushup videos, each filmed from two different perspectives, yielded 147,840 samples. The suggested strategy was modeled using half of the films, and its effectiveness was assessed using the remaining half. The research provided is entirely dependent on pushups counting with precision.

[4] Locating someone using MediaPipe Pose, which will provide a 33-landmark pose forecast. This technique's usefulness and prospective influence on the fields of

of computer vision and human posture assessment are shown by the thorough study and analysis of the technology. [5] A flexible and affordable tool for many platforms is Google MediaPipe. Without the need for high-performance hardware, it can be included in commonplace gadgets. In addition to recording body angles and landmark distances, it may also extract other data.

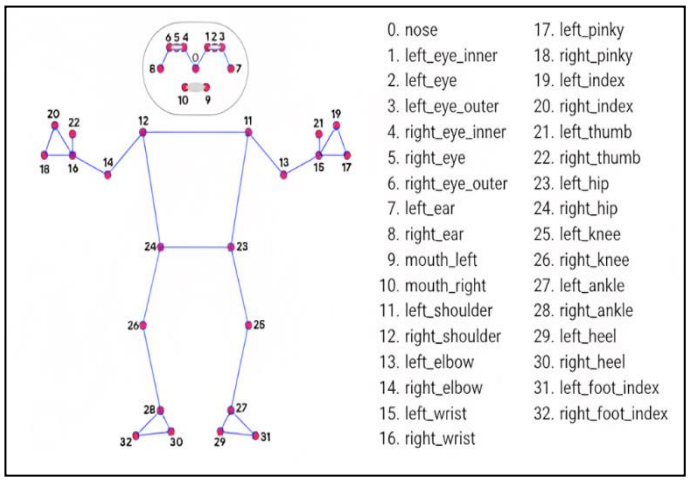

Figure 1: Landmarks Map of Google MediaPipe Pose [4]

[7] study introduces a program that classifies exercises, precisely counts repetitions, and collects sophisticated workout data from a 2D video feed of a person working out. The program aims to count each repetition like a human would, hence it does not have any specific guidelines about what counts as a repetition. It does not rate how well an exercise was carried out. The program acts as a base for upcoming programs that can offer comments on the effectiveness of exercise execution and suggestions for development. It operates in real-time, offering rapid user feedback, without knowing the intended workout or specific filming angles. The author used recordings of people doing push-ups and squats to conduct evaluations to gauge how well the program would operate in actual situations. Squat and push-up evaluations by the program showed outstanding accuracy rates of 95.57% and 93.69%, respectively. These outcomes confirm the program's dependability and efficiency in correctly identifying and evaluating these workouts.

III. STRUCTURE OF PROPOSED METHODOLOGY

The [8] Haar features are used to identify specific features or patterns in an image. These features work by calculating a value based on the difference in pixel counts between a white and a black rectangle. Haar features are often utilized as rectangular characteristics for quick human face detection, as shown in "Fig.2 ". In our proposed methodology, we utilize face detection using [9] OpenCV and Haar Cascade. Haar Cascade approach is utilized to detect faces in webcam video frames. Simultaneously, MediaPipe, particularly the posture estimation module, is utilized to estimate the person conducting the exercise's positions and landmarks for different exercises, we are especially focusing on push up counting.

The program combines the functions of Haar Cascade face detection and MediaPipe pose estimation. First, the algorithm initializes the face cascade classifier and takes the webcam video feed. The Haar Cascade approach is then used to detect faces in each frame, drawing rectangles around the discovered faces and displaying the name above each one.

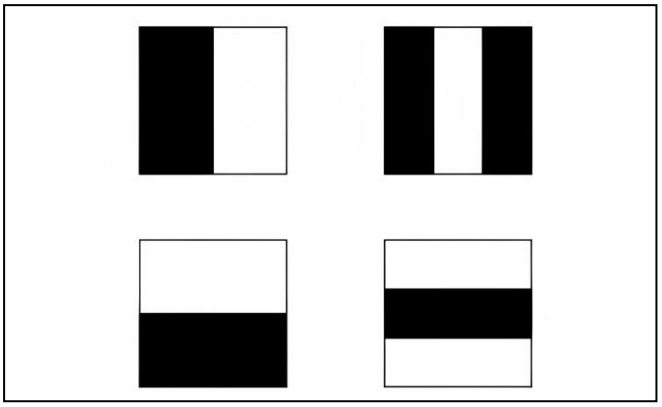

Figure 2: Haar features [1]

Following that, the code employs the MediaPipe pose estimation module to detect and track body landmarks. It draws pose landmarks on the image and checks the position of specified landmarks to see if a push-up is being executed. It distinguishes the "down" and "up" positions of the push-up movement by comparing the vertical positions of significant landmarks. The code keeps track of the number of push-ups performed and updates it when a full push-up is accomplished. It also keeps track of when the push-up exercise begins and ends. The total number of push-ups, as well as the start and end times, are saved to a CSV file. The code continually displays the video stream with the detected face rectangles and the updated push-up count throughout the procedure. Finally, when the user pushes the 'q' key, the program exits, and the webcam and [9] OpenCV windows are released and closed. Overall, this complete methodology demonstrates the use of [1] Haar Cascade for face identification and MediaPipe for position estimation to develop an exercise monitoring system capable of counting push-ups and gathering extra data for analysis and assessment.

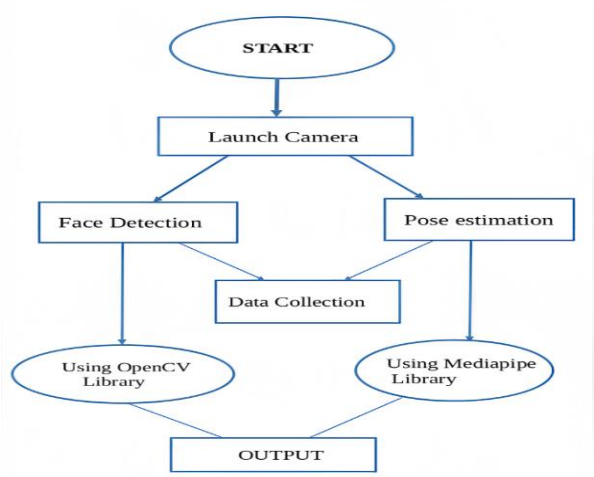

Figure 3: Flow Chart of Complete Methodology

IV.RESULTS AND DISCUSSION

In this unit, the simulation outcomes of proposed integration model accuracy is compared with existing work [10]. Face identification accuracy dropped by 10%, and there were cases where things other than actual faces were mistakenly detected as faces. But I case of pushup counting the accuracy nearly remain same with [7]. The following are the result ouput:

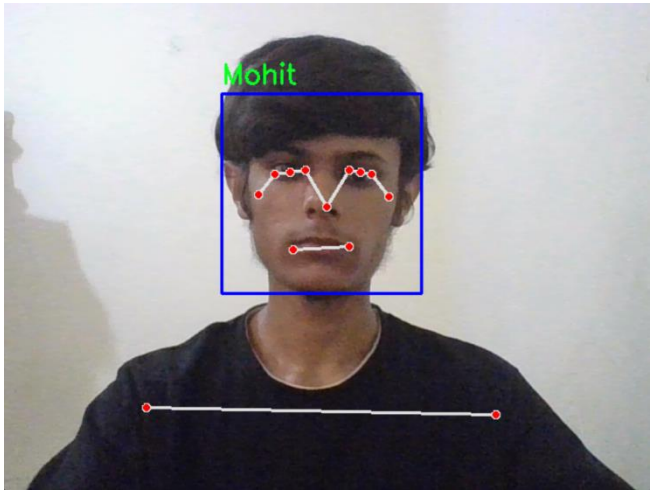

Figure 4: Haar Cascade Integration with Mediapipe Pose Estimation

Name	Total Push	Start Time	End Time
Mohit	16	17:54:35	17:55:53
Name	Total Push	Start Time	End Time
sahil	15	18:04:15	18:05:55
Name	Total Push	Start Time	End Time
Abhi	13	18:10:14	18:12:02
Name	Total Push	Start Time	End Time
Shivani	10	18:14:35	18:16:36
Name	Total Push	Start Time	End Time
Mehak	10	18:25:29	18:26:23

Figure 5: Data set collection in CSV file

V. CONCLUSION AND FUTURE WORK

In conclusion, the combination of position estimation using MediaPipe and face detection with the Haar Cascade approach provides a potent solution for improved exercise monitoring. We can revolutionize fitness tracking, data analysis, and personalized workout suggestions by combining these tools. This study advances exercise monitoring technologies and lays the door for better fitness outcomes for individuals and communities. Extension of the algorithm to recognize individual face and evaluation of every count image for detailed exercise monitoring remain as a subject for future work.

REFERENCES

1. Viola, P. and Jones, M., 2001. “Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition,” 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-I). IEEE.

2. Hatem, H., Beiji, Z., Majeed, R., Lutf, M., & Waleed, J. "Face Detection and Pose Estimation Based on Evaluating Facial Feature Selection" International Journal of Hybrid Information Technology, 8(2), 109-120. doi: http://dx.doi.org/10.14257/ijhit.2015.8.2.10 (2015).

3. Dhawle, T., Ukey, U., & Choudante, R. (2020). "Face Detection and Recognition using OpenCV and Python." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(10), e-ISSN: 2395-0056

4. “MediaPipe-Solutions-guide” https://google.github.io/mediapipe/solutions/pose

5. Zhang, Y. (2022). "Applications of Google MediaPipe Pose Estimation Using a Single Camera." ScholarWorks, California State University. Retrieved from “scholarworks.calstate.edu.”

6. H. -J. Park, J. -W. Baek and J. -H. Kim, "Imagery based Parametric Classification of Correct and Incorrect Motion for Push-up Counter Using OpenPose," 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), 2020, pp. 1389-1394, doi: 10.1109/CASE48305.2020.

7. Wallman, J. (2023). "Classifying Physical Exercises and Counting Repetitions using Three-Dimensional Pose Estimation." Institute of Architecture of Application Systems.

8. Li, Y., Xu, X., Mu, N. and Chen, L., 2016, June. Eye-gaze tracking system by haar cascade classifier. In Industrial Electronics and Applications (ICIEA), 2016 IEEE 11th Conference on (pp. 564-567). IEEE

9. OpenCV. [Online]; 2018. Available from: https://docs.opencv.org/3.4.1/d7/d8b/tutorial_py_face_detection.html

10. Priadana, A., Habibi, M. (2019). "Face Detection using Haar Cascades to Filter Selfie Face Image on Instagram." Department of Informatics, Universitas Jenderal Achmad Yani Yogyakarta. Retrieved from ResearchGate: [https://www.researchgate.net/publication/335794071]