e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

HUMAN ACTIVITY RECOGNITION USING OPENCV AND GOOGLE MEDIAPIPE

Shivam Nikam[*1], **Mohit Wadekar**[*2], **Aniket Borse**[*3]

Chinmay Mahajan[*4], **Nitin Shahane**[*5]

[*1,2,3,4]Student, Department of Computer Engineering, KK Wagh Institute of Engineering Education & Research, Nashik, Maharashtra, India.

[*5]Professor, Department of Computer Engineering, KK Wagh Institute of Engineering Education & Research, Nashik, Maharashtra, India.

DOI : https://www.doi.org/10.56726/IRJMETS36884

ABSTRACT

This paper focuses on Human Activity Recognition (HAR), a widely researched area in computer vision and machine learning. There are many different methods for the same. However, existing HAR methods are resource-intensive, prompting the need for a better alternative. In this paper, we propose a new approach to HAR using Google MediaPipe, a machine-learning framework designed to process time-series data. Our proposed method involves identifying body pose features, preprocessing input using the OpenCV library, analyzing input using a set of rules embedded in the algorithm, and obtaining output by comparing input features with predefined feature tuples with the help of the MediaPipe framework. We expect our proposed approach to improve the effectiveness of HAR with the availability of adequate activity information and the robustness of MediaPipe.

Keywords: Human Activity, MediaPipe, OpenCV, Pose Recognition, Real-time data processing, Machine learning pipeline, Deep Neural Networks.

I. INTRODUCTION

Human Activity Recognition is identified as an important research area in the field of computer science and machine learning. HAR finds applications in various fields, such as surveillance, human-robot interaction, and healthcare. In developing intelligent systems capable of assisting humans in everyday life, it is vital to recognize human activities. Existing methods of HAR are showing promising results, but they are resource intensive. The task is usually very complex and requires considerable computational power. Therefore, the need for a reliable and precise HAR system is essential. MediaPipe is a new approach for HAR, a machine-learning framework that deals with time series data in the form of real-time video inputs. It can be used for various applications such as hand tracking, face detection, pose estimation, object detection, augmented reality, etc. OpenCV is a free, open-source library for computer vision and machine learning software, containing many algorithms and tools to perform image processing, object detection, or other tasks related to computer vision. The MediaPipe analyzes the input after preprocessing with the help of OpenCV. By incorporating the strengths of both technologies for processing video data and detecting key features in people's bodies together MediaPipe and OpenCV can make real-time human activity recognition more robust and accurate. Deep learning-based approaches, particularly those utilizing convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promising results in HAR. These techniques can automatically determine features based on sensor or video frame data, thus eliminating the need for feature engineering. The recent research focuses on developing an efficient, scalable, and robust HAR system. One approach is to use lightweight neural networks that can run on low-power devices. Overall, HAR research is an active and rapidly evolving field, and new methods and techniques are continually being proposed and evaluated.

II. METHODOLOGY

The methodology for recognizing human activity used for this project was based on OpenCV and Google MediaPipe libraries. MediaPipe uses a 2-step deep neural network human detector. Detecting and localizing the human body in the video frame is the first step. This step was done by a pre-trained object detection model. Using the MediaPipe library, 33 landmark points were identified when the body was discovered. Then, these landmark points were connected to form the skeletal structure of the human body. The angles between prominent landmarks in the 3D space were determined by using the geometrical characteristics of triangles after a skeleton structure had been created. These angles were then used to determine which activity was carried out by a human subject. This was done using an algorithm that compares each activity's angle from landmark points to a predefined set of angles.

The entire process is depicted below:

Pose Detection:

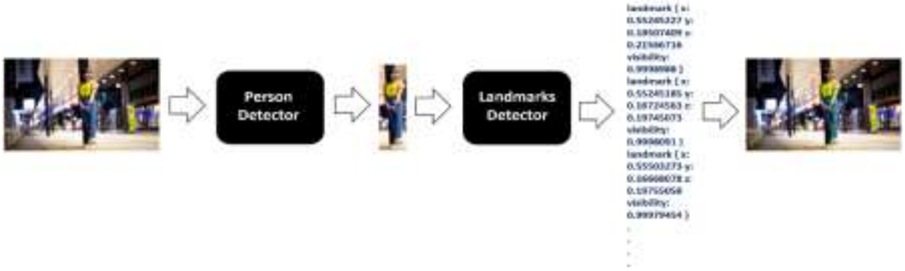

Person Detector
Landmarks Detector
landmark { x: 0.55245327 y: 0.18507409 z: 0.21566738 visibility: 0.9998868 }
landmark { x: 0.55245185 y: 0.16724583 z: 0.19745073 visibility: 0.9996091 }
landmark { x: 0.55503273 y: 0.16668078 z: 0.19755058 visibility: 0.99979454 }
.
.
.
.

Figure 1: MediaPipe 2-Step Detector Model

The 2-step machine learning pipeline of MediaPipe is used in this step. The input for the pose detection is a frame with a prominent person whose pose landmarks need to be detected. The first step involves the localization of a person within the given frame, and the second step helps predict the pose landmarks within the area of interest. For videos, the detector is used only for the first frame, and then the Region of Interest (ROI) is derived from the previous frame's pose landmarks using a tracking method. When the tracker loses the track to identify the body pose in the frame, the control shifts to the detector, and it is invoked again for the next frame. This process reduces computation and latency. The image below shows the 33 pose landmarks along with their indexes.

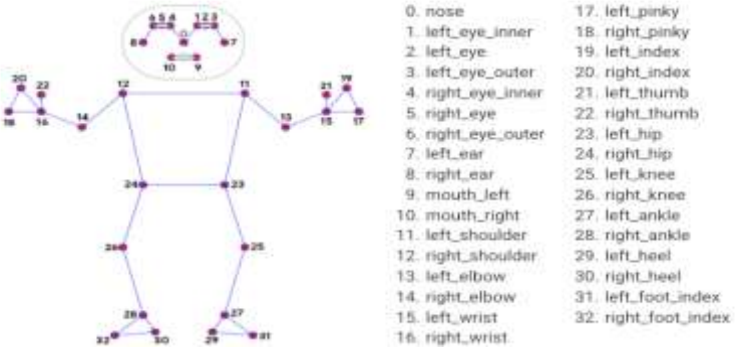
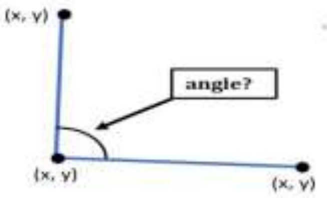


Figure 2: 33 Pose Landmarks

The detected landmarks are visualized in 3D by setting the hip landmark as the origin. Depending on the person's distance from the video recording instruments, there is a significant change in the relative distance of other points from the hip. The result of this step is the coordinates of the landmark points, which are necessary to calculate angles forming at some important landmarks in the human body posture. They are the x, y, and z-axis coordinates and one additional parameter named visibility, i.e., the extent to which the landmark is visible.

Activity Classification:



```
angle = math.degrees(math.atan2(y3 - y2, x3 - x2) - math.atan2(y1 - y2, x1 - x2))
```

Figure 3: Angles Calculation Between Landmarks

Some important landmarks in the human body pose are formed at the shoulder, knee, hip, and elbow. The angles formed at these landmarks are calculated using the mathematical equation. The values or parameters for this equation are already acquired in the previous step. It is the same as calculating angles between lines. The above figure shows the angles forming between landmarks. The first and second marks are regarded as starting lines, and the third mark is also considered an endpoint of the 1st line and a starting point on the 2nd line. The mathematical equation used to calculate the angle is shown above the figure.Initially, camera access is provided to the user with the help of the OpenCV library's 'cv2.videoCapture' object, which is essential to turn on the webcam or any other video recording instrument. Every first frame before the tracking, as mentioned in the first step, is used for the localization of the human body & the subsequent frames are provided for the tracker. In the next phase, the analysis of angles calculated from landmark points is done with the help of the pose classification algorithm. The parameter for this algorithm is the landmark points, a list of detected landmarks of the person whose activity needs to be classified. The angles are calculated with the help of the equation for each important landmark and the two connected landmark coordinates. These are compared to the set of predefined angles for each activity, and the activity is displayed as per the user's requirements.

III. MODELING AND ANALYSIS

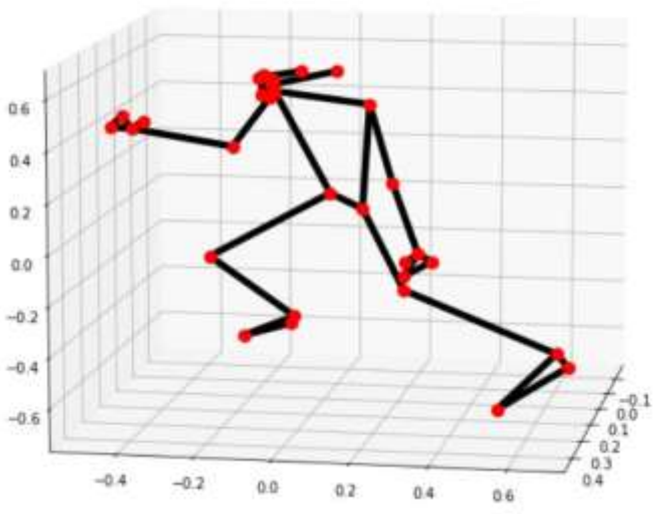


Figure 4: 3D Pose Estimation Model

The above figure shows a 3D pose model created using the hip landmark as the origin, which gives a sense of the depth of each landmark point. The 2-step detector deep neural network detector model of MediaPipe, which is used in this approach, is an innovative approach for human activity recognition. It has better performance with the use of fewer computational resources. The detection of 33 landmark points is considered the most comprehensive feature & is helpful for many applications. Considering the scope of this HAR project, the detection of angles, creating sets of their combinations & comparing them with natural human behaviour to recognize the correct human activity are the crucial steps in the approach.

IV. RESULTS AND DISCUSSION

Promising results have emerged to recognize human actions with Google MediaPipe and OpenCV. The 2-step pose detector, assisted by angle heuristics, detected, and tracked the key points with a precise accuracy that enabled the classification of activities in humans. The project was tested on a database that included various activities, e.g., walking, standing, sitting, bending, waving hands, sleeping, etc. The accuracy achieved by this approach was found to be approximately 95%. Such a high level of accuracy is very desirable as it can track and monitor people's activities in different areas such as sports training, health care, or security. In addition, the recognition speed made it possible to process videos in real-time. The potential application of computer vision and machine learning techniques to recognize individual activity has been demonstrated by the results achieved under this project.

The figures below show the results of this approach:

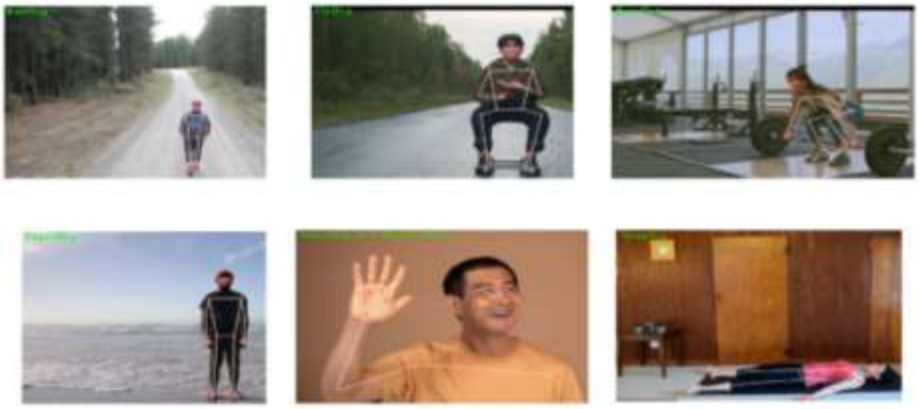


Figure 5: Activity Recognition Results

Even though the MediaPipe has performed well, still there are some areas for improvement. By increasing the variety of activities, the recognition model can be optimized.

V. CONCLUSION

Human activity recognition (HAR) has always been an area of research, and the emerging tools and technologies for HAR have the potential to accomplish its purpose with less computation time & fewer resources resulting in faster recognition with high accuracy. In this paper, we discussed one of the best approaches to HAR using the OpenCV library and Google MediaPipe. The preprocessing of video frames with OpenCV and the 2-step deep neural network detector model of MediaPipe to extract landmarks from the human body help achieve desired results. However, there are still opportunities for improvement. Further research must be conducted to establish a combination of activities and predict actions that do not show the complete body.

ACKNOWLEDGEMENTS

We thank **Prof. Nitin Shahane Sir** for his incredible contributions and advice in developing this research paper. Your dedication and expertise have been invaluable in shaping the course of our project. Astonishingly, you are passionate about and dedicated to excellence in this field. Your advice and suggestions have expanded our understanding of the subject. We thank you for sharing your expertise and knowledge with us. We are grateful for the chance to work with you.

VI. REFERENCES

[1] V. Agarwal, A. K. Rajpoot, and K. Sharma, “Ai based yoga trainer - simplifying home yoga using mediapipe and video streaming,” 2022.

[2] P. Prabu, K. Amrutha, and J. Paulose, “Human body pose estimation and applications,” 2021.

[3] D. L. Prasanna, V. Ramana, M. Tejasree, and C. Yasaswi, “Human activity recognition using opencv,” 2021.

[4] Somula Ramasubbareddy, N. Sudhakar Yadav & M. Ravikanth, “Neural Network-Based Activity Recognition System,” 2022.

[5] Nagesh U B, Abhishek V Doddagoudra, Adarsh K M, Mayoori K Bhat, Shreya L, “Human Action Recognition using Deep Learning Technique,” 2022

[6] P. Yi, Y. Cheng, R. Liu, J. Dong, D. Zhou, and Q. Zhang, “Human-robot interaction method combining human pose estimation and motion intention recognition,” 2021.

[7] Xuehao Xiang, Yongbin Gao, Naixue Xiong, Bo Huang, Hyo Jong Lee, Rad Alrifai, Xiaoyan Jiang, and Zhijun Fang “Human Action Monitoring for Healthcare Based on Deep Learning,” 2018.

[8] Lamiyah Khattar, Chinmay Kapoor, Garima Aggarwal “Analysis of Human Activity Recognition using Deep Learning,” in Proc. IEEE International Conference.

[9] Shubham Kumar, Satyam Kumar, Laxmi V, “Human Action Recognition Using Deep Learning,” 2022.