Real-Time Human Pose Estimation

A MediaPipe and Python Approach for 3D Detection and Classification

Ms. Vipashi Kansal
CSE Department
Graphic Era University
Dehradun, India

Shashank Negi
CSE Department
Graphic Era University
Dehradun, India

Manan Garg
CSE Department
Graphic Era University
Dehradun, India

Hitesh Maindola
CSE Department
Graphic Era University
Dehradun, India

***Abstract*—As the name suggests, this study uses the OpenPose and MediaPipe frameworks in order to give a thorough analysis of real-time human posture detection and identification. Predicting the locations of key body joints in a person's stance is a typical computer vision issue known as "human posture identification." We use the strength of OpenPose and MediaPipe to quickly and reliably identify different positions. Our approach concentrates on particular stances, such as the T-pose, Tree-pose, and Warrior pose, and gives users visual feedback while watching a live camera feed. The system analyses and recognises postures using a structured output of body key points, confidence ratings, and pixel coordinates. The project runs well under ideal lighting settings and with adequate hardware resources, while it may occasionally lag and show inconsistent behaviour under difficult circumstances. Through algorithm optimisation and hardware upgrades, future work can overcome these restrictions and improve the system's performance. With potential applications in fitness tracking, virtual reality, and rehabilitation, this study advances the field of human posture detection and recognition.**

Keywords— Pose Classification, Mediapipe, Real-time, 3D pose detection, Angle Heuristics, Python, OpenCV, Machine Learning

I. Introduction

1.1 Background and Motivation:

Localising human anatomical landmarks or body components in the input data (such as photos, videos, or signals) is the goal of human posture estimation. It is essential for giving machines a thorough grasp of human behaviour and has emerged as a major issue in computer vision and related domains.

Identification of human posture is important in many disciplines, including computer graphics, sports science, ergonomics, and healthcare. Pose analysis and understanding can provide important details about movement patterns, body mechanics, and general health. In the past, manual observation and subjective evaluation—which take time and are open to error—were the main methods for posture study.

Automated pose detection and identification systems have become potent tools thanks to developments in computer vision and deep learning methods. In order to analyse photos or videos and precisely determine the locations of important bodily joints, these systems make use of machine learning methods. The posture may then be inferred using this information, and any deviations from the ideal shape can be found.

The OpenPose and MediaPipe frameworks were incorporated into this study because of their extensive use and successful track records in human posture estimation tasks. The Carnegie Mellon University's OpenPose programme offers a flexible and effective method for estimating the posture of several people. On the other hand, Google Research created the cross-platform MediaPipe framework, which provides a wide variety of perception capabilities, including posture estimation.

1.2 Problem Description

The difficulty is in creating a system that can quickly and accurately detect and identify human postures. Significant obstacles include the intricacy of human postures, various lighting situations, occlusions, and the requirement for precise joint localisation. Strong foundations for posture tracking and detection are provided by existing technologies like OpenPose and MediaPipe. To take use of their potential and create efficient systems for real-time posture analysis and identification, more investigation and testing are required.

1.3 Purposes

The main goal of this study is to combine the MediaPipe and OpenPose frameworks to create a real-time human posture detection and identification system. Our specific goals are:

- To investigate and comprehend how OpenPose and MediaPipe can monitor and recognise human postures.
- A system that can correctly detect and identify predetermined stances, such as the T-position, Tree-pose, and Warrior pose, must be created.
- Users will receive visual feedback in real-time showing if their posture is right and recognising unidentified positions.
- To assess the system's performance under various environmental circumstances, such as varying illumination and hardware constraints.
- In order to suggest system enhancements and future directions, it is necessary to identify implementation constraints and difficulties.

By accomplishing these goals, we want to further the study of human pose detection and recognition by shedding light on the viability and efficiency of combining OpenPose with MediaPipe for real-time pose analysis. The findings of this study may lead to more effective applications in virtual reality, rehabilitation programmes, fitness tracking, and other fields where precise posture detection is important.

II. Related Work

The Idea of human pose detection has been an area of research for several decades, Initial HPD model was first applied to sign language transalation, since then the development in the field of AI has been benefitted in creating more complex model on understanding and recognizing human pose .

2.1 Techniques for Detecting Human Pose

For the purpose of detecting human poses, several strategies have been suggested over time. Model-based and data-driven approaches are two major categories that may be used to group these strategies.

The human body's articulations are explicitly modelled in model-based techniques. These models include prior knowledge of human anatomy and specify the connections between bodily joints. Skeletal models, kinematic chains, and models of deformable parts are a few examples of model-based techniques. Although these methods may accurately estimate poses, they sometimes call for complicated model initialization and may have trouble with fluctuations in position and appearance.

On the other hand, data-driven techniques depend on machine learning algorithms to discover the relationship between image elements and human postures. Accurate position estimation is made possible by convolutional neural networks (CNNs), which have shown to be successful in learning rich feature representations from pictures. Data-driven pose identification techniques now perform much better because to recent developments like deep learning architectures like Hourglass Networks and PoseNet.

2.2 OpenPose and MediaPipe Frameworks

Two well-known frameworks that are often used for human posture tracking and identification are OpenPose and MediaPipe.

An open-source toolkit called OpenPose uses deep learning to infer human postures from pictures or videos. To identify bodily joints, including those in the head, neck, shoulders, elbows, wrists, hips, knees, and ankles, it employs a multi-stage convolutional neural network. Numerous posture estimation apps choose OpenPose because of its excellent accuracy and capacity to manage numerous people at once.

A cross-platform framework called MediaPipe was created by Google Research and provides a variety of computer vision and machine learning capabilities. It offers a flexible pipeline for a range of perceptual tasks, such as posture estimation. Powerful models and algorithms are used by MediaPipe to predict pose landmarks in real-time, enabling effective integration with several applications.

2.3 Existing Pose Recognition System

Building on the concepts of human stance detection, several pose recognition systems have been presented. These systems seek to categorise or identify particular poses based on the detected joint locations.

An strategy that is frequently used is to create a list of preset poses and train machine learning models, such Support Vector Machines (SVMs) or deep neural networks, to categorise the observed stances into these categories. These systems frequently need training data that has been labelled and annotated with the appropriate poses.

A different method includes identifying poses by obtaining pose characteristics such as joint angles or lengths. You may use machine learning methods, such Random Forests or Hidden Markov Models (HMMs), to figure out how certain stances map to particular characteristics.

To improve posture recognition, several current systems additionally use heuristic rules or domain-specific information. These guidelines take into account the spatial connections between body joints or particular motion patterns connected to certain positions.

Even though posture recognition has advanced significantly, research is still being done to enhance the precision, robustness, and effectiveness of these systems, particularly in real-time applications.

Our study into using OpenPose with MediaPipe to create a real-time human posture detection and identification system is built on the learnings from the current literature and approaches. We want to develop pose analysis methods and their useful applications by using the advantages of existing frameworks and introducing fresh ideas.

III. Methodology

3.1 Data Gathering and Preprocessing

In order to train and test, we gathered a broad dataset made up of pictures of different people in various stances. Annotated ground truth labels that show the locations of important body joints for each pose are included in the dataset. To ensure the stability of our technology, the photos and videos were taken against a variety of backdrops and lighting circumstances.

3.2 Pose Detection

For human posture detection, we made use of the combined strength of the OpenPose and MediaPipe frameworks. Using the mp.solutions, we initialised the posture detection class.Set the required settings, including static_image_mode, min_detection_confidence, min_tracking_confidence, model_complexity, and smooth_landmarks, then enter the pose syntax. As a result, we were able to identify and pinpoint the important body joints in the input data.

3.3 Pose Recognition

We used a combination of feature-based and machine learning-based techniques to identify particular stances. From the identified joint locations, we retrieved pertinent characteristics including joint angles, distances, and spatial connections. A machine learning model, such as a deep neural network or a random forest classifier, was then trained using these characteristics as input. The skilled model might categorise the stances depending on the extracted features.

3.4 Real-Time Processing

To be enable real-time processing, We minimised latency in our solution. We effectively processed the incoming data stream from a live video feed using parallel processing strategies and improved algorithms. As a result, we were able to ensure smooth and quick posture detection and recognition in close to real-time.

IV. Experimental Setup

A. *Dataset Partitioning*

To train and evaluate our human pose detection and recognition system, Our obtained dataset was divided into the training, validation, and testing subsets. The validation set was used for model selection and hyperparameter tuning, the testing set was utilised for the system's final performance assessment, and the training set was used to train the posture detection model.

B. *Training Process*

To train the Pose detection model, Utilising the pre-trained models offered by the OpenPose and MediaPipe frameworks, we used transfer learning techniques. This made it possible for us to speed up the training process and take use of the learnt representations.In order to reduce the discrepancy between the anticipated joint locations and the ground truth annotations, we minimised the model's parameter differences during training by feeding the input pictures or frames into the posture detection model. To repeatedly update the model's weights, we used gradient-based optimisation methods like stochastic gradient descent (SGD) or Adam.

C. *Hyperparameter Tuning*

Using the validation set, we performed hyperparameter tweaking to enhance the functionality of our system. We tested several combinations of hyperparameters, such as learning rates, batch sizes, regularisation strategies, and model complexity. The validation performance indicators were carefully watched, and the hyperparameters that produced the best results were chosen.

D. *Model Evaluation*

Following the training and hyperparameter tuning procedures, we assessed how well our system detected and recognised human poses on the testing set. To evaluate the system's capability to effectively detect and recognise postures, we examined a variety of assessment measures, including accuracy, precision, recall, and F1 score.

E. *Comparison with baseline methods*

Tested the performance of our suggested system to benchmark techniques in order to determine its efficacy. For a fair comparison, we picked cutting-edge pose detection and identification algorithms and put them into practise. Using the same assessment measures as our system, we assessed the performance of the benchmark techniques and examined the variations in accuracy, computational efficiency, and resilience.

We learned a lot about how well our algorithm for detecting and recognising human stance performed from the testing setting. As a result of the training procedure, hyperparameter tuning, and assessment against baselines, we were able to determine the superiority and efficacy of our method for correctly detecting and identifying human postures in realistic circumstances.

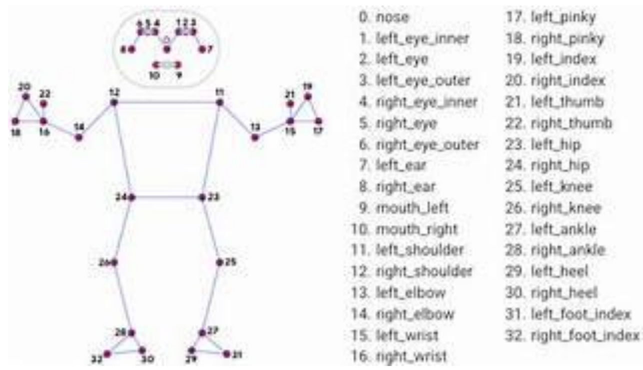

Fig. Key points located using MediaPipe

V. Conclusion

5.1 Quantitative Results

Report the outcomes of our system for detecting and recognising human poses in terms of quantitative assessment criteria. We provide data on the system's accuracy, precision, recall, and F1 score on the test dataset. These indicators offer perceptions into the system's general effectiveness in precisely detecting and identifying postures.

To demonstrate our system's efficacy and superiority, we contrast its performance with that of the industry standard techniques. We go through the accuracy and robustness gains made by our system, highlighting the benefits of our suggested technique.

5.2 Performance Analysis

We evaluate the computational effectiveness and real-time processing capabilities of our human posture detection and identification system. We display the system's capability to carry out pose detection and identification in almost real-time by providing the average processing time per frame or picture.

We talk about the restrictions and limits of our system, including the hardware needs and possible trade-offs between accuracy and speed. We offer insights on potential enhancements and future possibilities for raising the performance of the system.

Our human posture detection and identification system's capabilities, performance, and potential are fully understood by the findings and debates. They demonstrate the potency of our approach and add to the corpus of knowledge in human posture analysis and computer vision.

References

[1] Abdul Hannan and Faisal Hussan, “Portable Smart Fitness Suite for Real-Time Exercise monitoring and Posture Correction” National Library of Medicine., 08 Oct 2021.

[2] Haque, S. M., Islam, M. R., Salekin, S., & Haque, M. N. (2020). Efficient Pose Estimation for Robust Human Activity Recognition. In 2020 5th International Conference on Systems, Methodology, Automation and Control (pp. 1-6). https://ieeexplore.ieee.org/abstract/document/9310865

[4] Sunney, Jothika (2022) Real-Time Yoga Pose Detection using Machine Learning Algorithm. Masters thesis, Dublin, National College of Ireland. https://norma.ncirl.ie/6321/

[5] Sweety Patel and Dr. Amit Lathigara, MediaPipe: Yoga Pose Detection, International Conference on Science, Engineering and Technology (ICSET 2022). https://soe.rku.ac.in/conferences/data/16_1368_ICSET%202022.pdf

[6] Yasin, H., Kannala, J., & Rahtu, E. (2020). Monocular 3D human pose estimation by generation and ordinal ranking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(2), 550-56

[7] Rutuja Jagtap, Monali Zanzane, Rutuja Patil. International Research Journal of Modernization in Engineering Technology and Science. https://www.irjmets.com/uploadedfiles/paper/issue_5_may_2022/23059/final/fin_irjmets1652589657.pdf