Moving a character using body tracking

Adarsh Sahu
201IT203
Information Technology
NIT Karnataka
Surathkal, India

Kaustubh Vivek Khedkar
201IT128
Information Technology
NIT Karnataka
Surathkal, India

Vishruth M
201IT167
Information Technology
NIT Karnataka
Surathkal, India

***Abstract*—In this project, we extract the user's pose from a video feed using OpenCV and Mediapipe. With the advancements made in body pose recognition, optical motion tracking can be performed in almost all environments without any special equipment. Using the landmarks given by the pre-trained Mediapipe Pose model, the user's pose can be estimated. Using these estimations, key-presses are simulated as inputs for games**

***Index Terms*—VR, Motion Tracking, OpenCV, Mediapipe, Pynput, Pose, Game Simulation**

I. Introduction

With the recent improvements in virtual reality (VR) technology, the number of novel applications for entertainment, education, and rehabilitation has increased. The primary goal of these applications is to enhance the sense of belief that the user is "present" in the virtual environment. Motion tracking, the process of digitizing your movements for use in computer software, is incredibly important for virtual reality.

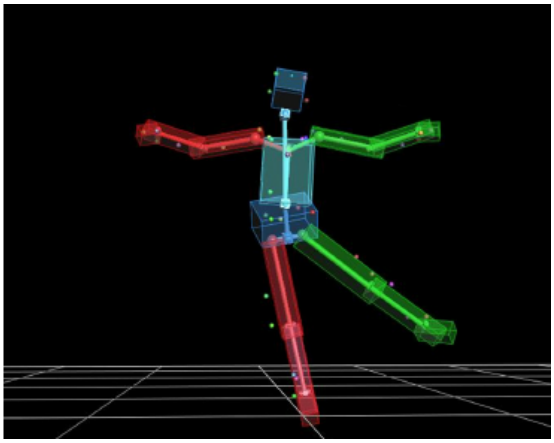

Fig. 1. Full Body Tracking

Optical methods of motion tracking usually use cameras of one sort or another. With the advancements made in body pose recognition, optical motion tracking can be performed in almost all environments without any special equipment. By tracking the user's skeleton in real-time, it is possible to synchronize the avatar's motions with the user's motions.

II. Literature Survey

The following journals and research papers were surveyed for the project.

Santosh Kumar Yadav et al. [1] built a system where they recognised the various Yoga asanas by using deep-learning algorithms. Here they used Convolutional Neural Networks (CNN) architecture in order to recognise yoga poses by using a stream of realtime images.

Matteo Rinalduzzi et al. [2] proposed a system that can be used by deaf people to communicate. Here magnetic positioning system is used for recognizing the static gestures associated with the sign language alphabet. Measured position data are then processed by a machine learning classification algorithm.

Shruthi Kothari et al. [3] proposed yoga pose detection using deep learning where in, with the help of DL and ML yoga poses are classified with the help of pre-recorded video and also in real time. The project talks about different pose estimation and methods of detection of key points in a detailed manner and explains various learning models (deep learning models) used for classification of poses.

Akshit Tayade et al. [4] proposed a system for Real-time accurate sign language and hand detection using Support Vector Machine (SVM) algorithm without any wearable sensors using Mediapipe.

S Khan et al. [5] proposed a system that can detect people with the help of multiple uncalibrated cameras. That is when a person is detected in one camera, the system is able to identify in which all cameras this same person is present. Because of the field of view of cameras, this system uses multiple cameras in order to detect a particular person.

Author	Methodology	Merits	Limitations
Arpita Haldera, Akshit Tayade	Real-time accurate detection using Support Vector Machine (SVM) algorithm without any wearable sensors makes use of this technology more comfortable and easier	Using Mediapipe and combining a ML model over it allows for accurate multi-class classification. Model can be retrained to fit other sign languages.	Does not work on a video stream. Using a ML model increases cost for very little increase in accuracy, only limited to hands.
Santosh Kumar Yadav, Amitojdeep Singh, Abhishek Gupta & Jagdish Lal Raheja	Convolutional Neural Networks (CNN) architecture to recognize yoga poses by using Images.	Addition of new poses by just retraining the model with new data.	Does not work on a video stream. CNN need big training dataset. Chances of overfitting/ underfitting.
Matteo Rinalduzzi, Alessio De Angelis, Francesco Santoni Emanuele, Buchicchio Antonio, Moschitta Paolo Carbone, Paolo Bellitti, and Mauro Serpelloni	Magnetic positioning system, wearable transmitting nodes, measures the 3D position and orientation of the fingers.	Tracking hand gestures in a 3D space. Allows flexibility and no camera calibration.	Special equipment required for operation. Higher cost of operation. Only limited to hands.
Shruti Kothari	Different pose extraction methods are then discussed along with deep learning-based models - Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)	The use of hybrid CNN and LSTM model on OpenPose data is seen to be highly effective and classifies all the yoga poses perfectly.	The performance of the models depends upon the quality of OpenPose pose estimation which may not perform well in cases of overlap between people or overlap between body parts
S. Khan; O. Javed; Z. Rasheed; M. Shah	Analyzing information from multiple cameras simultaneously.	This approach does not require feature matching, which is difficult in widely separated, uncalibrated cameras.	(The limitations of this system do not appl to our model)

III. Problem Statement

To design and implement a system to simulate directional key-press using real time pose estimating

A. Objectives

- Capturing and pre-processing a video feed.
- Extracting frames and using ML to locate body and landmarks using Mediapipes inbuilt model.
- Estimating poses and simulating key presses.

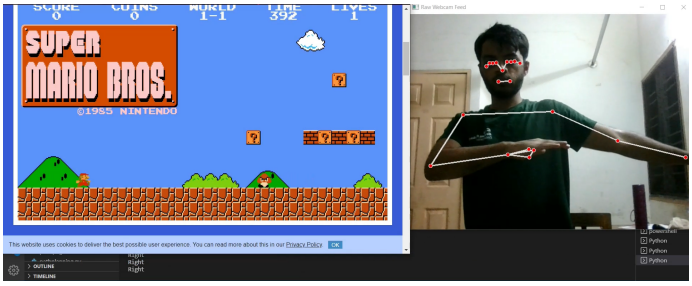

Fig. 2. Real time movement

IV. Methodology

To create a model which can move our game character around by recognizing certain poses that act as controls for character movement, we used some python libraries. These libraries allow easy implementation of video feed input and body recognition and tracking.

OpenCV is a library of python programming language and is used to capture and pre-process the video feed. The video feed is captured in BGR and colored to RGB for Mediapipe to process on the frames. This image is recoloured back to BGR for rendering as output.

For the purpose of detection of the body and tracking, the MediaPipe framework is used. MediaPipe Pose is a ML solution for high-fidelity body pose tracking, inferring 33 3D landmarks, whose output contains information about the Cartesian coordinates of the user's body assuming a 2D screen.

Using the given coordinates, the users pose can be estimated. Output is as a key-press simulation using pynput. The pynput library allows you to control your input devices such as they keyboard and mouse.

V. Result and Analysis

A. Results

- The model is successfully able to capture and preprocess the video feed.
- The movement of the user are being tracked and shown as output by Mediapipe's pipeline.
- Read certain postures of the user and provide the appropriate output such as up, left and right.
- Key presses are simulated based on pose. In-game characters can be controlled.

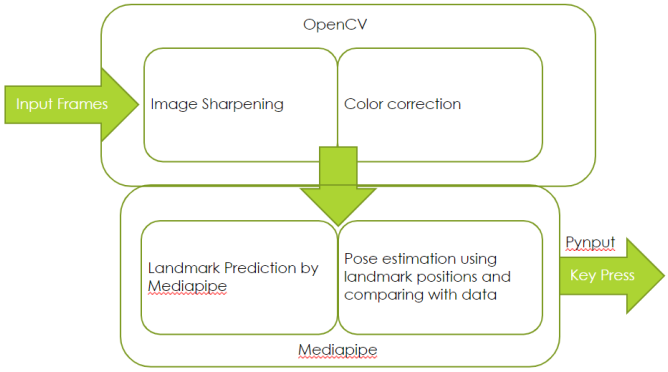

Fig. 3. Flow of System

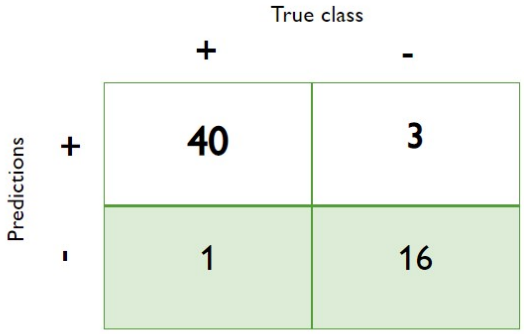

Fig. 4. Confusion matrix

B. Analysis of results

- On testing the model returns a 93.33 percent accuracy.
- The model has a precision of 93.02 percent.
- The model has a recall of 97.56 percent.
- Overall the model performed well on individual frames and also responds accurately in real time when tested on different systems.

VI. Conclusion

In this paper, we were able to build a system that is able to capture a video feed and process it to extract poses and simulate key-presses to control a game character. The model has impressive accuracy and works in real time. This sets a good baseline for further additions to the system.

VII. Individual Contribution

- Kaustubh Vivek Khedkar : Implementing the methodology as a python program. Incorporating the Mediapipe model and testing.
- Vishruth M : Research of related Previous papers, compiling report and presentation, and calculating error rate and accuracy of project model.

- Adarsh Sahu : Research of related articles and previous papers. Incorporating OpenCV code snippets. Finalizing report.

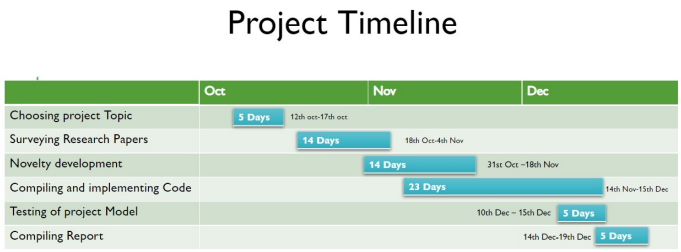

Fig. 5. Gantt Chart

VIII. Base Paper

Ardra Anilkumar, Athulya K.T., Sarath Sajan, Sreeja K.A. -Pose Estimated Yoga Monitoring System; 2nd International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS 2021); https://papers.ssrn.com/sol3/papers.cfm?abstractid=3882498

IX. References

[1] Santosh Kumar Yadav, —Real-time Yoga recognition using deep learning‖; 20 May 2019; https://doi.org/10.1007/s00521-019-04232-7;

[2] Matteo Rinalduzzi, -Gesture Recognition of Sign Language Alphabet Using a Magnetic Positioning System; 17 June 2021; https://www.mdpi.com/2076-3417/11/12/5594

[3] Shruthi Kothari, —Yoga Pose Classification Using Deep Learning‖, May 2020; https://doi.org/10.31979/etd.rkgu-pc9k

[4] Arpita Haldera, Akshit Tayadeb - Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning; https://www.ijrpr.com/uploads/V2ISSUE5/IJRPR462.pdf

[5] S. Kan; O Javed, Z. Raheed; M. Shah, —Human tracking in multiple cameras‖,7 July 2001, https://doi.org/10.1109/ICCV.2001.937537