

A project Report on Pose Estimation



CA 2

INT248

Name:- Subrojit Roy

Reg. no.:- 11801400

Roll no.:- RKM073A06

Introduction

Pose estimation is a computer vision technique to track the movements of a person or an object. This is usually performed by finding the location of key points for the given objects. Based on these key points we can compare various movements and postures and draw insights. Pose estimation is actively used in the field of augmented reality, animation, gaming, and robotics.

Media pipe is an open-source cross-platform framework for building multi-model machine learning pipelines. It can be used to implement cutting-edge models like human face detection, multi-hand tracking, hair segmentation, object detection and tracking, and so on. I have used Media Pipe and Open CV to make this project.

Literature Review

Humans can recognize the pose by looking at the location and positions of different parts of the human body. The same principle is applied in computer vision as the problem of HPE is defined as a problem of localization of human joints. Human body varies from simple to complex poses. Finding location of body parts and joints from a single image is not always a simple task because of variety of different poses as well as some external conditions like light, clothes, multiple people in a single picture etc. and estimating these different poses needs some systematic processes. Therefore, it has been considered as an interesting topic by researchers. Fischler and Elschlager used the concept of part based modelling for facial structure

estimation in the well-known PS model in 1973 . A general method consisting of two main modules; identification of parts and configuration of parts to form a structure by discretised the search space and use of dynamic programming. PS framework is a general model because it neither depends on the training scheme of the individual parts, nor on the type of connections between them. The base of this method is representation of objects by a set of parts organized in a deformable structure. P.F. Felzenszwalb and D.P. Huttenlocher made the PS more tractable and practical by using the distance transforms for object recognition.

Discussion:-

In this project I have used two libraries OpenCV and Media pipe which is available in python and are very powerful. What is media pipe is discussed above now coming to the OpenCV part. OpenCV is a great tool for image processing and performing computer vision tasks. It is an open-source library that can be used to perform tasks like face detection, objection tracking, landmark detection, and much more. The library is equipped with hundreds of useful functions and algorithms, which are all freely available to us. Some of these functions are really common and are used in almost every computer vision task.

In this project there are two files one is the hand tracking module which basically tracks the hand. There are 21 unique points in our hands which we can detect through the help of

media pipe. The project is of pose estimation so I had to add something so that we can really say that the computer had detected something and the program is useful. That is when the second file comes into play. Basically the first file tracks the hand and the unique points in our hands like here I have used the tip of two fingers, One is of the index finger and the other is of the thumb. The second file do the volume control part. First let me discuss what the program do then I will explain the working. Using this program we can control the volume of the device. We just need to run the program and when we use our two finger that is the index and the thumb and if we keep our fingers closer then it decrease the volume automatically we don't need to click a physical button and if we take the fingers far from each other it will increase the volume and the volume will change accordingly.

Now, the question is how it is working. As I had mentioned earlier that the program first detects the points in our hand so I have make the program in such a way that it will focus on the tip of the index and the middle finger and it will calculate the distance between those two tips. When the distance is more it means that we want to increase the volume so the program will increase the level of the volume in the system and vice versa.

Conclusion:-

This type of computer vision model are very important and very useful for the future generation. From this project, image a world where we can control numerous things of our devices just

using gestures, that will be great right. We can use computer vision in many fields such as military, health, etc.

Screenshots:-



