Human Pose Detection using Deep Learning

Deepak Maurya[1], Grandel Dsouza[2], Anoop Patel[3]

[1,2,3] *Student, Dept. of Information Technology, Xavier Institute of Engineering, Mahim, Mumbai, Maharashtra, India.*

Abstract - *Human pose detection is a crucial problem within the field of Computer Vision. Imagine having the ability to trace a person's every small movement and do a biomechanical analysis in real-time. This technology has a huge implication both now and in the future. Applications based on Human pose detection may including video surveillance, assisted living, healthy lifestyle and sports analysis. Formally speaking, Pose Estimation or detection is predicting the part or joint positions of an individual from a picture or a video. This technology is made possible by combining two major computer technologies i.e. Artificial Intelligence (A.I.) and Computer Graphics*

Key Words: Human pose detection, Computer vision, biomechanical analysis, video surveillance, assisted living, sport analysis, Artificial intelligence, computer graphics.

1. INTRODUCTION

Human body pose estimation or detection in computer vision/graphics is that the study of algorithms, systems, and pre-trained models that recover the pose of an articulated body, which consists of joints and rigid parts using image-based observations. it's one of the longest-lasting prevalent problems in computer vision the reason being the complexity of the models that relate observation with the pose, and since of the variability of situations during which it'd be useful.

Perception of the citizens in their neighbouring environment is a crucial capability that robots must possess. If an individual uses gestures to point to a specific object, then the interacting machine should be ready to understand things in the world context. Thus, pose estimation is a crucial and challenging problem in computer vision, and lots of algorithms are deployed in solving this problem over the last 25 years. Many solutions involve training complex models with large sample data sets and hence improvising and progressing further.

Pose estimation may be a difficult problem and a lively subject of research because the physical body has 244 degrees of freedom with 230 joints. Although not all movements between joints are evident, the physical body consists of 10 large parts with 20 degrees of freedom. Algorithms must account for giant variability introduced by differences in appearance thanks to clothing, body shape, size, and hairstyles. Additionally, the results could also be ambiguous thanks to partial occlusions from self-articulation, like an individual's hand covering their face, or occlusions from external objects. Finally, most algorithms estimate pose from monocular (two-dimensional) images, taken from a traditional camera. Other issues include varying lighting and camera configurations. The challenges are made even more complex to match up to additional performance requirements. These images lack the three-dimensional (3-D) information of an actual body pose, resulting in further ambiguities. there's recent add this area wherein images from RGBD cameras provide information about colour and depth.

2. How exactly did we achieve human pose detection using deep learning?

Firstly, to implement Human pose detection a computer system must precisely distinguish between a human being and the surrounding objects and the system must recognize the different body parts and differ each one from the other without any faults. This is where deep learning comes into the picture. What exactly is done is that the computer system is trained with different sample data in this context the system will be trained with different images like head, shoulders, arms etc. The training is performed using the Neural network concept of machine learning like CNN (Convolutional neural network) and Deep learning. Thus, different images of the different body parts and at different angles are fed on to the system and thus a trained model is achieved. Here it must be noted that the more sample data set the system is trained with the better performance the system shall yield. Also every body part here in the trained model will have a unique identity co-ordinate. This trained model is termed as COCO (common objects in context) model and MPII model. Functionalities of these models are: a) Image classification b) Object localization c) Semantic segmentation d) Segmenting individual object instances.

Upon achieving the above models, we now have an intelligent system that can identify and distinguish a human body from the surrounding. Now is where the computer graphics comes into picture. For computer graphics we have used OpenCV which we have used as a python library for image processing as well as video processing. Also, we have use Flask for designing a website (i.e. for Graphical User Interface)

3. Implementation Methodology

We approached the project in seven phases which are mentioned below:

1. Firstly, we downloaded the model weights
2. Then we loaded it into the network
3. Read video as a set of frames which we took as an input to the network
4. Predictions were made and key points were parsed
5. By using the key points, the skeletal structure was drawn
6. The skeletal structure gave the various angles that were being made
7. Finally using the angles, we got a result wherein certain instructions were given and then we plotted the graph

4. Details of Hardware and Software

Hardware:

a. CPU: 4 core processor

b. RAM: 4 GB

Software:

i. Python 3.6

ii. OpenCV

iii. Visual Studio Code

iv. Operating System: Windows 10

5. Results

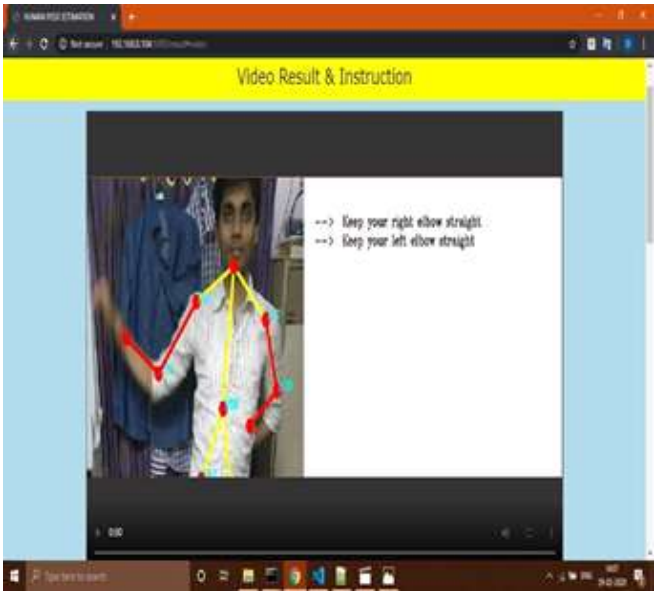

Figure-1: Skeletal structure of a human being and instructions corresponding to it

Here we can see the skeletal structure of a human being. Whenever a human being is doing certain exercises, there are predefined angles at which the movements should take place. Based on one's action, one would get the instructions. If one is not doing the movements properly, then one would be cautioned about that so that he can make the required improvements.

Figure-2: Graph describing the movements in the skeletal structure

6. Acknowledgement

We express our sincere gratitude to our Honorable Principal of Xavier institute of Engineering Mahim, Mumbai Dr. Y.D.Venkatesh for encouragement and facilities provided to us. We would like to place on record our deep sense of gratitude to Prof. Chhaya Narvekar, Head of Dept Of Information Technology, Xavier Institute of Engineering, Mahim, Mumbai, for her generous guidance help and useful suggestions. We would also like to thank our entire Information Technology staff who have willingly cooperated with us in resolving our queries and providing us all the required facilities on time.

7. Conclusion

We designed this project with an aim in mind to develop a low-cost software model for identifying the objects and so that we can carry out estimation of motion analysis of the objects and after trying various approaches we have successfully accomplished it.

8. References

[1] A survey on Model based approaches for 2D and 3D Visual Human Pose Recovery Xavier Perez-Sala 1, Sergio Escalera 2, Cecilio Angulo 3 and Jordi Gonz`alez

[2] Integrated recognition, localization, and detection using convolutional networks. L. Pishchulin, M. Andriluka, P. Gehler, and B. Schiele

[3] 2017 IEEE 12th International Conference on Automatic Face & Gesture Recognition Realtime Multiperson 2D Pose Estimation using Part Affinity Fields Zhe Cao Tomas Simon Shih-En Wei Yaser Sheikh, The Robotics Institute, Carnegie Mellon University

[4] Articulated body pose estimation - Wikipedia.html

[5] https://www.learnopencv.com/de ep-learning-based-human-pose- estimation- using -opencv -cpp- python/