Human Activity Recognition using Pose Estimation

Name: Onkar Pednekar Student ID: 801275237

Image Source

Problem Statement

A Human Activity Recognition (HAR) system's goal is to predict a person's action label from an image or video. Pose information is used by one of the most popular vision-based HAR systems. Poses reveal important information about human behavior. Using the BlazePose with MediaPipe library, we extract human poses (32 body keypoint locations in a two-dimensional plane) from images. Finally, using the pose information, activity is classified using a mathematical algorithm. We would be classifying various activities like Standing, Falling and Moving using the keypoint values for various body locations.

Motivation

- I have always wondered how the human activites can be detected through camera surveillance.
- Activity recognition is the basis for the development of many potential applications in health, wellness, or sports.
- Monitor Health: Analyze the activity of a person from the information collected by different devices. We can detect if the person is healthy based on the way he/she is walking or running.
- Fall Detection: We can use HAR to check if a patient has fallen from bed.
- Gym Reps Counting: Using keypoint location and activity we can count the number of reps done by a person/ number of pushups done by a person.
- Vast number of advantages of HAR in various domains aroused my interest in pursuing this topic.

Dataset

- We would not be using any dataset, since we already have a pretrained model provided by Mediapipe for detecting the Keypoint Locations.
- The dataset used by BlazePose implementation on MediaPipe is not publicly available, hence no link to provide about the dataset.
- For detecting the activity we would be using algorithmic approach to find the activity based on keypoint location along X and Y axis.
- Hence no dataset would be required for this task.

Summary of Method

- The selected method, BlazePose, gives us keypoint locations for various body parts. Using this keypoints we can then find the activity of the person.
- Currently I'm thinking of implementing three activities, ie Standing, Moving and Sleeping(Falling)
- Standing: For this activity, we compare the keypoint locations of consecutive frames and if the difference between them is not greater than a threshold, ie it is less, then the activity is categorized as Standing.
- Moving: We compare the keypoint locations of consecutive frames and if the difference between them is greater than a threshold, ie it is large, then the activity is categorized as Moving.
- Sleeping: We compare the keypoint location of head along the Y-axis. If it is larger than a certain Y- axis threshold, ie it is on the lower end of the frame, then it means the person has fallen down and the activity is categorized as Falling.

Research Paper

- Title: BlazePose: On-device Real-time Body Pose tracking
- Authors: Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, Matthias Grundmann
- Year: 2020
- Journal: IEEE (Found on ARXIV.org)
- Paper Link: BlazePose
- Implementation using Mediapipe: Link

Related Work

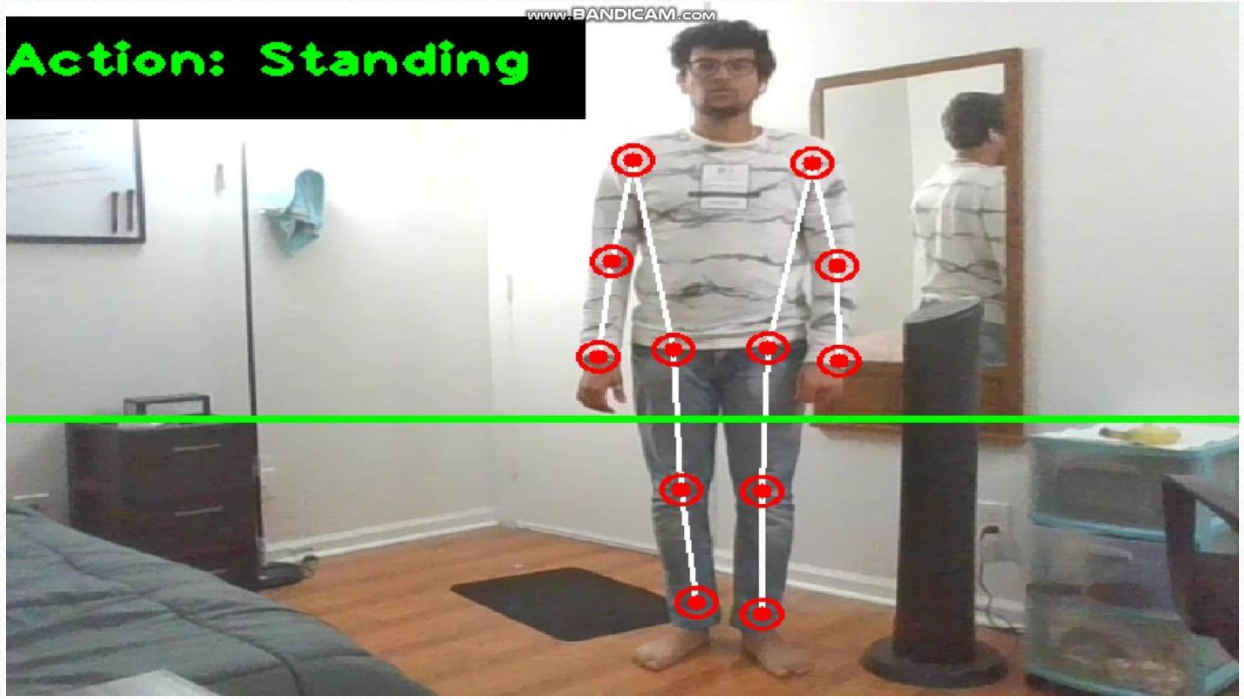
- Title: OpenPose- Realtime Multi-Person 2D Pose Estimation
- Author: Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, Yaser Sheikh
- Year: 2018
- Journal Name: IEEE (Found on ARXIV.org)
- This paper implements the same task of Pose Estimation using a different approach and neural network architecture.
- Paper Link: OpenPose

Related Work

- Title: DeepPose: Human Pose Estimation via Deep Neural Networks
- Author: Alexander Toshev, Christian Szegedy
- Year: 2013-2014
- Journal Name: IEEE (Found on ARXIV.org)
- This paper implements the Pose Estimation problem using Deep Neural Network.
- Paper Link: DeepPose

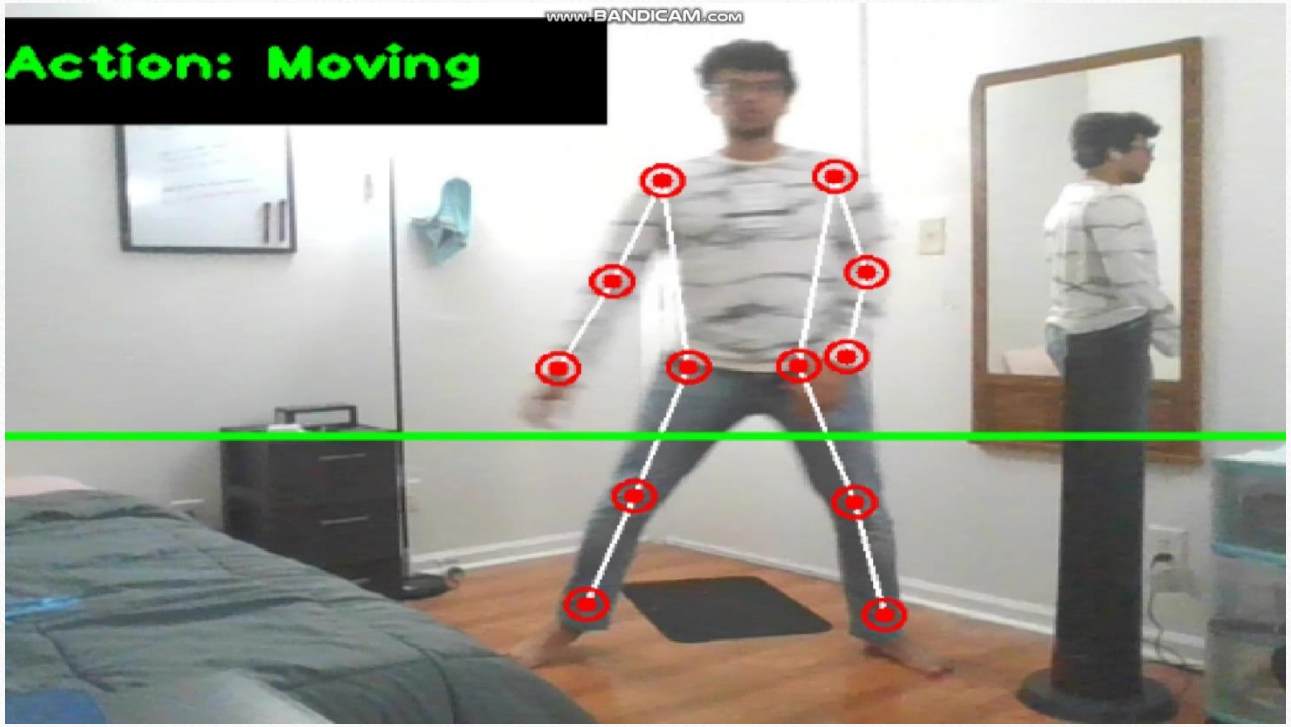
Results and Observation

Action: Standing



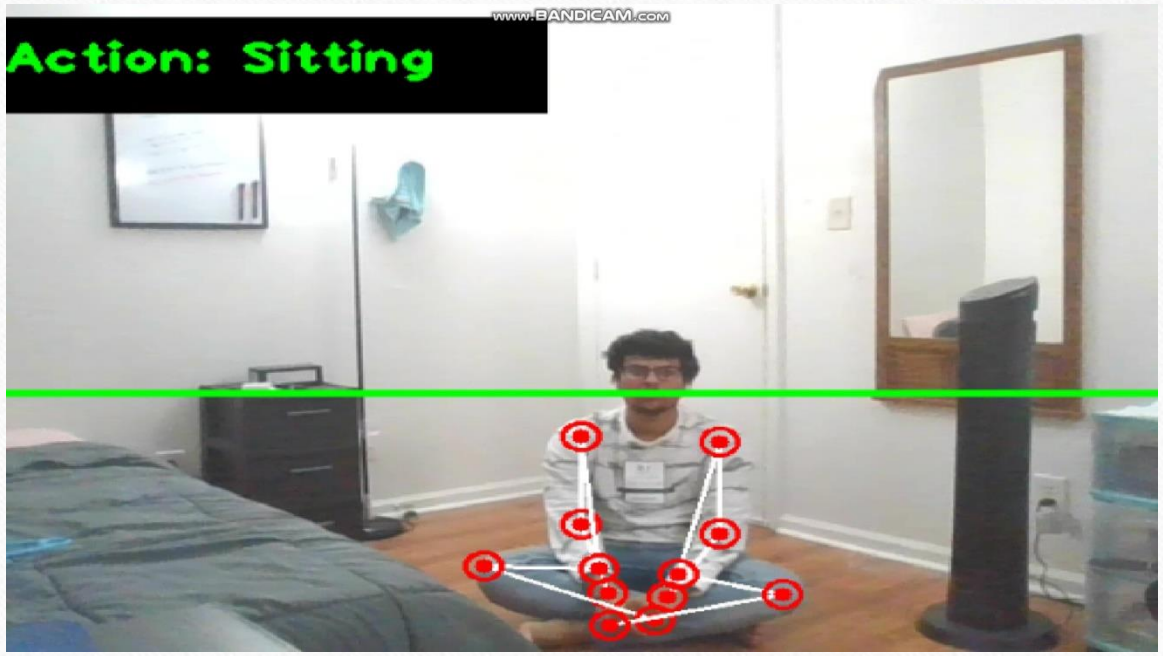
Results and Observation

Action: Moving



Results and Observation

Action: Sitting



Conclusion

- So to conclude, I will be using BlazePose to find the keypoints of various body part locations with respect to the frame.
- Based on these keypoints, I would be detecting the activity using a complex mathematical algorithm.
- The key learning of this project would be using transfer learning, or using pretrained models for our own use case and performing Pose Estimation on a real time video feed.