MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

Fifth Semester
BTech in CSE (AI & ML)
Department of Computer Science &Engineering

[Jul – Nov 2023]

Computer Vision Lab Project Guidelines
CSE 3181

Pose-Based Human Activity Classifier using OpenCV and MediaPipe

A report on
Computer Vision Lab Project
[CSE-3181]

Submitted By

Siddharth S Pai - 210962002
Om Bhayde - 210962032
Jenisha Melba - 210962031

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MANIPAL INSTITUTE OF TECHNOLOGY,
MANIPAL ACADEMY OF HIGHER EDUCATION
NOVEMBER 2023

Pose-Based Human Activity Classifier using OpenCV and MediaPipe

Siddharth S Pai[1], Om Bhayde[2], Jenisha Melba[3]

[1] Department of Computer Science &Engineering, Manipal Institute of Technology, India

[2] Department of Computer Science &Engineering, Manipal Institute of Technology, India

[3] Department of Computer Science &Engineering, Manipal Institute of Technology, India

[1st] Siddharth.t@learner.manipal.edu , [2]om.mcisma@learner.manipal.edu,
[3]jenisha.mendonca@learner.manipal.edu

Abstract— With the world population being more than 8.1 billion, Humans are everywhere. A checkpoint to all the human activities is more of a necessity than ever, with AI being the blooming sector with nearest cognisance to human brain, using AI to keep a track of human beings is the best way to follow up on things. Firstly in this report you will observe the literature review of object/human detection followed by human tracking and finally recognize the activity of a human being mostly done with the help of MediaPipe. Secondly discussion about the algorithms used in MediaPipe and how it makes a good library for Human activity recognition. The project's significance lies in its potential to revolutionize the Activity recognition sector, which may be used in real life situations such as Border Control [mainly in rough terrain], in High security Facilities, Industrial application, Anomaly Detection are few of its usages.

Keywords— "Human Detection", "Human Tracking", "Activity Recognition", "Mediapipe", "Automation"

I. INTRODUCTION

The effects of an increasing population are multifaceted and impact various aspects of society, the environment, and economies. As the global population continues to rise, several challenges and opportunities emerge. Urbanization is notable effect. Rising populations often result in rapid urban growth, leading to the development of megacities, with increase in the flux of population, challenges tend to get steeper and more difficult, Mainly safety of the society, people lurking in the unknown location tend to cause havoc and might endanger the lives of each other. So, Detecting Human presence, tracking a human being and recognizing the activity / movements becomes a mammoth task to be undertaken. This is where cognitive artificial intelligence comes into play, the recent models are trained to detect a human being, track, and recognize the activity and take the necessary measures needed. To do this we are using a pre-trained Python Library, MediaPipe which was developed by google which act as a base framework for the above tasks.

II. LITERATURE REVIEW

[1] FACE DETECTION AND RECOGNITION USING OPENCV

This research paper discuss methods used to derive Face detection. Face detection and picture or video recognition is a popular subject of research on biometrics. Face recognition in a real-time setting has an exciting area and a rapidly growing challenge. Framework for the use of face recognition application authentication. Current available face algorithms are Haarcascade_frontalface_default.xml, Eigenfaces, Fisher(Placeholder3)faces, Local Binary Patterns Histograms [LBPH].

As a part of detection algorithms they use Principal Component Analysis (PCA), PCA is used for calculating, reducing duplication, extracting functions, compression performance, etc. Because PCA is also a traditional linear domain technique, the method is suited to linear models. The PCA definition is to depict the large 1-D vector of pixels formed out of the two-dimensional face picture in the main components of the feature space. This is known as self-space projection.

In the last 20 years, facial recognition technology has come a long way. Today can check identity information automatically regarding safe transactions, tracking, security purposes and buildings access control.

[2] OBJECT DETECTION

Object detection usually consists of different subtasks such as face detection, pedestrian detection, and skeleton detection. As one of the fundamental computer vision problems, object detection can provide valuable information for semantic understanding of images and videos, and is related to many applications, including image classification, human behavior analysis, face recognition etc. The pipeline of object detection models can be divided into three stages:

1. Informative region selection: To scan the whole image with a multi-scale sliding window.
2. Feature extraction: To recognize different objects we use SIFT, HOG and Haar like feature detections.
3. Classification: A classifier is needed to distinguish a target object from all the other categories using techniques like SVM,Adaboost,and DPM.

The frameworks of object detection methods are of 2 types. One follows traditional object detection pipeline, generating region proposals at first and then classifying each proposal into different object categories. The region proposal-based methods include R-CNN, SPP-net, Fast R-CNN, FPN and Mask R-CNN. The other detection technique is a regression or classification. The regression methods include MultiBox , AttentionNet, G-CNN, YOLO, SSD, YOLOv2, DSSD and DSOD.

This paper provides a detailed review on object detection frameworks which handle different sub-problems, such as occlusion, clutter, and low resolution. . Finally, the paper proposes several promising future directions to gain a thorough understanding of object detections.

[3] Single Object Tracking: A Survey of Methods, Datasets, and Evaluation Metrics

Object tracking is one of the foremost assignments in computer vision that has numerous commonsense applications such as traffic monitoring, robotics, autonomous vehicle tracking, and so on. There are two types of object tracking: single object tracking and multi object tracking.

Single object tracking only tracks an individual target during a video, and the target is specified in the first frame and must be detected and tracked in the next frames of the video. In multi object tracking there are multiple objects to track. In this type of tracking, the tracker must first determine the number of objects. The methodology used over here is:

1. Optical Flow

Optical flow is the apparent motion of brightness patterns in the image. apparent motion can be caused by lighting changes without any actual motion.

2.Segmentation

Segmenting foreground objects from a video frame is fundamental and the most critical step in visual tracking. Foreground segmentation is done to separate foreground objects from the background scene.

3. Estimation Based Method

Estimation method formulates the tracking problem to an estimation problem in which an object is represented by a state vector.

So,in this paper, a comprehensive classification of object tracking algorithms is presented. In this category, tracking algorithms are divided into feature-based, segmentation-based, estimation-based, and learning-based categories.

[4] Accurate, Robust, and Flexible Real-time Hand Tracking

The human hand is remarkably dexterous, capable of high- bandwidth communication such as typing and sign language. Computer interfaces based on the human hand have so far been limited in their ability to track the detailed articulated motion of a user's hand accurately and reliably in real time. We believe that, if these limitations can be lifted, hand tracking will become a foundational interaction technology for a wide range of applications including immersive virtual reality, as- assistive technologies, robotics, home automation, and gaming.

Each input depth image is processed by a pipeline comprising three steps .

1. Hand RoI extraction: Identify a square region of interest (RoI) around the hand and segment the hand from background.

2. Reinitialization: Infer a hierarchical distribution over hand poses with a layered discriminative model applied to the RoI.

3. Model fitting: Optimize a 'population' of hand pose hypotheses ('particles') using a stochastic optimizer based on particle swarm optimization (PSO).

The evaluation demonstrated not only highly accurate hand pose estimates, but also dramatic improvements over the state of the art in robustness, recovering quickly from failure and tracking reliably over extended sequences, and flexibility, working for arbitrary global hand poses, at extreme distances from the camera, and for both static and moving cameras.

[5] Hand gesture recognition on python and opencv

In a day-to-day life, hand gesture recognition is one of the system that can detect the gesture of hand in a real time video. The gesture of hand is classify within a certain area of interest. Designing a system for hand gesture recognition is one of the goal of achieving the objectives of this project. The task of recognizing hand gestures is one of the main and important issues in computer vision. hand gesture recognition generally involves multiple levels such as image acquisition, pre-processing, feature extraction and gesture recognition. Image acquisition involvecapturing image in the video frame by frame using a webcam.

The captured images go through the image pre-processing process which involves color filtering, smoothing and thresholding. The list of objectives that will need to be achieve for this project:

(1) to establish a complete system for detecting, recognizing and interpreting hand gesture recognition
(2) through computer vision using Pythonand OpenCV, and (2) to create the numbers and sign languages of hand gesture shown in the systemthat will meets the name of the project.

The authors used Otsu method for Hand segmentation, Track Gesture Of Hand Using Haar-Cascade Classifier which is used to detect hand while not at motion.

Unique concept of Dual Region of interest (ROI) where one is active region while the region is static or not movable outside the frame of the realtime video and it highlights the area that involves.

This papers & source code reveals the number of fingers pointing towards the screen using the number of convex hull present in the frame.

[6] BlazePose: On-device Real-time Body Pose tracking

A comprehensive technical guide on how to extract human body keypoints using BlazePose for Pose Estimation .BlazePose is a model that extracts body keypoints from a single image. It exactly infers 33, 2D landmarks of a human body from a single frame such as shoulders, elbows, and knees as illustrated in the following figure. The user's face must be in the image to detect the pose. To have the best results, the person's entire body should be in the image.

Uses:

Its edge is that it is more suitable for applications like fitness ,rehabilitation and dance than already existing models. Why? It is more accurate. It also localizes more keypoints than other previous models do to be more suitable for applications where the scale and orientation of hands and feet are vital information.

Working:

BlazePose consists of two machine learning models: a Detector and an Estimator. The Detector cuts out the human region from the input image, while the Estimator takes a 256x256 resolution image of the detected person as input and outputs the keypoints.

III. METHODOLOGY

Methodology for Hand Gesture Recognition using Mediapipe:

1. Data Collection:

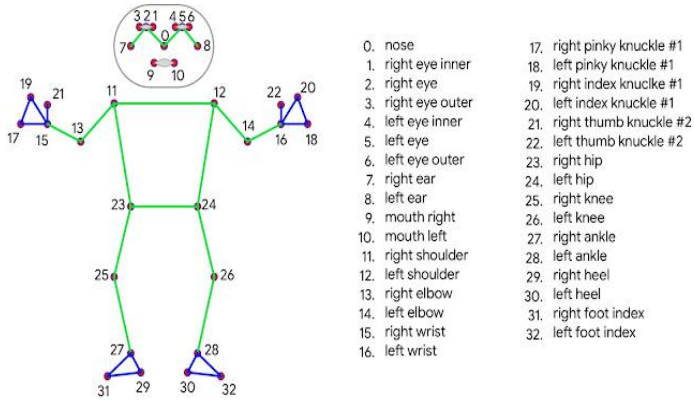
1.1 Dataset Selection:Identify or create a dataset suitable for hand gesture recognition. Consider diverse hand poses and lighting conditions.
1.2 Common Objects in Context (COCO) Dataset: COCO Dataset
1.3 Labeled Faces in the Wild (LFW) Dataset: LFW Dataset
1.4 Webcam/ camera / recorded video can also be used for this purpose.

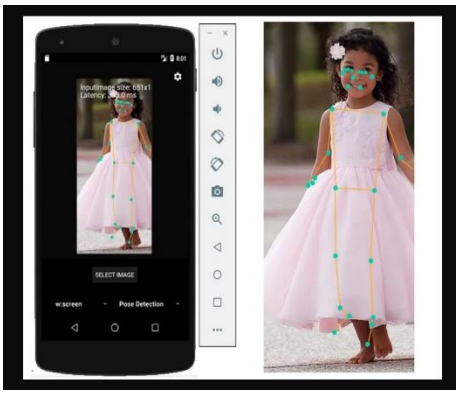
2. Pre-processing:

2.1 Data Cleaning:Remove any noisy or irrelevant data points from the dataset.
2.2 Image Resizing:Resize images to a consistent resolution suitable for input to the Mediapipe framework.

3. Hand Landmark Detection using Mediapipe:

3.1 Framework Integration:Integrate the Mediapipe framework into the project.
3.2 Hand Landmark Model: Utilize the hand landmark model provided by Mediapipe for detecting key points on the hand.

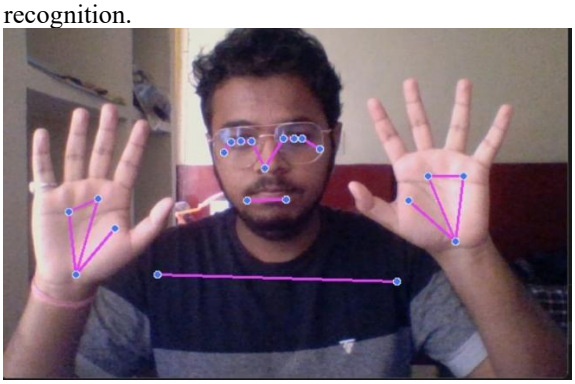




Extract 2D or 3D coordinates of hand landmarks for each frame.

4. Feature Extraction:

4.1 Hand Region of Interest (ROI):Define a region around the hand landmarks to focus on for gesture recognition.



Extract temporal features by considering the sequence of hand landmarks over time.

5. Gesture Recognition Model:

5.1 Model Selection: Choose a suitable machine learning or deep learning model for gesture recognition.

5.2 Training: Train the model using the pre-processed data, incorporating hand landmarks and temporal features.

5.3 Validation and Hyperparameter Tuning

Validate the model's performance using a separate validation set and fine-tune hyperparameters if necessary.

6. Evaluation Metrics:

6.1 Performance Metrics:

7. Future Work:

7.1 Enhancements:

Propose potential enhancements or modifications to improve the model's accuracy or real-time performance.

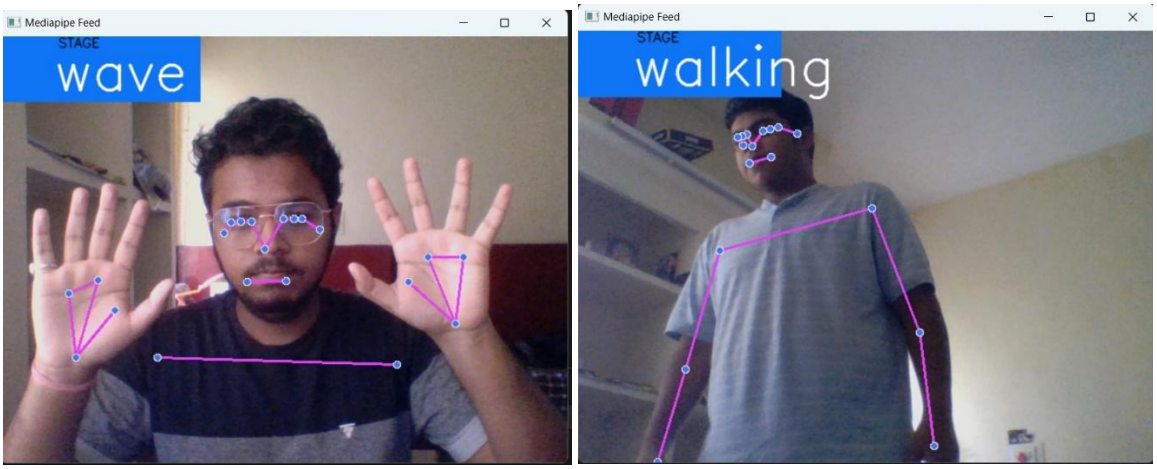
IV. EXPERIMENTAL SETUP

Our research primarily relies on smooth working of mediapipe with OpenCV , for this we had to install Python on a device, then install necessary libraries such as OpenCV, Numpy, Pandas, MediaPipe, Etc.

Once the necessary libraries are installed open python IDE, run the code for MediaPipe and as a part of test subject any human (Team Member) can walk & wave infront of the camera. You may save the result video or take necessary screenshots to support the documents.

V. RESULTS AND DISCUSSION

we achieved the target of detecting a human being , tracking a human being , and recognizing the activity the human is currently doing.



VI. CONCLUSIONS

The version of Our project successfully fulfed all the necessary objective and executes the code in a optimized manner. With the open-source framework such as MediaPipe working along with OpenCV we were able to Detect a Human being, track the movements of the Human being and we were able to achieve Activity recognition in real-time, hence completing all the Objectives of our project and paper.

VII. FUTUREWORK

1. Implement the same on a android & windows based platform which can easily be used as open-source Applications.
2. Create a CNN model for the same with better detection and accuracy.
3. Train a military grade model which can be used in used in the border petrol/security.

VIII. ACKNOWLEDGEMENT

We thank Manipal Institute of Technology for encouraging us to do this work.

IX. REFERENCES

[1] M. Khan, S. Chakraborty, R. Astya and S. Khepra, "Face Detection and Recognition Using OpenCV," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2019, pp. 116-119, doi: 10.1109/ICCCIS48478.2019.8974493.

[2] A. B. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," in IEEE Access, vol. 11, pp. 35479-35516, 2023, doi: 10.1109/ACCESS.2023.3266093..

[3] Single Object Tracking: A Survey of Methods, Datasets, and Evaluation Metrics
Zahra Soleimanitaleba, Mohammad Ali KeyvanradFaculty of Electrical & Computer Engineering Malek Ashtar University of Technology, Iran

[4] Accurate, Robust, and Flexible Real-time Hand Tracking Toby Sharp† David Kim Daniel Freedman
Cem Keskin† Duncan Robertson† Jonathan Taylor† Christoph Rhemann Ido Leichter Alon Vinnikov Pushmeet Kohli Eyal Krupka Andrew Fitzgibb⋆M.

[5] Ismail, A.P. & Aziz, Farah & Kasim, Nazirah & Daud, Kamarulazhar. (2021). Hand gesture recognition on python and opencv. IOP Conference Series: Materials Science and Engineering. 1045. 012043. 10.1088/1757-899X/1045/1/012043.

[6] BlazePose: On-device Real-time Body Pose trackingValentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, Matthias Grundmann https://doi.org/10.48550/arXiv.2006.10204