INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 3 - V7I3-2223)
Available online at: https://www.ijariit.com

AI Body Language Decoder using MediaPipe and Python

Sankeerthana Rajan Karem
sankeerthanakarem@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Sai Prathyusha Kanisetti
saiprathyusha9121@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Dr. K. Soumya
soumyacf@andhrauniversity.edu.in
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

J. Sri Gayathri Seelamanthula
sjsgayathri99@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Madhurima Kalivarapu
madhurimak.123@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

ABSTRACT

Body language are visual languages produced by the movement of the hands, face and body. In this project we evaluate representations based on skeleton poses, as these are explainable, person-independent, privacy-preserving, low-Dimensional representations. Basically, skeletal representations generalize over an individual's appearance and background, allowing us to focus on the recognition of motion. We present a real-time on-device body tracking pipeline that predicts hand skeleton and the whole-body notion. It is implemented via MediaPipe, a framework for building cross-platform ML solutions. We perform using pose estimation systems and analyze the applicability of the estimation systems to body language recognition by evaluating failure cases of the existing models. The proposed system and architecture demonstrate real-time inference and high prediction quality.

Keywords— Body Landmarks, MediaPipe, Body Language, Prediction, Accuracy, Real-time on-device Tracking, Pose Estimation, Recognition.

1. INTRODUCTION

Body Language Decoder helps detect and predict facial expressions, hand gestures and body pose. Facial expression recognition can help market research companies scale data and analyze quickly. The ability to perceive the shape and motion of hands can be a vital component in improving the user experience across a variety of technological domains and platforms truly. For example, it can form the basis for sign language understanding and hand posture control, and can also enable the overlay of digital content and information on top of the physical world in augmented reality.

The MediaPipe Body Landmark model gives for high-fidelity body pose tracking, inferring 33 2D landmarks on the body (or 25 upper-body landmarks) from RGB video frames, utilizing BlazePose. It detects the landmarks of a every single body pose, full-body by default, but it can be configured to cover the upper-body only, in such case it only predicts the first 25 landmarks.

MediaPipe Hands is a high-fidelity hand and finger tracking solution. In Just a single frame in can infer up to 21 3D hand Landmarks. It's a hybrid between a palm/hand detection model that operates on the full image and returns an oriented hand bounding box and a hand landmark model that operates on the image that is cropped region which is defined by the palm detector, which returns high-fidelity 3D hand key points. It detects landmarks of a single hand or both hands depending on the module type.

MediaPipe Face Mesh estimates 468 3D face landmarks accurately in real-time on-tracking mobile devices. It employs deep neural networks to infer the 3D surface geometry, requiring only a single camera input, without the need for a dedicated depth sensor.

Applications: Driver drowsiness detection, Sign language detection, Market research companies can use this technology to analyze data, Body language detection in interviews.

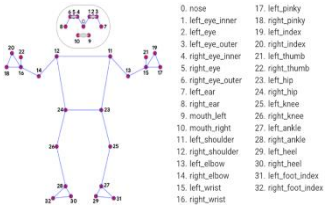

Fig 1.1 pose landmark

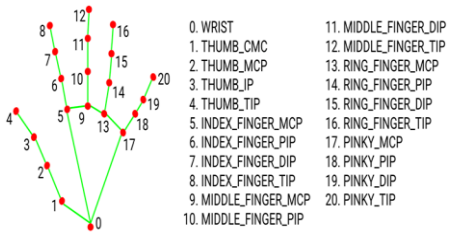

Fig 1.2 hand landmark

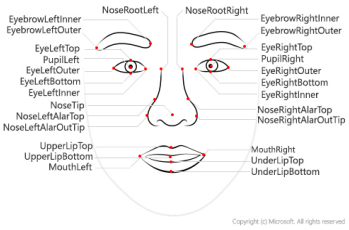

Fig 1.3 face landmark

Sentiment analysis is the process of detecting positive or negative sentiment in a sentence. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers. Sentiment Analysis is already widely used by different companies to use towards their product or brand in the digital world. However, in the offline world users are also interacting with the brands and products in retail stores, showrooms, social media etc. and solutions to measure user's reaction/expression automatically under such settings has remained a challenging task. Emotion Detection from facial expressions or a text using AI can be a viable alternative to automatically measure consumer's engagement with their content and brands. On the other hand detecting body language, which is considered as one of the most effective communication method, can help interviewers analyze the candidate during the process.

This can be achieved by creating an excel sheet of coordinate points of landmarks of desired poses and training the model on that. This trained model can then be stored and used later for predicting user's gestures.

2. BACKGROUND AND RELATED WORK

In this section, we present the main components of what we call a Body Gesture Recognition system. An important preparation step, which influences all the subsequent design decisions for such an automatic pipeline is the determination of the appropriate modelling of input (human body/gesture) and targets (emotion/expressions). Live perception of simultaneous human gesture, face landmarks, and hand tracking in real-time on mobile devices can enable various modern life applications: fitness and sport analysis, posture control and sign language recognition, augmented reality try-on and effects. MediaPipe already offers immediate, fast and accurate, yet separate, solutions for these complex tasks. Combining them all into a real-time semantically consistent end-to-end solution is a uniquely difficult problem as requiring simultaneous inference of multiple, dependent neural networks.

The MediaPipe Holistic pipeline integrates separate models for body i.e structure , face and hand components, each of which are optimized for their particular domain. The pose estimation model, for example, takes a lower resolutions and fixed resolution video frame (256x256) as input resolutions. But if one were to crop the hand and face regions/parts from that image to pass to their respective models, the image resolution would be too low for accurate articulation. Therefore, we designed MediaPipe Holistic as a multi-stage pipeline, which is more accurate than any other, which treats the different regions using a region appropriate image resolution.

First, we estimate the human pose with BlazePose's pose detector and subsequent landmark model. Then, using the inferred pose landmarks we derive three regions of interest (ROI) crops for each hand gesture (2x) and the face expression, and employ a re-crop model to improve the ROI. We then crop the full-resolution input frame/coordinates to these ROIs and apply task-specific face and hand models to estimate their corresponding landmarks. Finally, we merge all landmarks with those of the pose model to yield the full body landmarks.

The pipeline is implemented as a MediaPipe graph that uses a holistic MediaPipe landmark subgraph from the holistic landmark module and renders using a dedicated holistic renderer subgraph. The holistic landmark subgraph internally uses a pose/body landmark module, hand landmark module and face landmark module.

To summarize, the related work mentioned above, this system gives the recognition to the complete body/shape/ surface. Using holistic MediaPipe the face, hand and as well as body posture is detected. It collects the coordinates, processing image using opencv machine learning library and merges them giving the required output using scikit-learn machine learning library.

3. PROPOSED SYSTEM

The flowchart of the proposed system is given below.

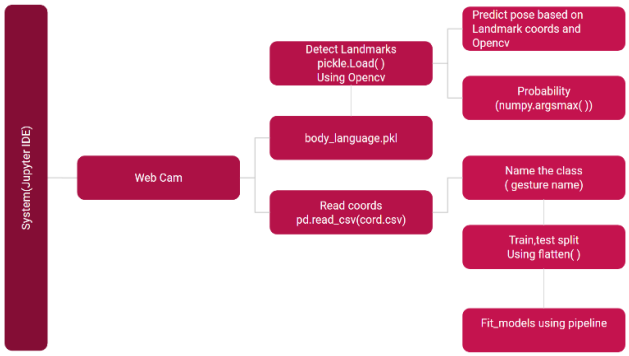

Fig 3.1: Flowchart of working model using IDE

3.1 Working

Initially we run code on jupyter IDE. when code is executed, camera starts running. It reads the coordinates using holistic MEDIAPIPE and then compares it with predefined body_language.pkl file when we trained earlier. Then it predicts the gesture. Also, it predicts accuracy of the pose comparing to pre-trained gesture using numpy-argsmax( ) function imported from Numpy library.

In case of training the model, we run the code and web cam starts running. It reads the co-ordinates from our body using holistic function imported from MediaPipe library and writes to coords.csv file. We name the class as the name of the gesture we are training. These coordinates are then clustered using flatten( ) function imported from scikit- learning ML library.

3.2 System Modules

(a) Install and import dependencies: In this step mediapipe, opencv, pandas and scikit-learn are installed and imported.

(b) Make some detections: Here a webcam window is opened using cv2 and specifications like color and thickness of our face, hand and pose landmarks are mentioned/modified.

(c) Capture landmarks and export to CSV: Here various coordinates of facial expressions and body gestures are captured and exported to a csv file named coords.csv.

(d) Train custom model using scikit learn: Here the collected data is read and processed to train our machine learning classification model on it.The model is then evaluated and serialized.

(e) Make detections with model: Now the code, when run, can make predictions of the user's gestures using the above trained model.

4. TESTING

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs (errors or other defects). Software testing can provide objective, independent information about the quality of software and risk of its failure to users and/or sponsors.

Table 4.1 Test Cases Representation

S.No	Description	Input	Expected Value	Actual Value	Result
1	Predicting a Happy face as happy	Happy face Through Web cam	Happy With probability	Happy Prob:0.97	PASS
2	Predicting a Sad face as sad	Sad face through web cam	Sad with probability	Sad Prob:0.99	PASS
3	Predicting a Victorious Gesture.	Victorious gestures through Web cam	Victory With probability	Victory Prob:0.92	PASS
4	Predicting Okay gesture.	Okay gesture Through web cam	Okay with probability	Okay Prob:0.97	PASS

Software testing can be conducted as soon as executable software/program (even if partially complete) exists. The overall approach to software testing or development often determines when and how testing is conducted and results. For example, in a phased process, most testing process occurs after system requirements have been defined and then implemented in testable.

5. RESULTS

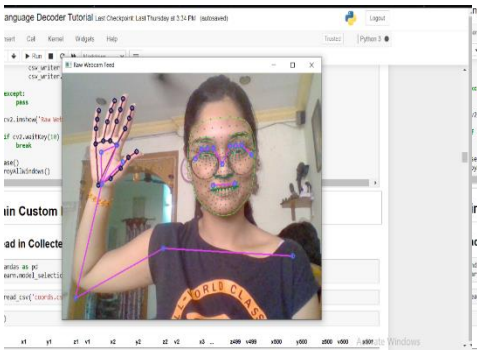

Fig 5.1 Hand landmarks

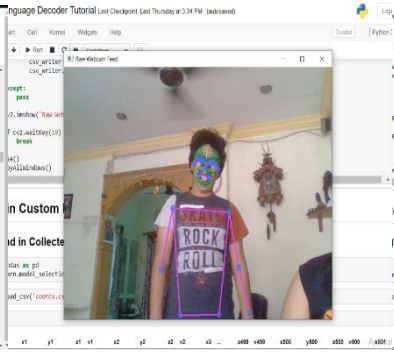

Fig 5.2 Body Landmarks

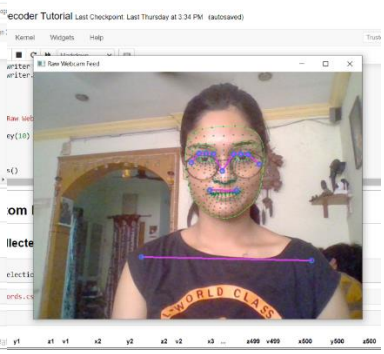

Fig 5.3 Face Landmarks

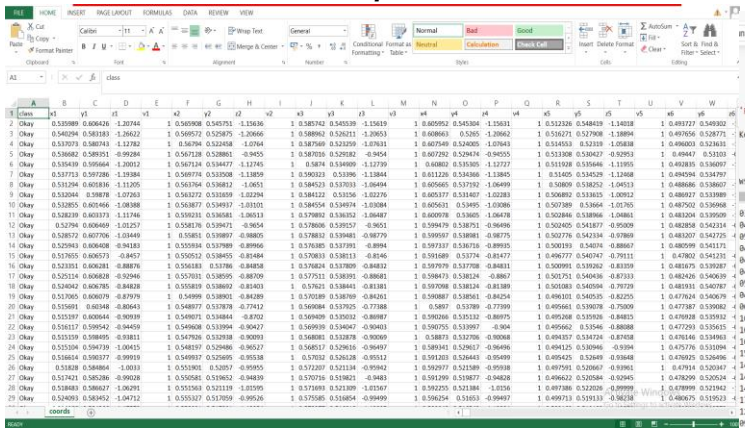

Fig 5.4 coord.csv(coordinates saved in excel sheet)

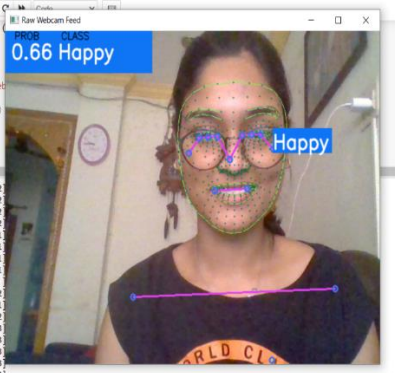

Fig 5.5 "happy" expression

6. CONCLUSION AND FUTURE SCOPE

Detecting and analyzing body language is gaining a lot of attention lately. Being able to detect and analyze facial expressions of client/customer helps businesses and marketing teams to get honest reviews and feedbacks. But facial expression is just a small part of body language. Body language consists of others elements like hand gestures and body poses. And body language plays a very important role in communication. For example in interviews, interviewers take candidate's body language into consideration. By enhancing this project, a tool can be provided to the interviewers which aids them in understanding how the candidate is responding when asked questions from different domains or put in different situations during HR rounds. Since this project supports real time hand landmark detection, hand sign language detection can also be implemented. Not only that, using this project, implementation of already existing projects like drowsiness detection of drivers, action detection etc can be made easier with much better results.

7. REFERENCES

[1] https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html
[2] https://heartbeat.fritz.ai/simultaneously-detecting-face-hand-motion-and-pose-in-real-time-on-mobile-devices-27849560fc4e
[3] https://medium.com/jstack-eu/using-machine-learning-to-analyse-body-language-and-facial-expressions-a779172cc98
[4] http://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html
[5] https://www.altexsoft.com/blog/business/functional-and-non-functional-requirements-specification-and-types/
[6] https://www.guru99.com/software-testing-introduction-importance.html
[7] https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?source=post_page----------
[8] https://link.springer.com/chapter/10.1007/978-3-642-41190-8_50
[9] https://medium.com/jstack-eu/using-machine-learning-to-analyse-body-language-and-facial-expressions-a779172cc98
[10] https://arxiv.org/pdf/1906.08172.pdf

