COMPUTER AUTOMATION USING MEDIAPIPE AND GESTURE RECOGNITION

B.NARSINGAM
Assistant Professor
Department of CSE
Teegala Krishna Reddy Engineering College, Hyderabad

SANGYAM NANDANA
Department of CSE
nandanasangam@gmail.com
Teegala Krishna Reddy Engineering College, Hyderabad

ZEESHAN TABASSUM
Department of CSE
zeeshantabassum1227@gmail.com
Teegala Krishna Reddy Engineering College, Hyderabad

BHARATH REDDY
Department of CSE
bharathreddy55355@gmail.com
Teegala Krishna Reddy Engineering College, Hyderabad

Abstract- Human-Computer Interaction(HCI) has been a multidisciplinary field of study focusing on the design of computer technology and, in particular, the interaction between humans (the users) and computers. There are multiple ways to interact with a computer and are not limited to physical hardware devices. Gesture recognition is a computing process that attempts to recognize and interpret human gestures through the use of mathematical algorithms. Gesture recognition is not limited to just human hand gestures, but rather can be used to recognize everything from head nods to different walking gaits. Computer automation is another area where scientists are trying to automate mundane, time taking tasks. Basic tasks like "Shutting down", "Opening apps", and "Visiting a particular URL" are tedious and can be automated. Utilizing the power of "Hand Tracking" and "Gesture Recognition", we can use our hands to control our system without ever touching a mouse or keyboard.

Index Terms – Hand Gesture Recognition, Facial Authentication, MediaPipe, Convolutional Neural Network.

I. INTRODUCTION

We are now in an era of industry 4.0 or the Fourth Industrial Revolution which requires automation and computerized that are realized from the consolidation between various physical and digital technologies such as sensors, embedded systems, Artificial Intelligence (AI), Cloud Computing, Big Data, Adaptive Robotic, Augmented Reality, Additive Manufacturing (AM), and Internet of Things (IoT). The enhanced digital technology connectivity made technology a crucial requirement in carrying out our daily activities like doing tasks or work, shopping, communication, entertainment, and even searching for information or news. The technology works more using machines and advances in interaction by using a broad range of gestures to recognize, communicate, or interact with each other. The gesture is known as a form of non-verbal communication or non-vocal communication that utilize the body's movement that can convey a particular message originating from parts of the human body, the hand or face are the most commonly adopted. Gesture-based interaction introduced by Krueger as a new type of Human-Computer Interaction (HCI) in the middle 1970s has become a magnetic area of the research. In the Human-Computer-Interaction (HCI), building interfaces of applications with managing each part of the human body to communicate naturally are the great attention to do research, especially the hands as the most effective-alternative for the interaction tool, considering their ability. Currently, many frameworks or library machine learning for hand gesture recognition have been built to make it easier for anyone to build AI (Artificial Intelligence) based applications. One of them is MediaPipe. The MediaPipe framework is presented by Google for solving the problem using machine learning such as Face Detection, Face Mesh, Iris, Hands, Pose, Holistic, Hair segmentation, Object detection, Box Tracking, Instant Motion Tracking, Objection, and KIFT. MediaPipe framework helps a developer focus on the algorithm and model development on the application, then support environment application through results reproducible across different devices and platforms which it is a few advantages of using features on the MediaPipe framework.

Scope of the Project

The scope of the project is this software provides the facility to use hand gestures to automate a few computer tasks quickly. This project uses a pre-trained "convolutional neural network" to predict the gesture and media pipe to perform hand tracking. The main feature of this project is 'Gesture Recognition' and 'Mediapipe. MediaPipe offers cross-platform, customizable ML solutions for live and streaming media. End-to-End acceleration: Built-in fast ML inference and processing accelerated even on common hardware. Build once, deploy anywhere: Gesture recognition is done using a pre-trained CNN, A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. Python is used for this project by Pycharm platform for stimulation.

II. LITERATURE SURVEY

In recent years, a lot of analysis has been done on Hand Gesture recognition. This recognition technology is split into 2 categories: -

II.I Vision-Based Approach

This methodology takes photos on camera as bit information. The vision-based approach focuses heavily on touch-captured pictures and brings out the most recognizable feature. Color belts were used at the start of the vision-based approach. The most disadvantage of this methodology was the quality color to be applied to the fingers. Then use blank hands rather than colored ribbons. This creates a difficult downside as these systems need background, uninterrupted lighting, personal frames, and a camera to realize period performance. Additionally, such systems should be developed to fulfill the necessities, as well as accuracy and strength. Theoretical analysis is predicated on however individuals understand data concerning their surroundings, nevertheless, it's in all probability the foremost tough to use effectively. Many completely different ways are tested up to now. The primary is to create a three-dimensional human hand model. The model is compared at hand pictures with one or additional cameras, and therefore the parameters are comparable to the form of the palm, and therefore the combined angles square measure calculable. These parameters square measure then accustomed produce the bit section. The second is to require an image mistreatment of the camera and extract bound options and people options square measure used as input within the partition algorithmic rule to separate.

II.II Sensor-Based Approach

This technique collects knowledge-generated mistreatment of completely different sensors. The info was then analyzed and conclusions were drawn in accordance with the popularity model. Within the case of hand perception differing types of sensors square measure used and placed on the hand, once the hand makes any bit, the info is recorded and analyzed. The primary sensing element used was knowledge gloves so LEDs. The introduction of the primary knowledge glove was created in 1977. The sensor-based approach impairs natural hand movements because of the utilization of external hardware. The good disadvantage is that advanced touches can't be created mistreatment of this technique.

III Methodology

MediaPipe Hands uses a Machine Learning Pipeline that integrates multiple co-working models: A palm-type acquisition model that works in an exceedingly complete image and returns a set hand-held binding box. A handwriting model that works with a cropped image location outlined by a palm detector and restores 3D reliable key points.

Existing System

- The use of python modules requires executing the code for automation. Hence, the use of hand gestures might be redundant.
- The use of OpenCV requires complex code to perform hand tracking.
- Manual automation using python can be time-consuming.

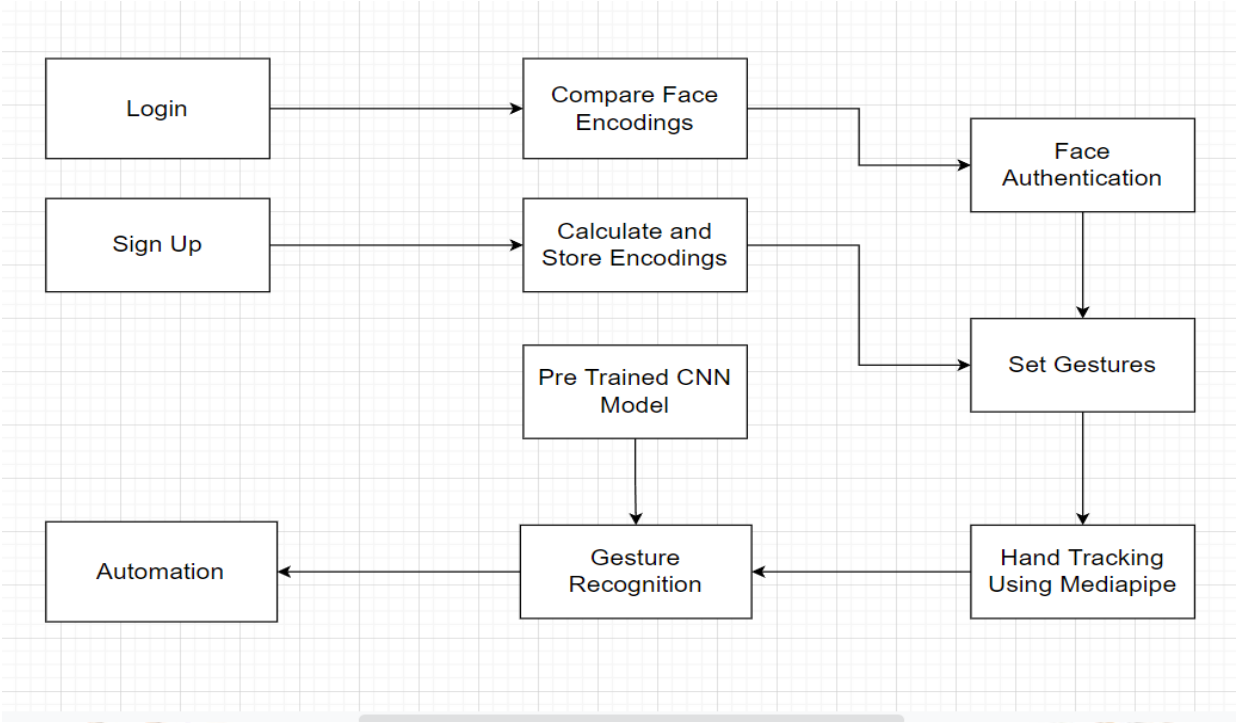
Proposed System

- In mediapipe Object localization is temporally consistent with the help of tracking, meaning less jitter is observable across frames.
- Mediapipe provides more accurate hand tracking when compared to OpenCV
- MediaPipe offers cross-platform, customizable ML solutions for live and streaming media.
- Pre-defined and changeable gestures.

IV Algorithms

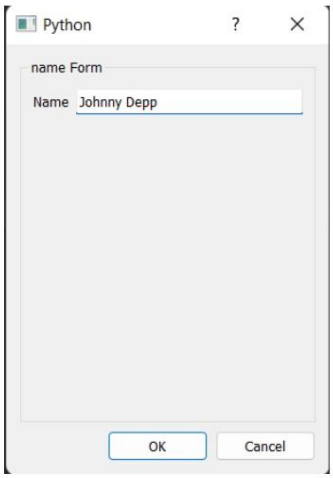
Convolutional Neural Network

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. A pre-trained model is created by someone else to solve a similar problem. Instead of building a model from scratch to solve a similar problem, you use the model trained on another problem as a starting point.

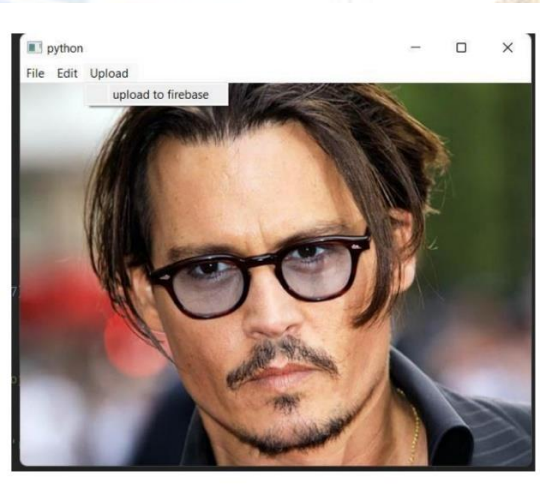


System Architecture

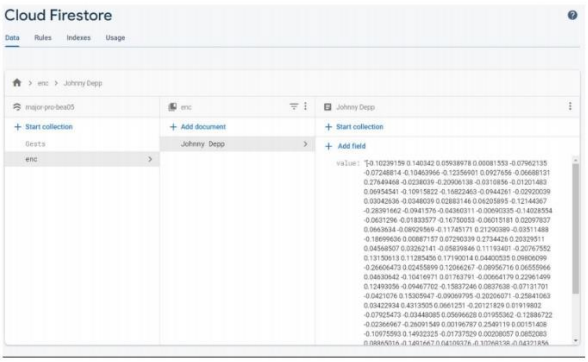
V RESULTS



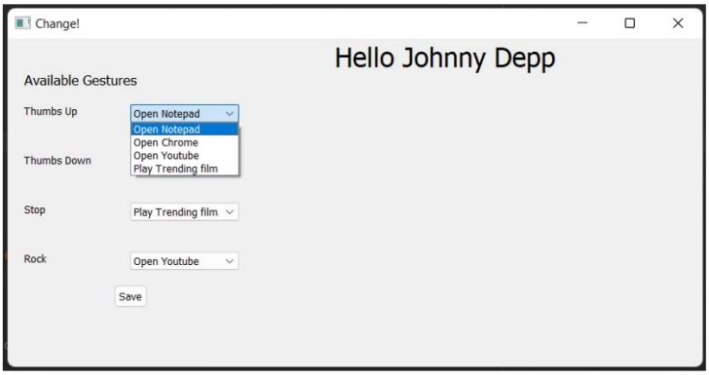
i.Upload Name GUI



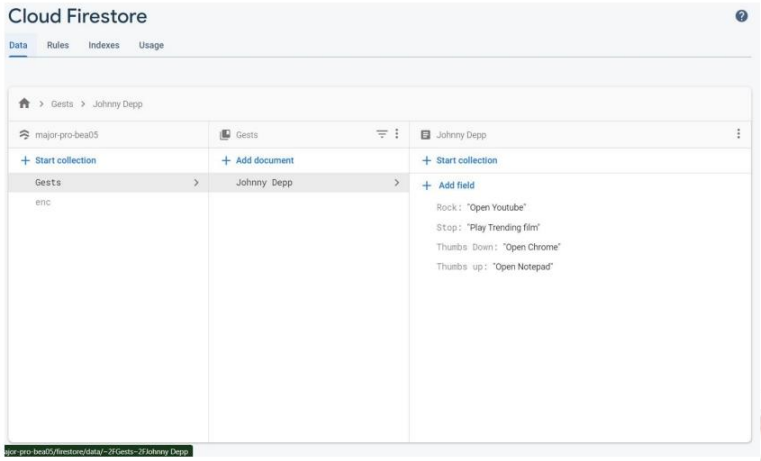
ii.Upload Image to Firebase



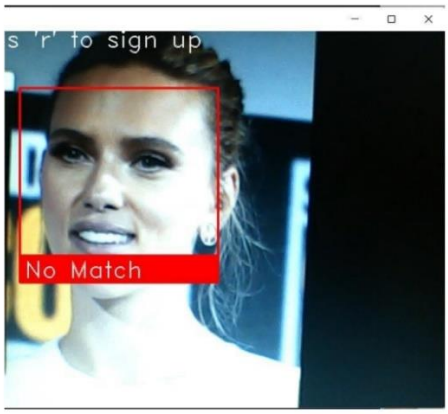
iii. Facial Encodings stored



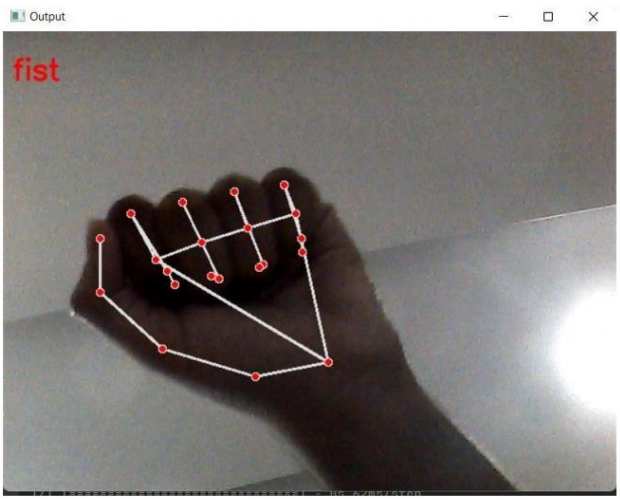
iv. Gesture GUI



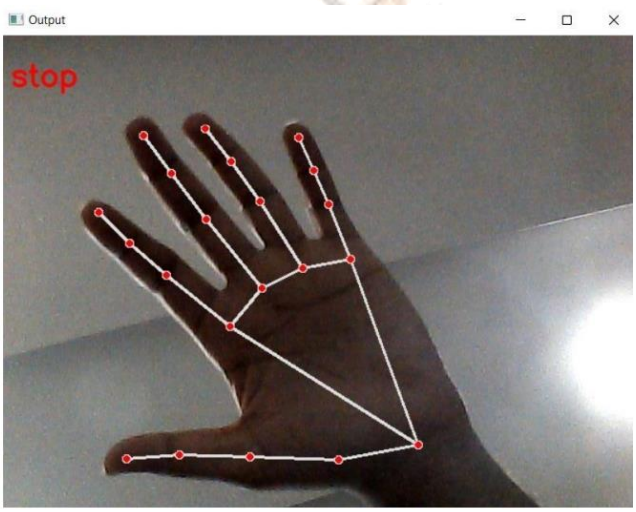
v. Gesture Data Stored



vi. Facial Match Not Found



Vii. Fist Gesture



viii.Stop Gesture

III. CONCLUSIONS

In the present thesis, we have presented a software capable of automating tasks usinggesture recognition. This is a project that incorporates several technologies to create an efficient software. We identified that numerous technologies/techniques could beaccompanied in this process namely:

- Mediapipe
- Convolutional Neural Network
- Facial Authentication
- Gesture Recognition

IV. REFERENCES

[1] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, Matthias Grundmann: MediaPipe: A Framework for Building Perception Pipelines

[2] Rafiqul Zaman Khan and Noor Adnan Ibraheem HAND GESTURE RECOGNITION: A LITERATURE REVIEW. International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012Jing-Jing Chen and Chong-Wah Ngo. Deep-based ingredient recognition for cooking recipe retrieval. In ACM Multimedia. ACM, 2016.

[3] Summerfield, Mark: Rapid GUI programming with Python and Qt: the definitive guide to PyQt programming

[4] Prabhu Ramachandran∗ automan: a simple, Python-based, automation framework
for numerical computing