Computer Automation Using Gesture Recognition and Mediapipe

Aditya Madhira[1], Naresh Mote[2], Voruganti Naresh Kumar[3]

1,2 B.Tech Student, 3Assistant Professor, Dept of CSE, CMR Technical Campus, Hyderabad, Telangana, India

Abstract: Automation is the use of technology to accomplish a task with as little human interaction as possible. In computing, automation is usually accomplished by a program, a script, or batch processing. Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures. Automation of tasks can be achieved with the help of "Gestures". Using Gestures to interact with the computer is a way of achieving Human Computer Interaction with less utilization of physical devices. Our system consists of four phases: Facial Authentication, Hand Tracking, Gesture Recognition, Automation

Keywords: Computer Automation, Gesture Recognition, Human Computer Interaction, Mediapipe, Facial Authentication

I. INTRODUCTION

HCI (human-computer interaction) is the study of how people interact with computers and to what extent computers are or are not developed for successful interaction with human beings. As an interdisciplinary field, HCI attracts researchers, educators, and practitioners from many different fields. Accordingly, many associations, special interest groups, and working groups focus on HCI or HCI-related studies [5]. Automation is important because it reduces time, effort and cost, whilst reducing manual errors. Repetitive tasks can be completed faster. Automating processes ensures high quality results as each task is performed identically, without human error. MediaPipe is a framework for building pipelines to perform inference over arbitrary sensory data [1]. By using a pre trained CNN, we are able to classify the gestures and perform the task associated with the gesture. Utilization of "Gesture Recognition" can aid us to interact with the computer with human gestures. Gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body with the intent of: conveying meaningful information or interacting with the environment. They constitute one interesting small subspace of possible human motion. A gesture may also be perceived by the environment as a compression technique for the information to be transmitted elsewhere and subsequently reconstructed by the receiver [4]. This will reduce the physical interaction and will provide a faster, easier way of computer interaction.

II. LITERATURE SURVEY

With over 20 billion electronic devices and around 10 billion people interacting with them, we need to find better ways of interaction with the devices. Automation can be achieved in multiple ways but gestures and gesture-based projects have been more utilized like we have seen in various other papers and projects.

1) "Gesture Storm", a product by Cybernet Systems Company enables weather reporters to use gestures to control the visual effects displayed in the background. This allows the reporter to display the weather picture in real time and also reduce effort and time. The idea of gesture can be implemented in other systems for everyone to interact with their own devices.

2) Utilization of "Computer Vision" library OpenCV is popular for hand and palm detection or tracking. This approach has been implemented by various developers. But with the introduction of Mediapipe which provides quick and efficient methods, OpenCV methods are convoluted.

3) Python is a high-level language which provides multiple modules which can be utilized to interact with the system. There are other frameworks which provide automation of tasks but require physical interaction with the computer.

4) Xbox Kinect is a motion sensing device for the Xbox gaming console which uses infrared projectors and detectors to perform real time gesture recognition for users to play games using their gestures. This was a good example of shifting the computer interaction method from physical/controller based to gesture based.

5) The Vision based Hand Gestures Interface for Operating VLC Media Player Application "program, in that the nearest K neighbor algorithm was used see various touches. Features of VLC media player which were driven by hand gestures including play, as well pause, Full screen, pause, increase volume, and decrease capacity. This program uses the database it contains various hand gestures and inputs compared with this image stored and appropriately VLC media player it was controlled. The current application is not very robust recognition phase [7].

III. OBJECTIVE AND SCOPE OF PROJECT

The main objective of this research is to find non-physical methods of interacting with the computer or to enhance 'Human Computer Interaction' and also to automate mundane computer tasks. The goal of this research is to utilize hand gestures to interact with computer or to automate certain tasks. The use of gestures to control a device makes it easier to use and automation reduces time and effort. Gesture recognition is beneficial as it is already implemented in certain aspects of mobile devices, there is more potential to it and this research demonstrates one of the ways of using gestures to automate tasks.

IV. PROPOSED SYSTEM

Our project is a GUI based software which utilizes multiple technologies/frameworks to automate certain possible tasks by recognizing the gestures made by the user. Python's 'face_recognition' module is used for facial authentication of the users. It produces a 128-bit vector consisting of the person's facial encodings. Users can login or sign up with the facial authentication system with just their face. A GUI window created with 'PyQT5' takes input from the user, it stores the task to be done associated with the gesture. 'Cloud Firestore', a realtime NoSQL cloud-based database is used to store the facial encodings and the gestures, the data is downloaded as long as there is internet connectivity. Mediapipe is responsible for seamless hand tracking, Mediapipe has a palm detector that operates on a full input image and locates palms via an oriented hand bounding box and a hand landmark model that operates on the cropped hand bounding box provided by the palm detector and returns high-fidelity 2.5D landmarks [2], the user then can make a hand gesture which is recognized by a pre trained CNN model. The model has high accuracy and can classify well. Finally, the associated task is automated as the gesture is recognized.
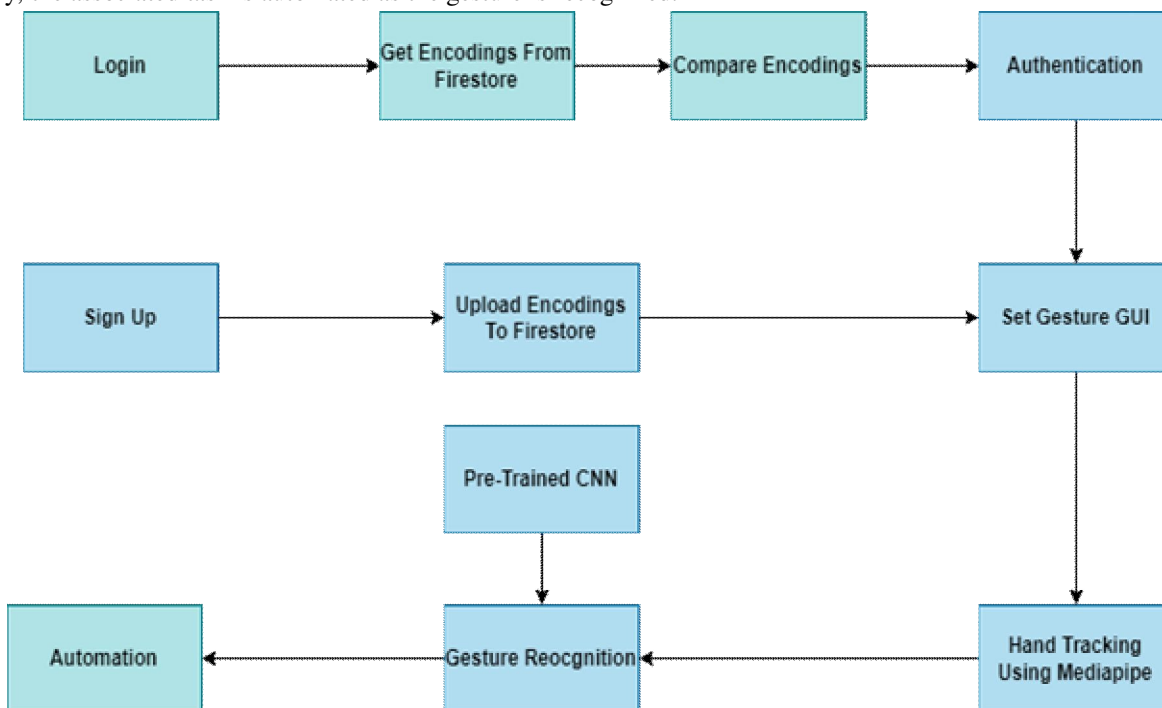

Fig. 1. Model Architecture.

V. RESULT AND ANALYSIS

The default mode is login mode where we try to perform facial recognition of the user's face. The user can start sign up mode by pressing 'r' key on the window. During sign up phase the user is required to upload his/her picture with clear visibility of face. If the user is a registered user, hand tracking and gesture recognition will be implemented, On the contrary, If the user is not registered, access will be denied for the user. After a successful authentication, user will be able to assign his/her own tasks to pre-defined gestures. Hand tracking is performed and user is free to use gestures to automate pre-defined tasks. We observed that the hand tracking performed by mediapipe is quick and efficient and can run without gpu support. The pre trained model has an accuracy of 90% in classifying the gestures made by the user.
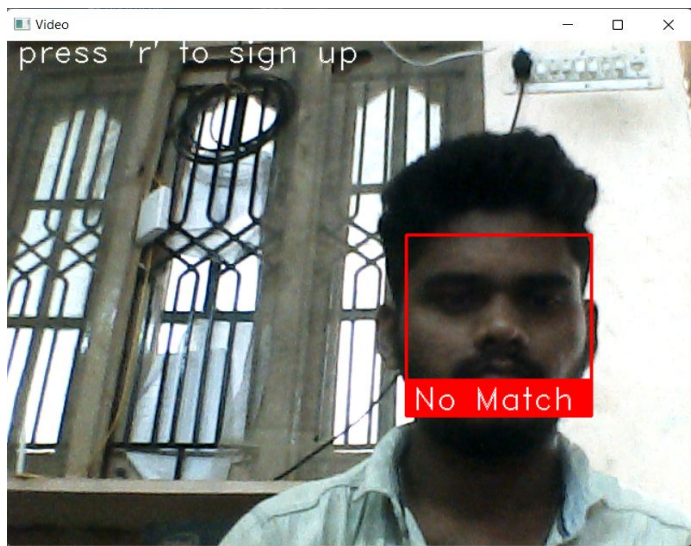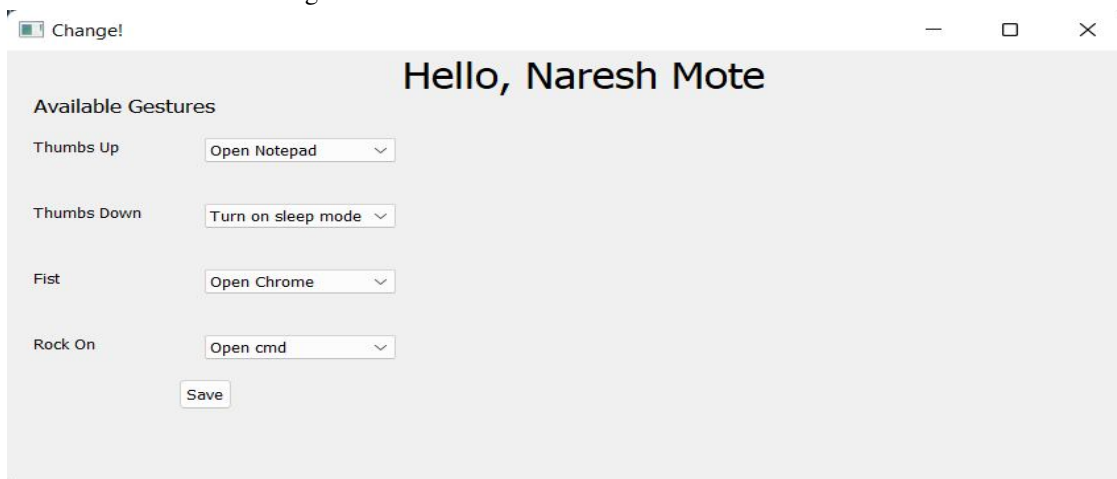
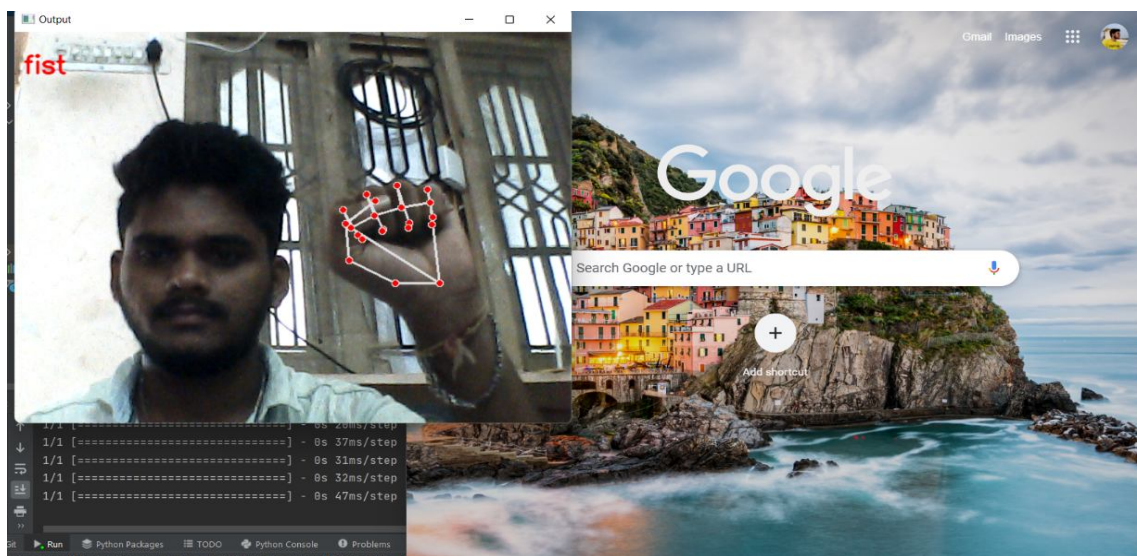Fig. 2. Facial Authentication Failed Condition.

Fig. 3. Gesture GUI

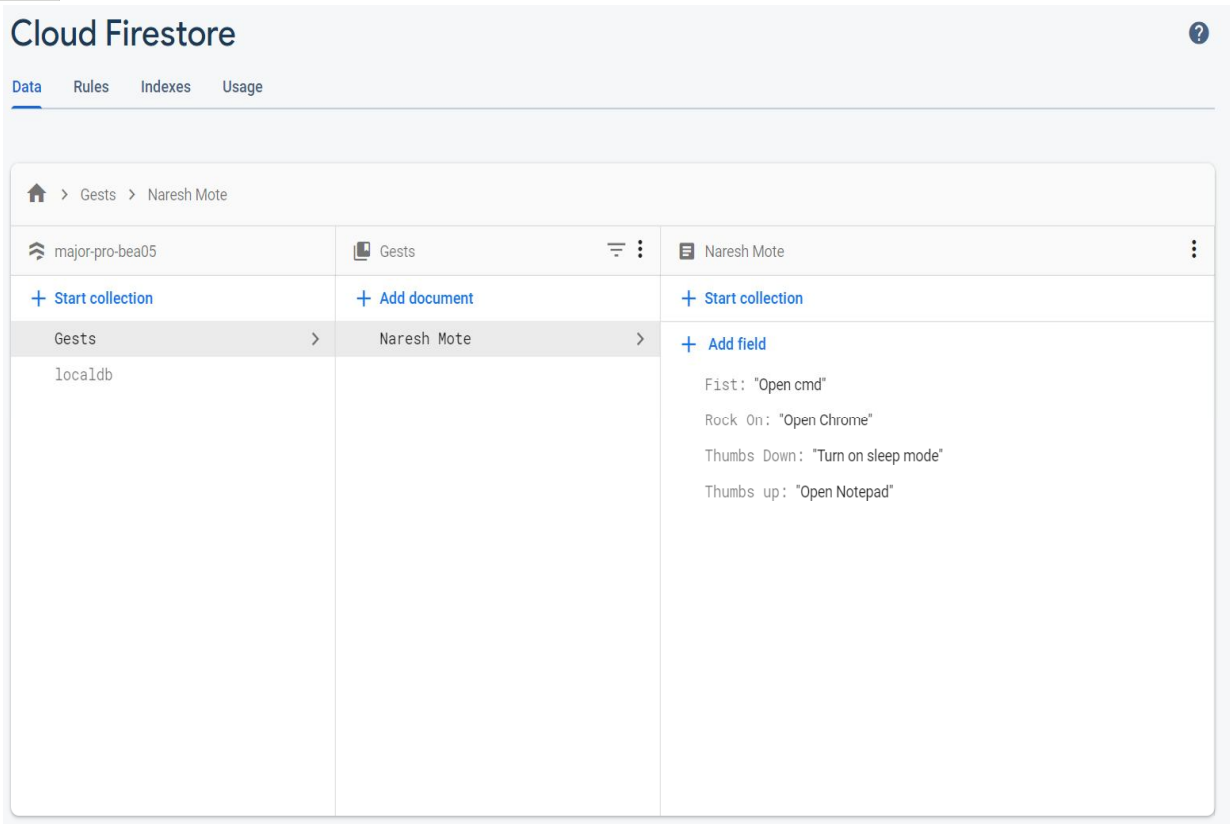Fig. 4. Opening Chrome Browser Using Fist Gesture

Fig. 5. Cloud Firestore

VI.   CONCLUSION AND FUTURE SCOPE

Computer Automation is a rapidly growing area of computer science. Engineers are trying to achieve automaton which can reduce human effort. Automatic reply of emails, webscraping, testing these tasks are being automated and are examples of automation of tasks. Human Computer Interaction is also a fast-growing field with new inventions and ideas to interact with the computer. Human computer interaction (HCI) also named Man-Machine Interaction (MMI) refers to the relation between the human and the computer or more precisely the machine, and since the machine is insignificant without suitable utilize by the human [3]. There are multiple projects and research journals about multiple ways to interact with a computer. Gesture recognition became popular and is being utilized as a way to interact with systems. Hand gesture recognition system received great attention in the recent few years because of its manifoldness applications and the ability to interact with machine efficiently through human computer interaction [3]. Gesture recognition is valuable and has proven it can be used to control devices. This idea was applied to control basic tasks of a computer with the powerful and accurate hand tracking capabilities of mediapipe. The demand for reliable personal identification in computerized access control has resulted in an increased interest in biometrics to replace password and identification (ID) card. The password and ID card can be easily breached since the password can be divulged to an unauthorized user, and the ID card can be stolen by an impostor [6]. Hence, facial recognition was used to provide reliable and secure authentication. There is a lot of potential for future implementations, it can be also deployed for mobile devices or other operating systems. Basic authentication can be added with facial authentication as to give users more options. More gestures can be added by training a neural net and with proper technology more tasks can also be automated in the future.

VII.   ACKNOWLEDGMENT

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, June 2019

[2] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, Matthias Grundmann: MediaPipe Hands: On-device Real-time Hand Tracking, June 2020

[3] 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 2012

[4] S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311- 324, doi: 10.1109/TSMCC.2007.893280

[5] Zhang, Ping; Benbasat, Izak; Carey, Jane; Davis, Fred; Galletta, Dennis; and Strong, Diane, "Human-Computer Interaction Research in the MIS Discipline" (2002). Former Departments, Centers, Institutes and Projects. Paper 40

[6] Nazeer, Shahrin & Omar, Normah & Khalid, Marzuki. (2007). Face Recognition System using Artificial Neural Networks Approach. Proceedings of ICSCN 2007: International Conference on Signal Processing Communications and Networking. 420 - 425. 10.1109/ICSCN.2007.350774.

[7] Vallabh Chapalgaonkar, Atharva Kulkarni, Amey Sonawale Media Control Using Hand Gestures: International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10 Issue IV Apr 2022

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

[9] J.Rekha, J.Bhattacharya, S.majumder   Hand Gesture Recognition for Sign Language: A New Hybrid Approach

[10] https://www.riverbankcomputing.com/static/Docs/PyQt5/

[11] https://firebase.google.com/docs/firestore

[12] https://www.optisolbusiness.com/insight/alphabet-hand-gestures-recognition-using- media-pipe

[13] https://docs.opencv.org/4.x/

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY