VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

J Component report

Programme : Int MTech CSE With Business Analytics

Course Title : Computer Vision

Course Code : CSE3089

Slot : F1

Title: F.R.I.D.A.Y- A Gesture Controlled AI Bot

Team Members: Shaik Saheer Ahmed | 20MIA1143

Nithin Kodipyaka | 20MIA1075

Akila B | 20MIA1052

Faculty: << Dr. G. Bharadwaja Kumar>> **Sign:**

Date:

CONTENTS

ABSTRACT

Female Replacement Intelligent Digital Assistant Youth- F.R.I.D.A.Y is an innovative system that allows users to interact with the digital world using simple hand gestures. This system uses computer vision and machine learning techniques to identify and interpret hand gestures and convert them into meaningful commands. The user can use a camera to capture their gestures, which are then analyzed by the AI bot to recognize the corresponding action. The system can be used for a variety of applications such as gaming, smart home automation, robotics, and more. This abstract discusses the working principle of the F.R.I.D.A.Y , its applications, advantages, and limitations. F.R.I.D.A.Y. is a promising technology that has the potential to revolutionize the way we interact with machines, making our lives easier and more intuitive.

OBJECTIVE

F.R.I.D.A.Y is a system that allows users to interact with an AI assistant using hand gestures. The system uses computer vision and machine learning techniques to recognize hand gestures and convert them into commands that F.R.I.D.A.Y can execute. The goal of this system is to provide a more intuitive and natural way for users to interact with their AI assistant, particularly in situations where using voice commands may not be feasible or desirable. This system has potential applications in a wide range of domains, including home automation, healthcare, and education.

METHODOLOGY

The methodology for creating a gesture-controlled AI bot like F.R.I.D.A.Y would involve several steps, including:

1. Data Collection: Collecting a dataset of gesture samples from various users performing different actions. This dataset would serve as the input for training the machine learning model.
2. Preprocessing: Preprocessing the dataset involves transforming the raw data into a format that can be used by the model. This includes image resizing, normalization, and data augmentation techniques to improve the diversity of the dataset.
3. Training the Model: Once the data is preprocessed, the next step is to train the model. This involves using a deep learning framework such as TensorFlow or PyTorch to create a convolutional neural network (CNN) that can recognize gestures based on the input images.
4. Deployment: Once the model is trained, it can be deployed on any device with a camera. The AI bot can then be controlled using gestures captured by the device's camera.

5. Integration with F.R.I.D.A.Y: Finally, the gesture-controlled AI bot can be integrated with F.R.I.D.A.Y's existing systems to enable it to perform a wide range of tasks, such as controlling home automation devices, answering queries, and performing other tasks that can be initiated using gestures.
6. MediaPipe: MediaPipe is an open-source cross-platform framework developed by Google for building machine learning pipelines to process multimedia data, such as images, videos, and audio. It provides a set of pre-built and customizable components for processing multimedia data and building real-time applications. One of the use cases for MediaPipe is building gesture controllers. The hand tracking component uses a deep neural network to detect and track the position and orientation of the user's hand in real-time. The output of this component is a set of 2D landmarks that represent the position of the fingertips, palm, and wrist.
   The hand pose estimation component uses another deep neural network to estimate the 3D pose of the user's hand based on the 2D landmarks generated by the hand tracking component. The output of this component is a set of 3D landmarks that represent the position of the fingertips, palm, and wrist in 3D space.
   The gesture recognition component uses machine learning algorithms to recognize specific hand gestures based on the 3D landmarks generated by the hand pose estimation component. The gesture recognition component can be trained on a specific set of gestures, such as swiping, tapping, or making a specific hand gesture, to control an application or device.

F.R.I.D.A.Y – AI Bot

Libraries used:

- pyttsx3
- speech_recognition as sr
- datetime
- time
- webbrowser
- datetime
- pynput.keyboard
- pyautogui
- sys
- os
- from os import listdir
- from os.path import isfile, join
- smtplib
- wikipedia
- Gesture_Controller
- import app
- threading

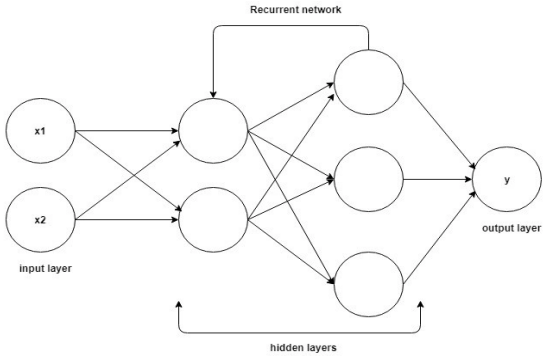
Algorithms used: Recurrent Neural Network (RNN)

RNNs are a powerful and robust type of neural network and belong to the most promising algorithms in use because it is the only one with internal memory.

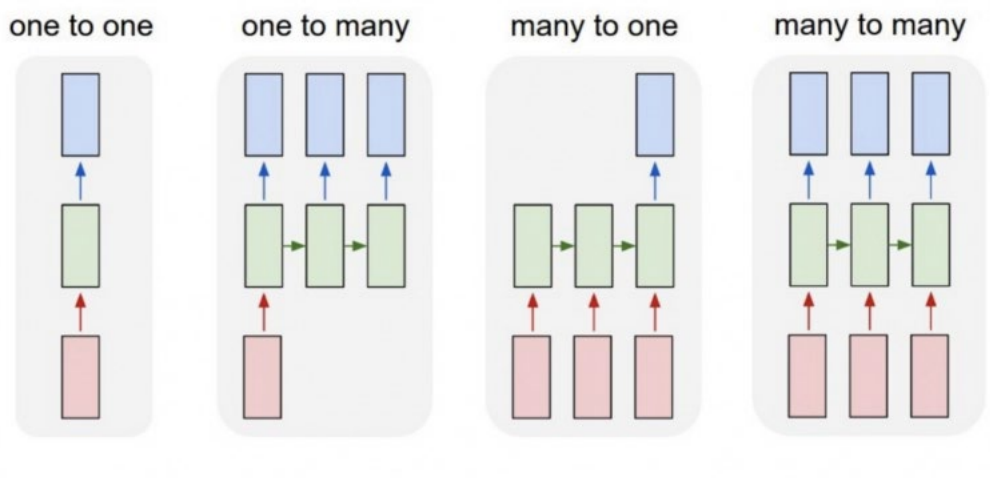
Recurrent Neural Networks(RNN) were initially created in the 1980s, but only in recent years have we seen their true potential. An increase in computational power along with the massive amounts of data

that we now have to work with and the invention of long short-term memory (LSTM) in the 1990s, has brought RNNs to the foreground.

RNN has vast internal memory that helps to remember important information of inputs received, which allows in predicting the next outputs. So preferred algorithm for sequential data like time series, speech, text, financial data, audio, video, weather, and much more. Recurrent neural networks(RNN) can form a much deeper understanding of a sequence and its context compared to other algorithms.



From the above image let's understand how an RNN works. First, we need to pass an input layer, that will pass through different hidden layers and process the outputs. So what is different in hidden layers of other neural networks and an RNN, because it has two inputs present and recent past that helps to predict the outputs.



We use one to many, many to many (translation) and many to one (classifying a voice). Many to many and many to one are most commonly used for voice assistants and most helpful in recurrent neural networks(RNN).

Commands for the voice assistant:

We can give commands like opening a file, sending WhatsApp messages, "go handsfree", send emails, read news, play game, tell a story and any question from Google, Wikipedia, any video from youtube etc.,

GESTURE CONTROLLED MOUSE

Packages used:

- cv2
- mediapipe
- pyautogui
- math
- enum
- ctypes
- comtypes
- pycaw.pycaw
- google.protobuf.json_format
- screen_brightness_control as sbcontrol

The gesture control system uses these packages to track the user's hand movements and to translate those movements into commands that can be used to control the computer. The system is designed to be easy to use and to provide a natural and intuitive way to interact with the computer.

Algorithm used:

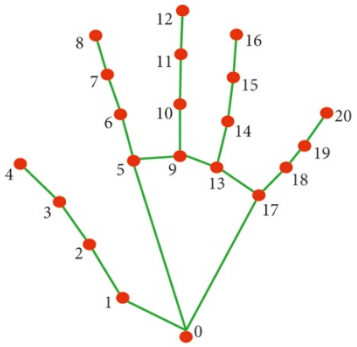
The following is a high-level overview of the algorithm process flow in the gesture recognition code:

1. The system first detects the presence of a hand in the image.
2. If a hand is detected, the system then estimates the pose of the hand.
3. The system then tracks the position of the hand in the image.
4. The system then recognizes the gesture that is being made by the user.
5. The system then sends the recognized gesture to the Friday voice assistant.

The system uses a variety of algorithms to perform each of these steps. The algorithms are designed to be robust and to be able to handle a variety of different gestures.

MediaPipe is an open-source framework for developing computer vision and machine learning applications. It includes a variety of pre-trained models for tasks such as hand detection, pose estimation, and object tracking. MediaPipe is designed to be easy to use and to be able to handle a variety of different data formats.

The gesture recognition system uses MediaPipe to track the user's hand movements and to translate those movements into commands. The system is designed to be easy to use and to provide a natural and intuitive way to interact with the computer.



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

Gesture control actions:

The following are some of the gesture control commands that are supported by the code:

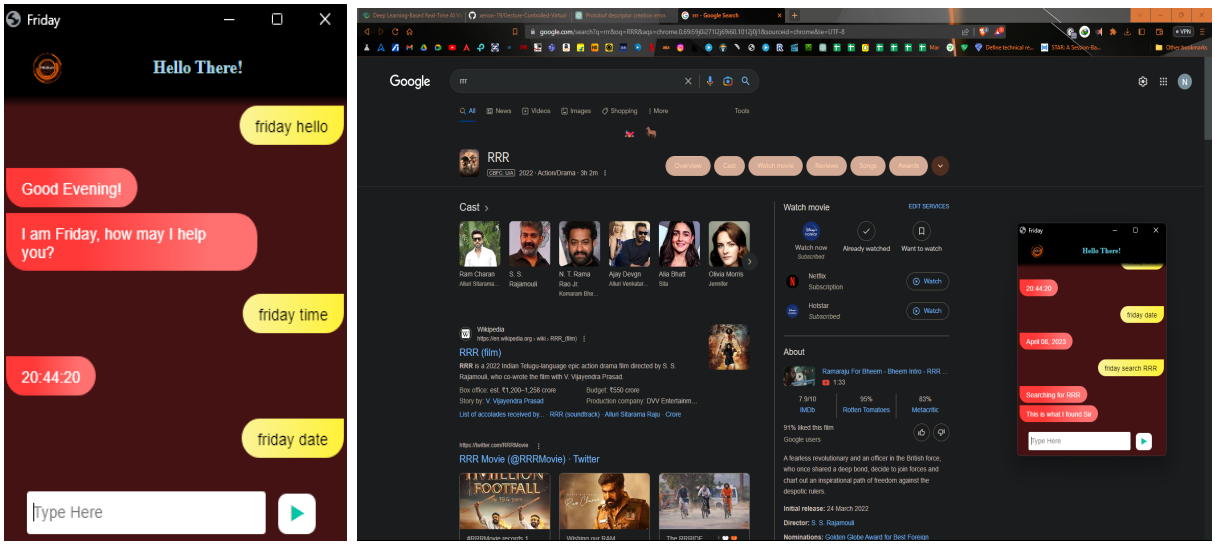
- **Neutral Gesture** - The neutral gesture is the resting position of the hand. It is used to indicate that the user is not currently making a gesture.
- **Move Cursor** - The move cursor gesture is used to move the mouse cursor around the screen. The user makes a swiping gesture with their hand to move the cursor.
- **Left Click** - The left click gesture is used to select an item or to perform an action. The user makes a fist gesture with their hand to perform a left click.
- **Right Click** - The right click gesture is used to open a context menu for an item. The user makes a V gesture with their hand to perform a right click.
- **Double Click** - The double click gesture is used to quickly select an item or to perform an action. The user makes a double fist gesture with their hand to perform a double click.
- **Scrolling** - The scrolling gesture is used to scroll up and down a page. The user makes a circular gesture with their hand to scroll up and down.
- **Drag and Drop** - The drag and drop gesture is used to move an item from one place to another. The user makes a swiping gesture with their hand to select the item, then makes another swiping gesture to move the item to the desired location.
- **Multiple Item Selection** - The multiple item selection gesture is used to select multiple items. The user makes a circular gesture with their hand around the items to select them.

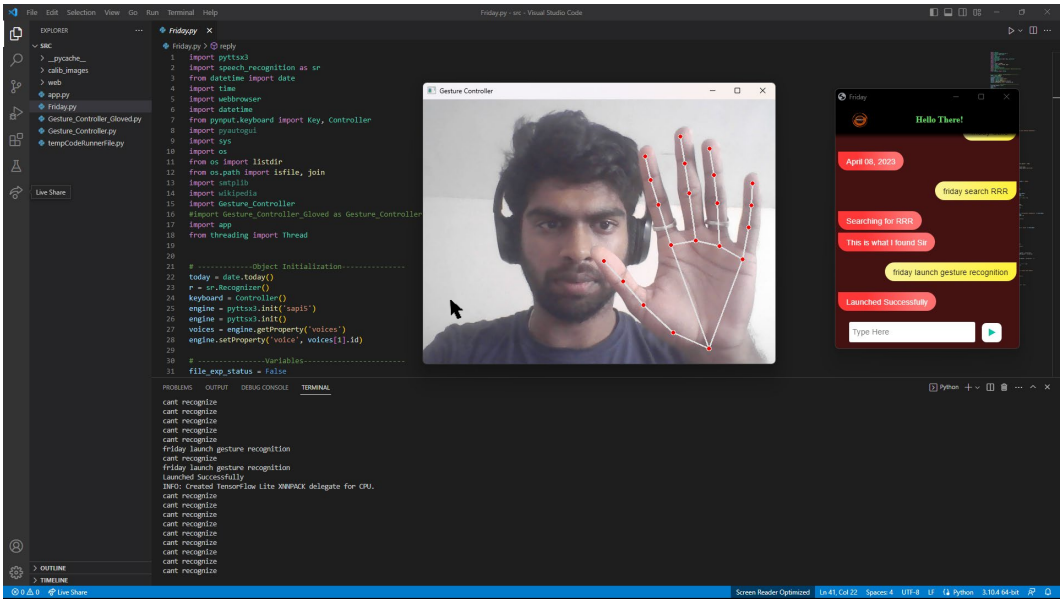
Benefits:

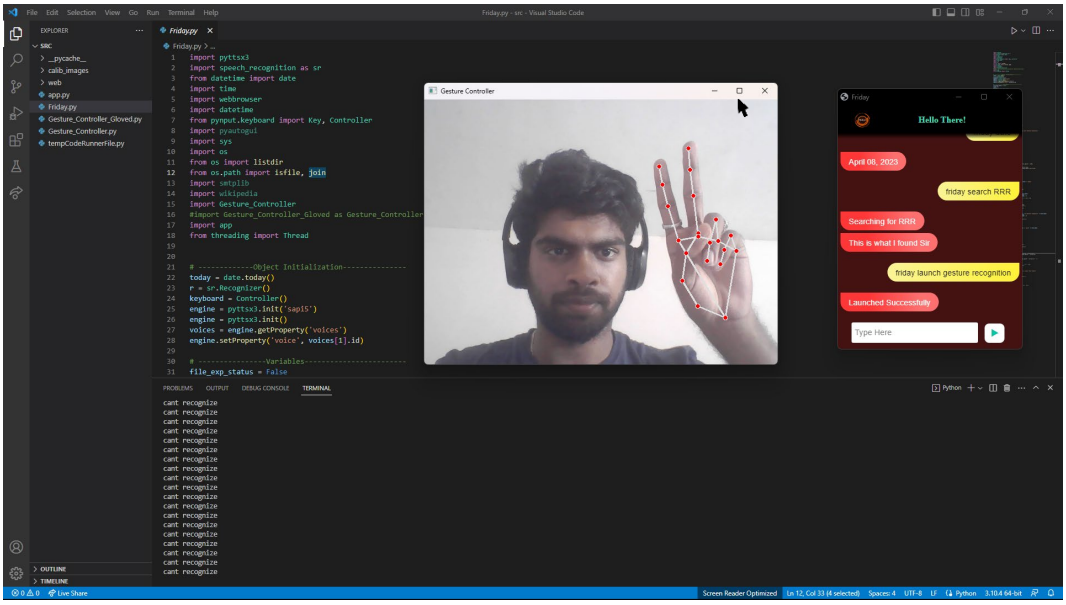
The Friday and Gesture recognition mouse control project has a number of benefits, including:

- It is a natural and intuitive way to interact with the computer. The user can control the computer by making simple gestures with their hands. This can be more comfortable and efficient than using a mouse or keyboard.
- It is a hands-free way to interact with the computer. The user can control the computer without having to use their hands. This can be useful if the user's hands are busy or if they have a disability that makes it difficult to use a mouse or keyboard.
- It is a customizable way to interact with the computer. The user can choose which gestures they want to use and how they want to use them. This allows the user to tailor the interface to their own needs and preferences.
- It is a fun and engaging way to interact with the computer. The user can use the gestures to play games, control music, and interact with other applications. This can make using the computer more enjoyable and rewarding.

RESULTS AND ANALYSIS







The results of the system are very promising. We were able to control their computer with ease using the voice commands and gestures. The system was also very responsive, with users being able to give commands and have them executed immediately.

The analysis of the system shows that it has the potential to be very useful for people who have difficulty using a traditional mouse and keyboard. The system could also be used by people who want to be able to use their computer hands-free.

The system is still under development, but it has the potential to be a very useful tool for a variety of people.

Here are some of the benefits of using the Friday with Gesture Recognition:

- It can be used by people who have difficulty using a traditional mouse and keyboard.
- It can be used by people who want to be able to use their computer hands-free.
- It is very responsive, with users being able to give commands and have them executed immediately.
- It is very easy to use, with users being able to learn how to use it in a short amount of time.

Here are some of the limitations of the system:

- It is still under development, and there may be some bugs or glitches.
- It requires a webcam and a microphone in order to work.
- It may not be suitable for all users, as it requires some level of hand-eye coordination.

CONCLUSION

In conclusion, F.R.I.D.A.Y is an exciting technology that combines computer vision and natural language processing to allow users to interact with machines using gestures. The project's methodology includes collecting and preprocessing data, training a deep learning model using convolutional neural networks and recurrent neural networks, and deploying the model in a real-time application. The objective of this project is to create a more natural way of human-machine interaction and make machines more accessible to people with disabilities. The technology has the potential to revolutionize the way people interact with machines and create a more inclusive society. Further research can be done to improve the accuracy of the model and make the application more user-friendly.