MEDIA PLAYER CONTROLLER USING HAND GESTURE & SPEECH RECOGNIZATION

Kaminee. S. Patil[1] , Kaustubh Hemant Joshi[2], Aniket Datta Adavkar[3], Varad Kiran Naik[4] , Rutwik Dashrath Jadhav[4]

[1] *Professor, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, Maharashtra, India*
[2] *Student, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, Maharashtra, India*
[3] *Student, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, Maharashtra, India*
[4] *Student Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, Maharashtra, India*
[5] *Student Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, Maharashtra, India*

ABSTRACT

The goal for the project was to develop a new type of Human Computer Interaction system that subdues the problems that users have been facing with the current system.

The project is implemented on a Linux system but could be implemented on a windows system by downloading some modules for python. The algorithm applied is resistant to change in background image as it is not based on background image subtraction and is not programmed for a specific hand type; the algorithm used can process different hand types, recognizes no of fingers, and can carry out tasks as per requirement.

Keywords: - *Machine Learning Algorithms, Media Player Controller Using Hand. Gesture etc.*

1.INTRODUCTION:

Nowadays, gesture recognition plays an important part in the interaction between humans and computers. To facilitate simple yet user-friendly communication between humans and computers hand Gestures can be used which enable us humans to interact with machines without having to use devices like keyboards, laser pens, etc. In the proposed system, users can use four simple gestures to control the Media Player without physically touching the PC. Gesture is a symbol of physical behavior or emotional expression. It includes body gesture and hand gesture.

Gesture recognition determines the user intent through the recognition of the gesture or movement of the body or body parts. In the past decades, many researchers have strived to improve the hand gesture recognition technology. Hand gesture recognition has great value in many applications such as sign language recognition, augmented reality (virtual reality), sign language interpreters for the disabled, and robot control.

The User interface has a good understanding of human hand gestures. By using the gesture, Feelings and thoughts can also be expressed. Users generally use hand gestures to express their feelings and notifications of their thoughts. Hand gesture and hand posture are related to the human hands in hand gesture recognition.

2. LITERATURE SURVEY:

In 2015, Chong Wang, "Super pixel-Based Hand Gesture Recognition with Kinect Depth Camera" proposed the system which uses kinect depth camera. It is based on a compact representation in the form of super pixels, which efficiently capture the shape, texture and depth features of the gestures. Since this system uses kinect depth camera, the cost of system is more.

In 2014, Swapnil D. Badgujar,"Hand Gesture Recognition System" proposed the system which recognize the unknown input gestures by using hand tracking and extraction method. This system is applied to recognize the single gesture. There is assumption of stationary background so that system will have smaller search region for tracking. This system only control mouse with the finger using it on web cam.

In 2014, Viraj Shinde, Tushar Bacchav, Jitendra Pawar and Mangesh Sanap developed "Hand Gesture Recognition System Using Camera". They focus on using pointing behaviors for a natural interface to classify the dynamic hand gesture, they developed a simple and fast motion history image based method. This paper presents low complexity algorithm and gestures recognition complexity and more suitable for controlling real time computer system. It is applicable only for the application Of power point presentation.

In 2014, N. Krishna Chaitanya and R. Janardhan Rao presents "Controlling of windows media player application using hand gesture recognition", this system uses various hand gestures as input to operate the windows media player application. This system uses single hand gestures and its directional motion which defines a particular gesture for the above mentioned application. In this system decision tree has been used for classification. This system only supports windows media player application and not any others.

In 2012, Ram Rajesh J., Sudharshan R., Nagarjunan D. and Aarthi R., "Remotely controlled PowerPoint presentation navigation using hand gestures" developed the system in which slides of power point presentation are controlled without using any marker and gloves. In this system the developer used the segmentation algorithm for hand detection. After detecting hand calculation is for active figures. If the fingers are not stretched properly while making a gesture then application did not work properly.

In 2006, Erol Ozgur and Asanterabi Malima, build a "A Fast Algorithm for Vision- Based Hand Gestures Recognition for Robot Control" which controlled robot using hand gestures but considered limited gestures. Firstly segmentation of hand region was carried followed by locating the fingers and then finally classifying the gestures. The algorithm used is invariant to translation, rotation and scale of the hand. This system is applicable to robot control application with reliable performance.

3. PROBLEM STATEMENT:

Build a system to control media player using hand gesture and voice commands. The problem is to control media player using hand gesture and voice commands. Media player has different action to perform.

3.1 PROBLEM DESCRIPTION:

Media player can have different actions which are as follows:
- Play
- Pause
- Forward
- Backward
- Resume
- Volume up
- Volume down

On the basis of hand gesture and above voice commands. The designed system will perform appropriate action .

INPUT:

The live hand gesture from webcam. And voice command from microphone

OUTPUT:

According to given input the system will perform appropriate operation.

3.2 REQUIREMENT ANALYSIS:

There are 3 steps to execute

STEP 1: Review the hand gesture,

STEP 2: Then Detecting the gesture of hand.

STEP 3: Play, pause, resume video in media player as per hand gestures.

4 METHODOLOGY:

The overall methodology followed in the proposed technique has three stages.

• Gathering the images or dataset from the user

At first we are using 'OpenCV' library to have a wide range of image and video processing functions, including image filtering, edge detection, feature detection, image segmentation and object tracking. Then we move on to gathering the images with the functions in the cv2 library. While gathering the images we try to gather as many images as possible as more the amount the data the better he model will be eg. We create a folder 'right' in the folder of gathering images and in that folder we gathered 1000 images which would signify 'right'. Once we are done with gathering images we are then having the 'NumPy' package, which are used for saving the files.

• Training the Model

First of all we are trying to use the Squeezenet' model which is a part of CNN algorithm. We are using gitnore to call the squeezenet module from github, integrate in our model . Then we use 'Keras' and 'Tensorflow' packages which are used to train our model. They are used for feature extraction and classification. Tensorflow in this code provides certain operations like pooling, convolution and activation functions. Keras here provides a high-level API for defining , compiling and training the model.

• Using the trained module for predictions

We are using 'pyautogui' to integrate the labels with a particular media controlling function and then using the model trained to predict the gestures and thus the label and producing a media controller function attached to it. Here we attach the functions such as 'nothing', 'rewind' and 'Forward' to the labels of data. Then after that it switches on the video feed and takes in live video inputs and functions the media functions attached to it to handle the media player.

• Accuracy testing

From 'tensorflow.keras.models' package we import the 'load_model' file, we then use the 'Evaluate' function to find the accuracy which is coming to be 89%. The model evaluate function is used to evaluate the model's accuracy on the test dataset. This function takes in the test data and labels and returns the test loss and accuracy as output. The verbose argument is set to 2 to print the evaluation process and results to the console.

5 REQUIREMENT SPECIFICATION:

Python:

Python is commonly used for developing websites and software, task automation, data analysis,and data visualization. Since it's relatively easy to learn, Python has been adopted by many non- programmers such as accountants and scientists, for a variety of everyday tasks, like organizing finances.

Mediapipe:

Mediapipe is a cross-platform library developed by Google. Provides amazing ready-to-use ML solutions for computer vision tasks.

Open CV:

OpenCV is a great tool for image processing and performing computer vision tasks. It is an open-source library that can be used to perform tasks like face detection, objection tracking, landmark detection, and much more. It supports multiple languages including python, java C++.

PYTTSX3:

pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline and is compatible with both Python 2 and 3. An application invokes the pyttsx3.init() factory function to get a reference to a pyttsx3. Engine instance. it is a very easy to use tool which converts the entered text into speech. The pyttsx3 module supports two voices first is female and the second is male which is provided by "sapi5" for windows.

6 CONCLUSION :

In current world many facilities are available for providing input to any application some needs physical touch and some without using physical touch (speech, hand gesture etc.).But not many applications are available which are controlled using current and smart facility of providing input which is by hand gesture .By this method user can handle application from distance without using keyboard and mouse.This application provides a novel human computer interface by which a user can control media player (windows) using hand gesture.The application defines some gesture for controlling the functions of windows player.The user will provide gesture as an input according to interested function.The application provides a flexibility of defining user interest gestures for specific command which make the application more useful for physically challenged people, as they can define the gesture according to their feasibility.

7 REFERENCES:

[1] "A Real-Time Gesture Recognition System for Media Players using Convolutional Neural Networks" by J. Jung, H. Kim, and H. Park. This paper proposes a real-time gesture recognition system for controlling media players using a CNN.

[2] "Real-Time Hand Gesture Recognition for Controlling Media Players using Convolutional Neural Networks" by M. M. Asghar and F. Hussain. The authors proposed a real-time hand gesture recognition system for controlling media players using a CNN and OpenCV .

[3] "Hand Gesture Recognition for TV Remote Control Using Convolutional Neural Networks" by T. H. Nguyen and N. T. Nguyen. This paper proposes a hand gesture recognition system using a CNN for TV remote control, including media player playback control.

[4] "A Novel Hand Gesture Recognition Technique for Human-Computer Interaction" by S. S. Rajput and V. K. Dixit. This paper proposes a hand gesture recognition technique using a CNN and a depth camera for controlling media players and other computer applications.

[5] "A Convolutional Neural Network-Based Hand Gesture Recognition System for Home Automation" by V. K. Meena and S. M. Rathod. This paper proposes a CNN-based hand gesture recognition system for controlling home automation devices, including media players.