Research Paper on Smart Video Player

Tejas Gavankar, Siddesh Kanekar, Shreyas Kinhekar, Chirag Kamble

Department of Electronics and Telecommunication Engineering, Atharva College of Engineering ,Mumbai , India.
(Email :gavankartejas-extc@atharvacoe.ac.in, kanekarsiddhesh-extc@atharvacoe.ac.in, kinhekarshreyas-Extc@atharvacoe.ac.in, kamblechirag-extc@atharvacoe.ac.in)

Abstract – In this paper , we have proposed a model for a working smart video player which would have two main features .One is the simulation using hand gesture technique and second one is the emotion detection of the video using CNN (ML Algorithm).This model gives an accuracy of 90-96% and hence can be studied for further process .This model yields better results than other models in all aspects.

Keywords – Arduino UNO, Python, Machine Learning , Convolution Neural Network ,Emotion Detection**.**

I.   Introduction

Designing hand gesture and facial expression based HCI systems remains one of the most difficult tasks among the many computer visions-based interactive systems. Job functions like image categorization, localization, and detection are at the heart of many of these systems. Convolutional Neural Networks (CNN, or ConvNet) are a family of deep, feedforward artificial neural networks that have been effectively used in machine learning to analyze the recognition of human action. Our fundamental goal is to discover a non-physical method of computer interaction. Our research uses a convolutional neural network to create a video player that can be operated by hand gestures. The movements would act as a direct command for actions like playing or pausing the film.For gesture-based controls, our video player employs the HC-SR04 Ultrasonic Sensor to identify the user's hand. Additionally, the user can utilize hand movements to play/pause, adjust the volume, and fast-forward/rewind video.

Human beings regularly have different moods and facial expressions changes consequently.Humans frequently experience a range of emotions, which causes their facial expressions to alter. In our approach, a deep learning system can identify a user's emotions based on their facial expression. Facial expression recognition in real time, including happy, sad, angry, terrified, surprised, disgusted, neutral, and many others. Six distinct human emotions can be found using this technique. The trained model has the ability to instantly evoke all emotions. Face identification and localization in a chaotic scene, facial feature extraction, and facial emotion classification are tasks that an automated facial expression recognition system must complete.

II.   Problem Statement

The ideal snapshot of a still hand making only one gesture against a plain background in well-lit conditions should serve as the basis for gesture detection. But that hardly ever happens in real life. When exhibiting gestures, we don't always have the luxury of using solid, transparent backgrounds.

Part of the purpose of machine learning in gesture detection is to solve some of the major technical problems involved in correctly identifying gesture photos.

i.   Non-standard   backgrounds: Gesture recognition   ought   to   perform   well

regardless of the environment; it ought to function whether you're driving, at home, or strolling along the street. You can train a machine to consistently distinguish between the hand and the background using machine learning.

ii. Movement: Logic would dictate that a gesture is more of a movement than a still image. Gesture recognition ought to therefore be able to discern patterns. For instance, rather than only identifying a picture of an open palm, we may identify a wave motion as a command, for example, to close the present program.

iii. Combination of movements: In addition, gestures may include multiple motions. As a result, we must provide context and detect certain patterns, such as the employment of a thumb to indicate a small region or a small number of files by moving fingers in a clockwise direction.

iv. Gesture diversity: Humans perform various gestures in a variety of ways. Although humans are very forgiving of mistakes, this unpredictability may make it more challenging for machines to recognize and classify gestures. Additionally, machine learning is useful here.

v. Fighting the lag: The gesture detection system needs to be built so that there is no delay between making a gesture and it being recognized. The only way to promote the use of hand gestures is to demonstrate how quick and reliable they can be. If gestures don't speed up and simplify your interaction, there really isn't any other incentive to start utilizing them. The ideal latency is one that is negative.

Even for humans, it can be challenging to identify a facial expression as being indicative of a certain emotion. According to studies, different persons are able to identify various emotions in the same facial expression. And AI has it even worse. The accuracy of current emotion identification techniques is a subject of continuing discussion.

i. Technical difficulties: Identifying an item, continuous detection, incomplete or unpredictable actions, etc. are all difficulties associated with emotion recognition. Let's examine the most common technical obstacles to deploying an ER solution and potential solutions.

ii. Face occlusion and lighting problems: When working with unprepared data, face occlusion from pose changes is a common problem for motion detection in video. Using a fractalization technique, which recognizes facial traits in video, generates pertinent landmarks, then extrapolates them to a 3D representation of a human face, is a popular way to get around the result.

III. Proposed Work

This model utilizes face recognition and hand movements for regulating the media player system. Face acknowledgment is utilized to stop and play the system and other features Various hand movements are utilized for regulating the different elements of media player system. First we will begin with the working of Simulation of Video Player

i. Video Simulating using Hand Gestures
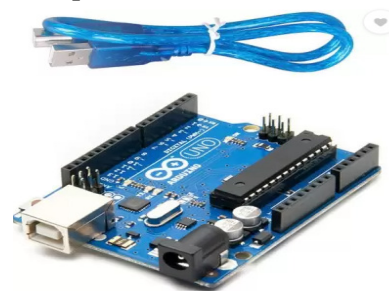
A) Components

Fig 1 : Arduino UNO

Fig2: HC-SR04 Ultrasonic Sensor

Fig3: Breadboard and Wires

B) Coding Environment

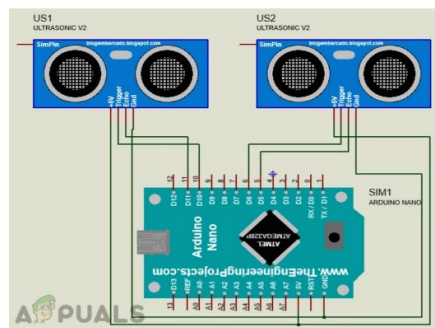Fig 4 : Arduino IDE

Fig5: Python

C) Block Diagram

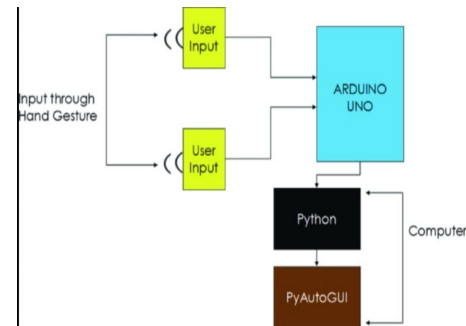Fig 6 : Connection of Ultrasonic Sensors and Arduino UNO

Fig7: Connection of Arduino and Laptop

We connect our Project components in the diagram shown above and can successfully achieve our feature 1.

ii. Emotion Detection in Video
For emotion detection , Our process will begin with training of images from FER2013 dataset ,then image classification and extraction will be the procedures followed one after the other .Below are some of the Python libraries we should have knowledge .

a. NumPy

NumPy is a python library used for working with arrays.It also has functions for working in domain of linear algebra, Fourier transform, and matrices.We are using NumPy in our testing Model to perform some numeric operations on images.We are fetching each face in x, y, w and h. After that we are drawing a rectangle and whatever size we are getting we are cropping it and storing it in gray frame.After cropping the face we need to resize or pre-process the particular face images before passing it to our Emotion Detection Model.

b. Open CV-Python

OpenCV is an open source computer vision and machine learning software library.It is designed to solve computer vision problems and it provides a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.We are using OpenCV to access the live feeds from camera and to read the video content.We are using OpenCV library's Functions Like Videocapture – For

capturing the video or accessing live the feed.Resize -- For resizing the Frame.CascadeClassifier – Used to detect the face in the Video.CvtColor – Converting the reading image in the Grayscale.Rectangle – To draw a rectangle with the help of parameters in the image.

c. Keras Model

Keras is an open source software library that provides a Python interface for artificial neural networks. It acts as an interface for the TensorFlow library. Keras is used for creating deep models which can be productized on smartphones. Keras is also used for disturbed training of deep learning models. We are going to use few tools and packages from keras library to build our model. Flow from directory is one of the tool in keras library which fetches the data images from directory by flowing through the directory. Keras packages we are going to use are – Keras.models – It is use to build sequential model. Keras.layers – It is use to create the layers by using functions like Conv2D, maxpooling2D, Dense, Dropout, flatten. Keras.optimizers. – It is use to optimize the model by using adam optimizers.

d. Pillow

Pillow library contains all the basic image processing functionality. It does the image rotation, resizing and transformation. Pillow modules allows you to pull some statistics data out of image using histogram methods, which later can be used for statistical analysis.

The final step of our methodology would be to deploy a Convolution Neural Network , which would be the interface to any video and our deployed Algorithm which would give us accurate results and proper experiment analysis.
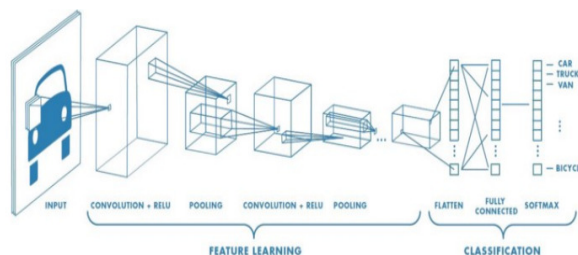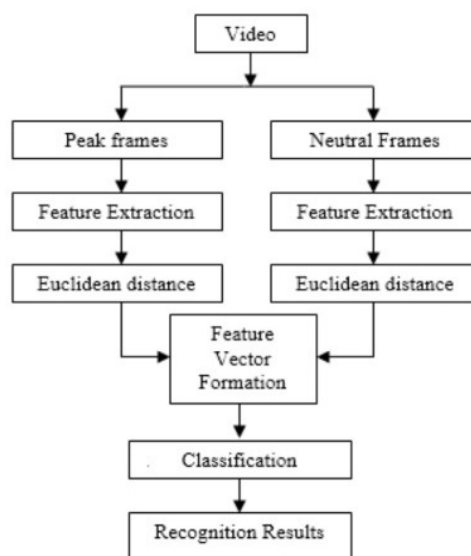

Fig 8:Convolution Neural Network

Fig 9 : Workflow of CNN Algorithm to Final Results (Emotions detected)

IV. Results

We run the application and test the system by using various hand gestures in front of the sensor that is implemented on the laptop/computer after successfully implementing all the components for the smart video player. Our hand gesture was successfully detected by the laptop, and it responded appropriately. We employ a variety of hand gestures, such as the full palm signifying a pause, the pointing fingers on the left and right signifying moving backward and forward, respectively. Waves upward and downward signify an increase in volume, respectively.

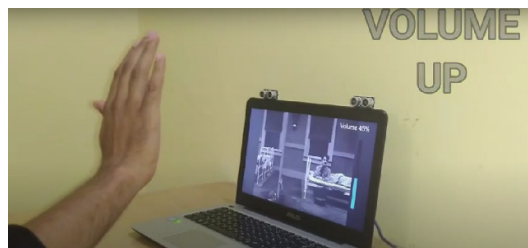Fig10 : Represents rewind feature using hand gesture

Fig 11 : Represents Volume up feature

Fig 12 : Represents Forward Feature

We use three steps in the emotion detection process—face detection, feature extraction, and emotion classification—and our suggested model, which uses deep learning, produces results that are superior to those of the previous model. The proposed strategy saves computing time while increasing validation accuracy, decreasing loss, and achieving performance evaluation that contrasts our model with an earlier existing model. On the FERC-2013 and JAFFE databases, which include seven basic emotions like sadness, fear, happiness, anger, neutrality, surprise, and disgust, we tested our neural network architectures.
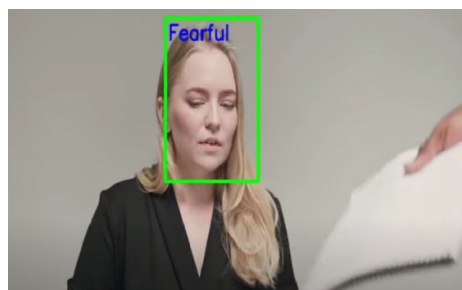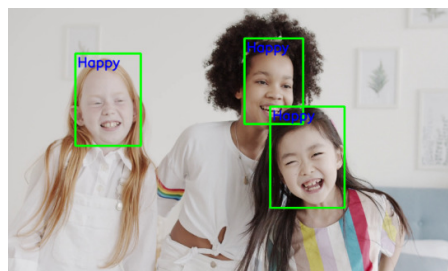

Fig 13 : Fear emotion detected

Fig 14 :Happy emotion detected on multiple faces

V.   Conclusion

Our main objective was achieved to make a controlled and smart media player considerably easier for the user. To a certain extent, the video player is roboticized in order to achieve the desired result. Utilization of face recognition and hand gesture recognition for this purpose in order to control the media player's highlights, such as pausing the video stream when the user is not looking at the screen for a certain amount of time until they do, and also controlling features like volume up or down, forward and backward (using hand gestures) was achieved.

The purpose of this research is to give a comparative analysis of various algorithms for video-based facial and emotional identification. Since the previous decade, human-computer interaction systems, biometrics, security, and other areas of computer science have all benefited greatly from automatic facial expression and emotion identification.

VI.   References

1) Bhavana .V,G Mohana Surya Mouli , G Venkata Lakshmi Lokesh, " Hand Gesture Recognition using Otsu's Method ", 2017 IEEE International Conference.
2) Akash Kumar Panda, Rommel Chakravarty , Soumen Moulik, "Hand Gesture Recognition using Flex sensor and ML Algorithms",2020 IEEE EMBS Conference.
3) Shravani Belgamwar , Sahil Agarwal , "Arduino Based Gesture Control System For

Human Computer Interface, 2018 Fourth International Conference.

4) Alvin Sarraga Alon , Julie Ann B.Susa , "Hand Gesture Recognition for an automatic Fan Speed Control System " , 2020 16th IEEE International Colloquium.

5) Sarwesh Giri , Babul Kumar , "Emotion Detection with Facial Feature Recognition Using CNN & OpenCV" ,2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).

6) Jiawen Deng , Fuji Ren ,"Multi-label Emotion Detection via Emotion-Specified Feature Extraction and Emotion Correlation Learning" , 2020 IEEE Transactions on Affective Computing.

7) Kathi Mohan Goud , Shaik Jakeer Hussain , "Estimation of Emotion Using CNN " , 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC).