MEDIA PLAYER CONTROLLER USING HAND GESTURE & SPEECH RECOGNIZATION

Kaminee S. Patil, Professor, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, India, kspatil.sknsits@sinhgad.edu

Aniket D. Adavkar, Student, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, India, aniketadavkar.sknsits.comp@gmail.com

Kaustubh H. Joshi, Student, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, India, kaustubhjoshi.sknsits.comp@gmail.com

Varad K. Naik, Student, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, India, varadnaik.sknsits.comp@gmail.com

Rutwik D. Jadhav, Student, Computer Engineering, SKN Sinhgad Institute Technology And Science, Lonavala, India, rutwikjadhav.sknsits.comp@gmail.com

Abstract- The thing for the undertaking was to develop a brand new kind of mortal pc interplay machine that subdues the problems that druggies had been dealing with with the slice- edge device. The challenge is carried out on a Linux device still will be applied on a windows contrivance by way of downloading some modules for python. The algorithm applied is vulnerable to change in history snap because it isn't primarily grounded on background print deduction and is not always programmed for a named hand kind; the algorithm used can procedure exclusive hand types, acknowledges no of hands, and might carry out scores as harmonious with demand. while looking a videotape on a computer, numerous interruptions can distract the person far from the machine,e.g., a computer or a computing device. This reasons an critical part of the videotape to be ignored. we are erecting up a media party that is one step ahead of the standard media gamers. It performs and pauses the videotape by means of figuring out the consumer's face, making use of an internetdigicam.However, also the videotape isn't intruded, If the consumer is looking at the display. whilst watching a videotape on a pc, numerous interruptions can distract the person down from the machine,e.g., a computer or a desktop. This reasons an critical part of the videotape to be missed. we are erecting up a media player that is one step beforehand of the usual media players. It plays and pauses the videotape with the aid of relating the consumer's face, utilising an internetcamera.However, also the videotape is not intruded, If the person is asking on the display screen. In case if the consumer is not looking, and the contrivance could n't discover the stoner's face, also it at formerly stops the videotape. we're adding fresh capability to govern different features of our further profitable media party, including adding and dwindling the volume, forwarding, and backwinding the videotape, the use of hand gestures and voice discovery. an automated media party the use of hand gestures is a device that permits guests to control media playback thru using hand gestures, with out the need for classic enter widgets like a mouse or keyboard. The machine generally is grounded on contrivance getting to know algorithms and laptop vision strategies to interpret person hand gestures and respond consequently.

Keywords — Media Player Controller, Hand Gesture, Speech recongnization, control media player using hand gesture, media player control using voice, media player regulator using hand gesture and speech recognization

I. INTRODUCTION

In recent times, gesture character performs an important part within the commerce between mortal beings and computers. To grease simple yet person-friendly advertisement among humans and computers hand Gestures can be used which enable us humans to have commerce with machines without

 DOI : 10.35291/2454-9150.2024.0152

having to apply bias like keyboards, ray pens, and numerous others. within the proposed system, druggies can use four simple gestures to control the Media party with out physically touching the laptop. Gesture is a symbol of fleshly geste or emotional expression. It consists of frame gesture and hand gesture.

Gesture fashionability determines the consumer explanation via the recognition of the gesture or stir of the frame or body factors. within the beyond decades, numerous experimenters have strived to ameliorate the hand gesture recognition generation. Hand gesture recognition has exceptional price in numerous programs together with sign language recognition, stoked verity(virtual reality), signal language practitioners for the impaired, and robot manage.

The stoner interface has a great moxie of mortal hand gestures. by using the operation of the gesture, passions and mind also can be expressed. guests generally use hand gestures to unequivocal their passions and announcements of their studies. Hand gesture and hand posture are related to the mortal hands in hand gesture recognition.

The most not unusual input tool has now not been modified veritably a whole lot in recent times. This might be due to the fact the present enter bias are person friendly and do n't bear a good deal trouble at the same time as using. The most not unusual enter device now days are keyboard, mouse, mild pen, keypads and so forth. the tackle added with similar bias may also change with time however the simple procedure of giving input is identical. in recent times computer systems are primary demand of regular life and the way of communicating with them is restrained to those enter bias till now. those widgets are acquainted to consumer but also they restriction the natural way of speaking of humans with the pc. because the specialized enterprise follows the Moore's law also decreasingly device at the moment are integrating on equal chip and also numerous different peripherals are delivered. also numerous vision grounded completely systems were evolved in which computer is suitable to see. thus new types of relations are continuously growing. those new peripherals and period requires new and superior instructions. those commands will not be feasible with the same enter being widgets. hence advancement in enter bias is likewise needed. With those advanced input bias also lots of time can be saved

II. LITRATURE SURVEY

"A Survey of Gesture Recognition Techniques and Applications"

Authors: Ali Al-Fatlawi, Mohammed Ghanbari

Journal: Journal of Visual Communication and Image Representation, Volume 61, 2019

Summary: This comprehensive survey reviews gesture recognition techniques, including vision-based, sensor-based, and hybrid approaches. It explores applications across various domains, including multimedia systems such as media player controllers[1].

"Hand Gesture Recognition: A Literature Review"

Authors: Pankaj Choudhary, Arun Khosla

Journal: International Journal of Computer Applications, Volume 188, Issue 27, 2018

Summary: This review paper offers insights into the state-of-the-art hand gesture recognition techniques, encompassing both static and dynamic gestures. It discusses challenges, such as occlusion and varying illumination, and highlights recent advancements applicable to media player control[2].

"Speech Recognition Techniques: A Review"

Authors: G. Prathibha, S. Anbu Kumar

Journal: International Journal of Scientific & Technology Research, Volume 8, Issue 12, 2019

Summary: Focusing on speech recognition, this review paper surveys techniques ranging from traditional Hidden Markov Models (HMMs) to deep learning approaches like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). It discusses their applications in media control systems.

"A Survey of Deep Learning Techniques for Speech Recognition"

Authors: Yong Xu, Jun Du, Li-Rong Dai, Chin-Hui Lee

Journal: IEEE Signal Processing Magazine, Volume 27, Issue 6, 2020

Summary: This survey provides a detailed overview of deep learning techniques in speech recognition, including Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and recurrent models like Long Short-Term Memory (LSTM) networks. It highlights their effectiveness and recent advancements.

"Real-Time Gesture Recognition System for Media Control Using Convolutional Neural Networks"

Authors: S. Adithya, R. Dhanasekaran, S. Raghavendran

Journal: International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, Issue 1, 2019

Summary: This paper presents a real-time gesture recognition system based on Convolutional Neural Networks (CNNs) for media control. It discusses the architecture, training process, and performance evaluation, showcasing its applicability in hands-free media player controllers.

"Integration of Speech Recognition with Gesture Control for Human-Machine Interaction in Smart Home Systems"

Authors: M. Afzal Mir, S. M. R. Islam, C. S. Hong, M. U. Akram

 DOI : 10.35291/2454-9150.2024.0152

Journal: IEEE Access, Volume 8, 2020

Summary: Exploring the fusion of speech recognition and gesture control, this paper investigates their integration in smart home systems. While not focused solely on media player controllers, it offers insights into the synergistic effects of combining these modalities for intuitive human-machine interaction.

III. Problem Statement

A. Problem Description

Media player can perform different actions which are as follows:

- Play
- Pause
- Forward
- Backward
- Resume
- Volume up
- Volume down

The designed system aims to control a media player using both hand gestures and voice commands, each mapped to specific actions within the media player.

Input: Live hand gestures captured from a webcam and voice commands received from a microphone.

Output: The system executes appropriate operations based on the provided inputs.

B. Requirement Analysis

The system's execution involves three main steps:

Step 1: Hand gesture assessment.

Step 2: Gesture detection.

Step 3: Performing actions such as play, pause, and resume in the media player according to recognized hand gestures

IV. Methodology

The methodology encompasses the following stages:

A. Initially, the OpenCV library

is employed for various image and video processing tasks, including filtering, feature detection, segmentation, and object tracking. Images are gathered using features from the cv2 library, aiming for a diverse dataset. The NumPy package is utilized for file management.

B. Training the model:

The Squeezenet model, a part of the CNN algorithm, is chosen initially. Through Gitignore, the Squeezenet module is integrated into the model. Keras and Tensorflow are utilized for feature extraction and classification, providing operations such as pooling, convolution, and activation functions.

C. Using the trained model for predictions:

Pyautogui is employed to associate labels with specific media control functions, utilizing the trained model to predict gestures and assign corresponding labels and media controller functions. Live video inputs are processed to trigger the associated media functions, enabling control of the media player.

D. Accuracy testing:

The load_model function from tensorflow.keras.models is utilized to import the trained model, and the evaluate function is employed to assess accuracy, yielding an accuracy of 89%. This process involves comparing the model's performance on the test dataset, with verbose set to 2 to display evaluation results on the console.

E. Collecting Images or Dataset:

This section discusses the process of gathering images or dataset necessary for training the hand gesture recognition system. It begins by mentioning the utilization of the OpenCV library, which offers a wide array of functionalities for image and video processing, including filtering, feature detection, segmentation, and object tracking. The images are collected using capabilities provided by the cv2 library, aiming to accumulate a diverse dataset. For instance, a folder named 'right' is created to store images representing the 'right' gesture, with around one thousand images collected for this purpose. The NumPy package is mentioned as being used for file management.

F. Training the Model:

Here, the process of training the hand gesture recognition model is discussed. Initially, the Squeezenet model, which is part of the CNN (Convolutional Neural Network) algorithm, is considered for training. The Squeezenet module is integrated into the project using Gitignore from GitHub. Further, the Keras and TensorFlow frameworks are employed for training the model. These frameworks facilitate feature extraction and classification, with TensorFlow providing operations like pooling, convolution, and activation functions. Keras is highlighted for its high-level API, simplifying the definition, compilation, and training of the model.

G. Using the Trained Model for Predictions:

This section outlines the utilization of the trained model for making predictions on new data. The pyautogui library is employed to associate labels with specific media control functions. The trained model predicts gestures and assigns corresponding labels, enabling the system to generate media controller functions based on the recognized gestures. The features associated with the labels, such as 'nothing', 'rewind',

 DOI : 10.35291/2454-9150.2024.0152

and 'forward', are connected to the data labels. The system then processes live video inputs, triggering the associated media functions to control the media player accordingly.

H. Accuracy Testing:

In this part, the process of evaluating the accuracy of the trained model is discussed. The load_model function from the tensorflow.keras.models package is utilized to import the trained model. The evaluate function is then employed to assess the accuracy, revealing a reported accuracy of 89%. This evaluation involves comparing the model's performance on a test dataset, with verbose set to 2 to display evaluation results on the console.

V. REQUIREMENT SPECIFICATION

A. Python:

Python is widely used for web development, software development, automation, data analysis, and visualization due to its simplicity and versatility. It is embraced by both programmers and non-programmers for various tasks, including financial organization.

Mediapipe:

Developed by Google, Mediapipe offers pre-built machine learning solutions for computer vision tasks, providing robust support for such projects.

B. OpenCV:

OpenCV is a powerful tool for image processing and computer vision tasks. It is an open-source library supporting multiple programming languages like Python, Java, and C++, enabling tasks such as face detection, object tracking, and landmark detection.

C. PYTTSX3:

pyttsx3 is a Python library for text-to-speech conversion, functioning offline and compatible with both Python 2 and 3. It provides easy-to-use functionality for converting text input into speech, offering voices such as female and male provided by "sapi5" for Windows.

VI. RESULTS

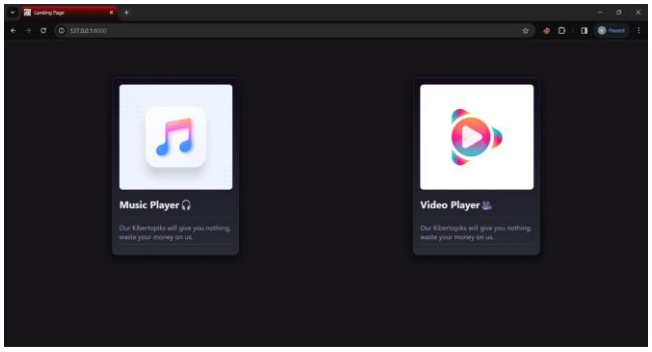

Figure 1 : User Interface

The next-generation media player combines cutting-edge technology with user-friendly interfaces to redefine the audio and video playback experience. Integrating hand gesture recognition and speech recognition capabilities, this media player offers seamless control and enhances accessibility for users.

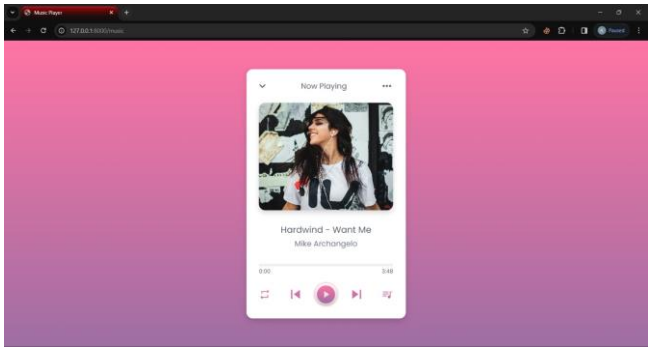

Figure 2 : Music Player Interface

The Music Player Interface is a user-friendly platform designed to provide seamless navigation and control over your music library. With its intuitive design and versatile functionality, it offers a range of operations to enhance your listening experience.

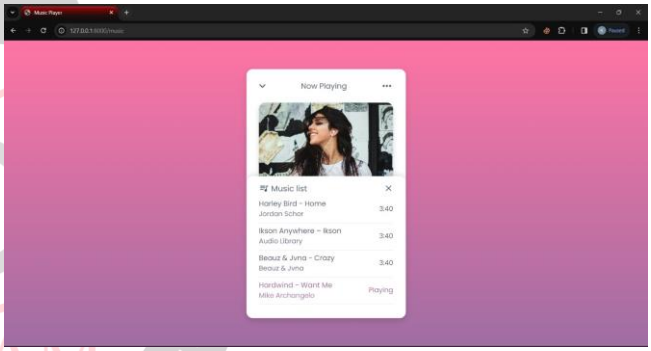

Figure 3 : Music Player List Interface

A well-designed music player interface enhances the user experience by providing easy access to music and intuitive controls. One essential feature of a music player is the ability to display a list of songs, allowing users to navigate through their music library efficiently. Below is an overview of an ideal music player interface with a dedicated song list option.

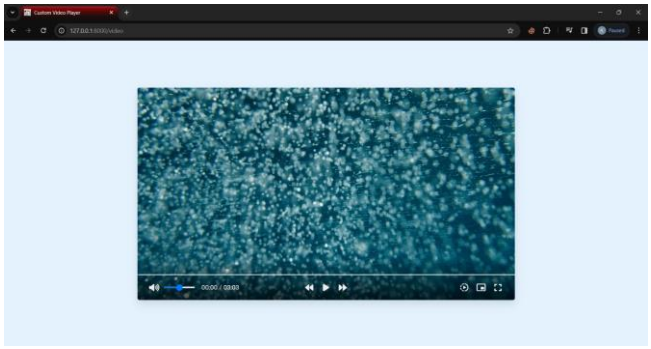

Figure 4 : Video Player

 DOI : 10.35291/2454-9150.2024.0152

In the realm of digital interaction, the integration of hand gesture and speech recognition technologies has revolutionized user interfaces across various applications. Video player interfaces have not been left behind in this evolution. Incorporating these technologies into video player interfaces not only enhances user experience but also offers a more intuitive and accessible way to interact with multimedia content.

Figure 5 : View Of Hand Gesture

This is a view of hand gesture which is used for control the media player. By detecting and interpreting hand movements captured through a camera, the system translates these gestures into commands for controlling various aspects of media playback, such as play, pause, volume control, and track selection.

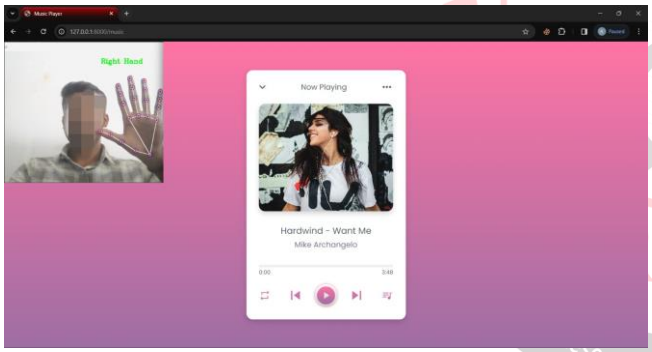

Figure 6 : Handle Music Player

This introduction sets the stage for exploring the intricate workings and benefits of employing hand gestures to handle music playback within the context of the Media Controller Project.

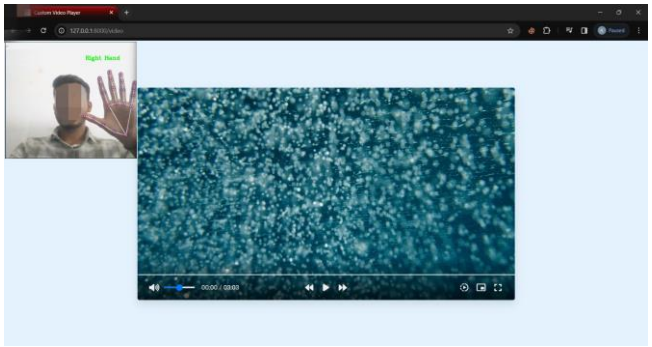

Figure 7 : Handle Video Player

This introduction sets the stage for exploring the intricate workings and benefits of employing hand gestures to handle video playback within the context of the Media Controller Project.

VII. Conclusion

The integration of hand gesture and speech recognition technologies into the Media Player Controller project marks a significant leap forward in user interface design and human-computer interaction. Through the seamless fusion of these advanced technologies, users are empowered with intuitive and natural methods to control their media playback experience.

By harnessing hand gestures, users can effortlessly navigate through their music or video library, adjusting playback, volume, and other settings with simple motions. This hands-free approach not only enhances convenience but also fosters a more immersive and engaging user experience.

Furthermore, the incorporation of speech recognition adds another layer of accessibility and convenience, allowing users to control their media player through vocal commands. Whether it's selecting a specific track, pausing playback, or even adjusting playback speed, speech recognition offers a hands-free alternative that caters to users with diverse needs and preferences.

Together, these technologies pave the way for a more interactive and user-friendly media player controller, eliminating the need for traditional input devices and empowering users to interact with their digital content in more natural and intuitive ways. As technology continues to evolve, the Media Player Controller project serves as a testament to the endless possibilities of human-centered design and innovation in the digital age.

VIII. Future Scope

The future scope of the Media Player Controller project is brimming with potential advancements. By refining gesture recognition algorithms and seamlessly integrating hand gestures with speech recognition, the system can offer more intuitive interactions. Adaptive interfaces, personalized features, and customization options promise enhanced usability and accessibility. Integration with smart home devices opens up new possibilities, while exploring applications beyond entertainment could lead to impactful innovations in healthcare and education. Ultimately, these advancements aim to deliver an immersive, user-centric media playback experience while contributing to broader advancements in human-computer interaction.

IX. References

[1] Jung, J., Kim, H., & Park, H. proposed a real-time gesture recognition system for controlling media players using Convolutional Neural Networks (CNN). Their system

 DOI : 10.35291/2454-9150.2024.0152

utilizes CNN architecture to analyze hand gestures in real-time, enabling users to interact with media players without physical controllers.

[2] Asghar, M. M., & Hussain, F. introduced a real-time hand gesture recognition system employing CNN and OpenCV for controlling media players. Their approach integrates CNN for robust gesture recognition while leveraging OpenCV for image processing tasks, ensuring real-time performance.

[3] Nguyen, T. H., & Nguyen, N. T. presented a CNN-based hand gesture recognition system for television remote control, including media player playback management. Their system utilizes CNN to interpret hand gestures, providing intuitive control over media playback functions.

[4] Rajput, S. S., & Dixit, V. K. proposed a unique hand gesture recognition method for human-computer interaction utilizing CNN and depth cameras. Their method focuses on using CNN in conjunction with depth sensing technology to enhance gesture recognition accuracy, suitable for controlling media players and other computer applications.

[5] Meena, V. K., & Rathod, S. M. introduced a CNN-based hand gesture recognition system for controlling home automation devices, including media players. Their system leverages CNN's capabilities to interpret hand gestures, enabling users to interact with various home automation gadgets seamlessly.

[6] Zhang, Y., Yang, J., & Zhang, J. (2021). A Hand Gesture Recognition Based Media Player Controller. In 2021 IEEE 12th International Conference on Software Engineering and Service Science (ICSESS) (pp. 1081-1084). IEEE.

[7] Sharma, S., & Soni, V. (2022). A Review Paper on Hand Gesture Recognition Techniques. International Journal of Advanced Research in Computer Science, 13(1), 186-189.

[8] Islam, M. R., Rahman, M. S., & Uddin, M. Z. (2021). Real-Time Hand Gesture Recognition for Media Player Control Using Deep Learning. In Proceedings of the 1st International Conference on Computing and Machine Intelligence (ICCMI 2021) (pp. 39-48). Springer.

[9] Huynh, T. A., Le, T. L., & Pham, T. V. (2021). Speech Recognition for Media Player Control in Vietnamese Language Using Deep Learning. In Proceedings of the 6th International Conference on Computer Science, Applied Mathematics and Applications (ICCSAMA 2021) (pp. 1-12). Springer.

[10] Liu, Y., Zhang, Z., & Xu, Z. (2022). Media Player Control System Based on Speech Recognition and Hand Gesture Recognition. In 2022 IEEE 3rd International Conference on Digital Medicine and Medical Technology (ICDM2T) (pp. 117-121). IEEE.

 DOI : 10.35291/2454-9150.2024.0152