Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue-8, October 2024.
DOI: *https://doi.org/10.46379/jscer.2023.07080*

Gesture-Controlled Media Player With Real-Time Facial Authentication Using Deep Learning

S.Yoheswari, J.Siva Prakash Pandi,B.K.Vignesh Babu, K.L.Vishnu Vardhan

[1] Assistant Professor, Computer Science Engineering, K.L.N College of Engineering, Sivagangai, India

[2,3,4]UG Students, Computer Science Engineering, K.L.N College of Engineering, Sivagangai,India.

Article Information

Received	:	16 Oct 2024
Revised	:	20 Oct 2024
Accepted	:	21 Oct 2024
Published	:	31 Oct 2024

Corresponding Author:

S.Yoheswari,

Email: yoheswari1988@gmail.com @gmail.com

Abstract— This system would look into making interaction with the system smoother and more user-friendly by integrating gestural control, based on hand movements, in addition to traditional tools such as keyboards and mice. Using computer vision and learning techniques, this project would assist users to control media players through real-time hand gestures and add richness to the user experience. The system is able to identify hand gestures for different media-play functions such as play, pause, or changing volumes. Besides, the application ensures that only selected users are allowed to gain control over the media player since it includes facial detection and recognition to avert privacy invasion and potential misuse. The gesture control is activated only if there is a known face recognized with the help of the camera. In all other cases, such a state provides users with the possibility to operate the media player in a simplified way that is still convenient. Recognition combined with real-time face authentication will allow this approach to add a secure user experience to currently rising demand for more natural and safe user interfaces.

Keywords: ***Gesture Recognition, Computer Vision, Deep Learning, Real-Time Control, Face Detection, Media Player Interaction, User Authentication.***

Citation: S.Yoheswari, J.Siva Prakash Pandi,B.K.Vignesh Babu, K.L.Vishnu Vardhan, "Gesture-Controlled Media Player With Real-Time Facial Authentication Using Deep Learning", Journal of Science, Computing and Engineering Research, 7(10), October 2024.

I. INTRODUCTION

In recent years, user-friendly interfaces have become really popular: they have replaced the traditional use of keyboards and mice. Keyboard and mouse interaction is very good for many tasks, but speaking in general, it is less efficient when it comes to media playback. Gesture-based device interaction is much more intuitive and convenient. The objective of this project is to make totally gesture-controlled media players, which make no use of any physical devices.

With computer vision and machine learning, the system interprets hand gestures to control the action-play, pause, and volume control-and allows for the assurance of rightful use through real-time facial recognition. It continuously monitors the system through the internal camera of the device: on gestures and facial recognition, providing for an entirely seamless and secure experience in accessing the system. Importantly, the utility is to enable users to have gestured controls without the need for special hardware

II. PROPOSE SYSTEM

The system offers an efficient and non-contact manner of controlling media playback without any sensors or wearable devices. In this project, the device's inbuilt camera identifies hand gestures in real time to ensure smooth media control. Through the use of CNNs, the system can perfectly and dynamically recognize hand gestures even in complex environments, thereby making media-interaction intuitive and responsive. In addition, RNNs are used to process gesture sequences that involve recognition of time-dependent gestures, like volume shifting or skipping tracks, so it becomes more intuitive to interact with it. The system further provides security using real-time facial recognition; only an authorized user would be capable of controlling the media player. This avoids unauthorized use and grants full security. Since the system has the traits of adapting to time and gradually learning from the gradual changes in lighting conditions and the behavior of users through deep techniques, it always maintains high accuracy over time.

Advantages of the proposed system:

- It has no external hardware, which makes it easier and cheaper to use.
- More secure, intuitive user media control experience.
- It is adaptive in varying environments and user conditions to provide constant performance.
- It is a gesture-controlled hardware-free media player with better security via facial recognition, a good real-time use gadget.

III. SYSTEM IMPLEMENTATION

The proposed system, Gesture-Controlled Media Player with Real-Time Facial Authentication, leverages computer vision and deep learning algorithms to enable a secure and intuitive media control interface. The input data is collected through the device's built-in webcam, which captures both facial features and hand gestures. This data undergoes real-time processing for gesture recognition and facial authentication to ensure that media controls are only accessible to authorized users.

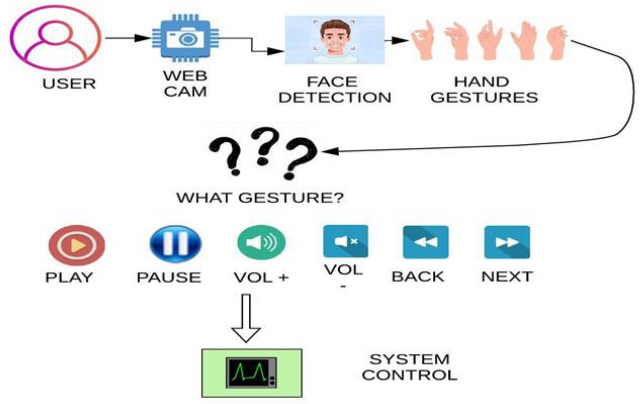


Figure 1: Architecture Diagram

Threshold Calculation for Finger Detection

Threshold Formula:

thresh = (lst.landmark[0].y * 100 - lst.landmark[9].y * 100)

Explanation:

This equation combines two y-coordinates of two specific hand landmarks, namely landmark 0-the wrist and landmark 9-the base of the middle finger, and returns their difference. Then, these two y-coordinates multiplied by 100 provide scaled-up versions of themselves, which can be used for comparison purposes. The final step in this process is to divide that number by 2 in order to set up the threshold utilized in the detection process of finger bending.

Use:

This threshold is later used to determine whether fingers are raised or curled based on the positions of other landmarks relative to this threshold.

Real-Time Hand Gesture Recognition

The system utilizes Convolutional Neural Networks (CNNs) to accurately detect and interpret hand gestures.

Convolutional Neural Networks (CNNs)

Overview:

CNNs are one of the particular classes of deep models designed especially to handle structured grid data, like images. This class of network is highly effective at drawing out spatial hierarchies and features from input data in such a manner, hence CNNs are highly useful for such tasks as image classification, object detection, and segmentation.

Architecture:

CNNs consist of several key layers:

- **Convolutional Layer:** A layer that applies filtering, which is done with a combination of convolution operations applied to the input images with the use of filters or kernels, each detecting different features - edges, textures, and so on. Each output layer is essentially a feature map, meaning each cell in this feature map represents the presence of a particular feature in the input image.

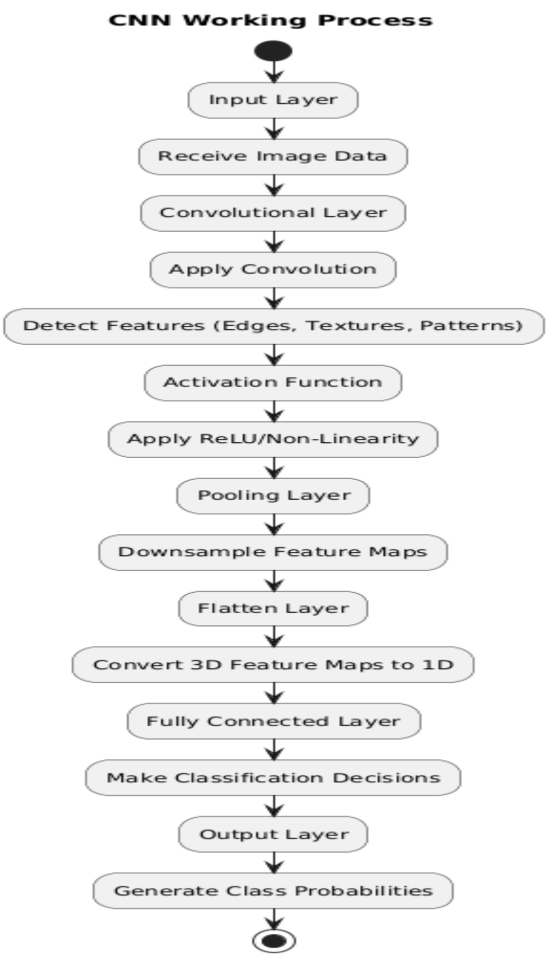


Figure 2: workflow of CNN

- **Activation Function (ReLU):** However, after convolution, the ReLU (Rectified Linear Unit) activation function introduces nonlinearity to the model so that it learns the more complex patterns.

- **Pooling Layer:** Pooling layers downsample feature maps to reduce their spatial dimensions while preserving the most important features. The most popular technique here is max pooling which takes a maximum value in a feature window as achieving translational invariance
- **Fully Connected Layer:** Several convolutional and pooling layers lead to fully connected layers. In such a layer, each neuron is connected with every neuron of the layer above it. Here is where the final predictions are made concerning the features obtained from previous layers.

Application in Hand Gesture Recognition: In the context of hand gesture recognition for media control:

- **Training Data:** They are trained on large datasets of hand gestures, where each image represents some gesture, like fingers or open palm.
- **Feature Extraction:** Convolutional layers learn the uniqueness features related to each gesture even when there is noise and changing lighting conditions
- **Gesture Classification:** Once trained, the CNN can classify real-time input images from a webcam into predefined gesture categories, enabling seamless media control.

Advantages of CNNs:

- **Robustness to Variations:** CNNs can effectively recognize gestures despite variations in hand orientation, lighting, and background.
- **Efficiency:** They are computationally efficient, capable of processing large amounts of image data quickly.

Gesture Command Mapping

Each detected hand gesture is mapped to specific media player commands, allowing for seamless media control:

- **Palm Gesture**: Triggers play or pause.
- **One Finge**: Skips forward.
- **Two Fingers**: Skips forward by a specific number of seconds.
- **Three Fingers**: Increases the volume.
- **Four Fingers**: Decreases the volume.

This mapping relies on the precise finger detection and counting logic, as described in the provided formula for detecting the number of raised fingers. By checking the relative positions of landmarks for different fingers, the system can differentiate between gestures like a fist or an open palm.

Face Recognition for User Authentication

It embeds real-time face detection and recognition in such a way that only authorized users can control the media player. It keeps checking by using face recognition on deep learning models to ensure that some recognized face is existing in the frame. This will give one more security layer and restrict the use of this system to the children who might unintentionally control the player by themselves. In a scenario where no face is detected, it blocks gesture-based controls.

Real-Time Processing

The entire system operates in real-time, ensuring minimal latency between gestures and media control.

Recurrent Neural Networks (RNNs)

Overview:

It is RNNs or Recurrent Neural Networks, a class of neural networks intended to process sequential information. Where the output at any instance depends not only on the input but also on the inputs received earlier. RNNs are very effective for tasks related to time-series data, natural language sequences, and gesture recognition sequences. The output at a given time can depend not just on a current input, but on previous inputs as well. This makes RNNs especially good for tasks involving time series data, natural-language sequences, and gesture recognition sequences..

Architecture:

The key characteristic of RNNs is that they have the ability to hold a hidden state that contains information from previous inputs:

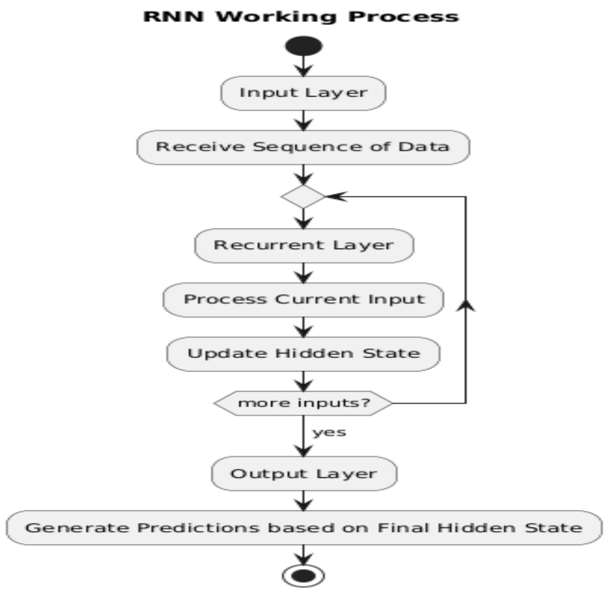


Figure 3: Workflow of RNN

- **Input** Layer**:** The RNN accepts input data at each time step, for example, a sequence of gestures.
- **Recurrent Layer:** The recurrent layer also maintains a hidden state while processing input data. Due to these updates of the hidden state at each time step, it results in processing of input data. It captures the context of previous inputs and helps in remembering information over time.
- **Output Layer:** Detection of w gesture and its duration.

Application in Hand Gesture Recognition: In real-time gesture recognition:

- **Temporal Sequence Processing:** RNNs process sequences of frames captured from the video feed, enabling the system to recognize gestures that require temporal context (e.g., holding a gesture to adjust volume).
- **Gesture Sequences:** RNNs can learn patterns in gestures that change over time, providing a richer understanding of user intentions based on the sequence of detected gestures.

Advantages of RNNs:

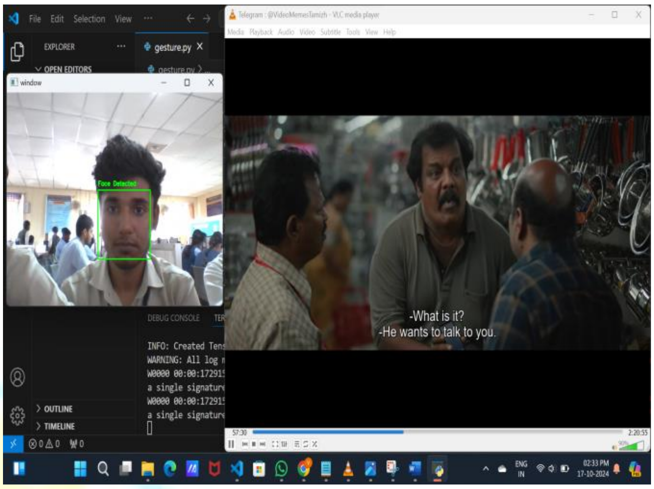
- **Handling Sequential Data:** RNNs are adept at handling sequences, making them suitable for recognizing gestures that involve continuous movement or temporal dependencies.
- **Contextual Awareness:** By maintaining a hidden state, RNNs can leverage context from previous frames to make more informed predictions.

- **Combined Use of CNNs and RNNs in Gesture Recognition**

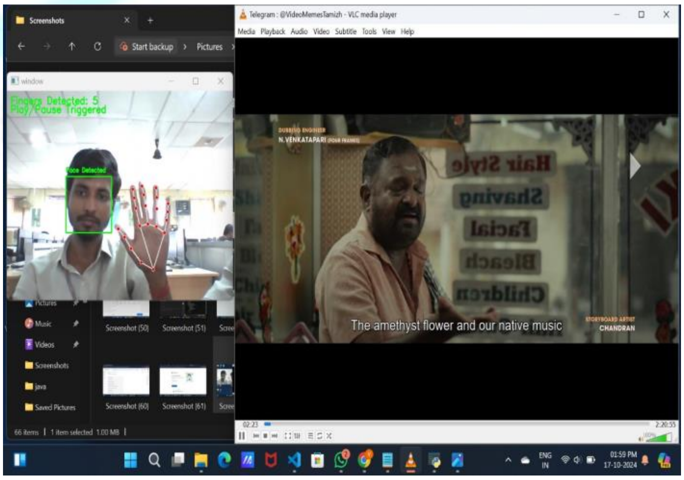
In a gesture recognition system:

- **CNNs** are typically used to extract spatial features from individual frames of video input, detecting specific hand gestures.
- **RNNs** take the output from CNNs across multiple frames to analyze the temporal patterns and sequences of gestures, ensuring accurate recognition of gestures that occur over time.
- The combination of CNNs and RNNs creates a powerful system for hand gesture recognition, allowing for accurate, real-time interpretation of user intentions. This dual approach enables the system to handle the complexities of gesture recognition in dynamic environments, making it suitable for applications like media control through hand gestures and providing a seamless user experience.

IV. RESULT



Once a face is recognized, the system allows control of the media player through natural gestures as shown in the figure below. This output demonstrates how the shift from face recognition to gestural commands might be made and allow users to nearly casually play, pause, and change the volume. The successful demonstration shows how the integration of facial authentication with gesture recognition may be a way to ensure not only highly secure but also extremely user-friendly media interaction.



Once a face is authenticated, the system enables control of the media player through intuitive hand gestures, as illustrated in the accompanying diagram. After successful facial recognition, the user can perform specific gestures to control media playback effortlessly. For example, a palm gesture triggers play or pause, while raising one finger skips forward, and two fingers skip ahead by a set number of seconds. Additionally, three fingers increase the volume, and four fingers decrease it. This seamless integration of facial authentication and gesture recognition ensures secure, responsive, and user-friendly media control.

V. CONCLUSION

The Gesture-Controlled Media Player with Real-Time Facial Authentication uses deep learning and computer vision to enable natural and intuitive control over a media player via hand gestures that could carry out actives such as play, pause, and volume control without the use of input devices. It ensures secure access by only letting authorized people control the media player since, due to continuous face detection, one cannot view and use the player while he or she is not in view of the camera. It makes use of a local camera and requires no external hardware. Thus, intensive testing of this becomes assuredly reliable and responsive, showing the strength that gesture recognition and deep learning bring to user interaction and security within media applications.

REFERENCES

[1]. Mu-Chun Su, Jieh-Haur Chen, Achmad Muhyidin Arifai, Sung-Yang Tsai, Hsi-Hsien Wei. "Smart Living: An Interactive Control System for Household Appliances." Vol no:9, Issue: January 2021, IEEE Explore

[2]. Monisha Sampath, Priyadarshini Velraj, Vaishnavii Raghavendran, M. Sumithra. "Controlling Media Player Using Hand Gestures with VLC Media Player." Vol no:14, Issue: 3, June 2022, WJAR..

[3]. Bashir, A., Malik, F., Haider, F., Ehatisham-ul-Haq, M., Raheel, A., & Arsalan, A. (2020). A Smart Sensor-based Gesture Recognition System for Media Player Control. In 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)

[4]. Nagalapuram, G. D., Shailendra, R., Dasuratha, V., & (2021). Controlling Media Player with Hand Gestures using Convolutional Neural Network. In Proceedings of the Mysuru Conference on Computing, Mathematics and Engineering Technologies. IEEE.

[5]. Rashtchian, C., & Ghaffari, A. (2021). "Hand Gesture Recognition Techniques: A Survey." Journal of Robotics and Control (JRC), 2(1), 1-9. doi:10.18196/jrc.21001.

[6]. Zhang, Y., & Zhang, H. (2021). "A Survey on Hand Gesture Recognition Based on Deep Learning." Journal of Visual Communication and Image Representation, 73, 102894. doi:10.1016/j.jvci.2020.102894.

[7]. Arifin, S. A., et al. (2020). "Gesture Recognition System for Multimedia Control using Machine Learning." 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 1-6. doi:10.1109/ICACSIS50353.2020.9234986.