International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 159-162
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0022

Hand Gesture Based Video Playback and Volume Adjustment System Using Computer Vision

Pallavi Patil[1], Prakruti Sakpal[2], Pratiksha Zilli[3], Preet Mehta[4]
[1]Assistant Professor, Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Kalaburagi – 585 102, Karnataka, India.
[2,3,4]Under Graduate, Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Kalaburagi – 585 102, Karnataka, India.
Emails: *pallavirpatil040@gmail.com[1], prakrutisakpal184@gmail.com[2], pratizilli@gmail.com[3], preetmehta432@gmail.com[4]*

Abstract
Gesture-based interaction provides a natural and contactless way to control multimedia systems, reducing reliance on traditional input devices. This project presents a Hand Gesture Based Video Playback and Volume Adjustment System using Computer Vision techniques. The system leverages real-time hand tracking and gesture recognition to perform essential video control functions such as play, pause, scroll down, screenshot and volume adjustment. By mapping specific hand gestures to playback commands, users can interact with the system naturally without relying on physical devices such as remote controls or keyboards. Volume control is achieved by measuring the distance between the index finger and the thumb, enabling smooth and continuous audio adjustment.

Keywords: *Hand Gesture Recognition, Human-Computer Interaction, Computer Vision, Gesture-Controlled Multimedia, Video Playback Control, Contactless Interaction, Dynamic Volume Adjustment, Real-Time Hand Tracking, Image Processing, Smart Multimedia Systems.*

1. Introduction

The rapid advancement of technology has significantly transformed the interaction between humans and digital multimedia systems. Conventional input devices such as keyboards, mice, and remote controls, though widely used, often limit flexibility and natural user interaction. With the growing demand for contactless systems in smart environments, healthcare, and public spaces, more intuitive interaction methods are required. Gesture recognition has emerged as an effective solution by enabling users to control systems using natural hand movements. This approach enhances accessibility, hygiene, and overall user experience [1].

1.1. Background and Need for Gesture-Based Interaction

Traditional human–computer interaction methods rely heavily on physical input devices, which may not be suitable in all environments. Gesture-based interaction offers a contactless, natural, and intuitive alternative for controlling digital systems. By using computer vision techniques, hand gestures can be detected without additional hardware, making the system cost-effective and scalable. The increasing adoption of smart environments has further emphasized the need for reliable gesture recognition systems [2].

1.2. Gesture-Based Multimedia Control System

Gesture-based multimedia systems allow users to control video playback and audio settings using predefined hand gestures. Basic actions such as play, pause, scrolling, and screenshot capture can be performed easily, while dynamic gestures enable smooth volume adjustment. The proposed system employs real-time computer vision algorithms to recognize gestures and map them to multimedia commands. Its robustness to varying lighting conditions and backgrounds makes it suitable for real-world applications such as smart classrooms, home entertainment, and public display systems [3].

2. Method

The Hand Gesture Based Video Playback, System Control, and Navigation System operates using real-time computer-vision processing to detect and classify predefined hand gestures. The methodology is divided into four stages: frame acquisition, hand landmark detection, gesture feature extraction, and action execution. The system is implemented using Python, OpenCV, and MediaPipe [4].

Algorithm 1: Real-Time Gesture Recognition and Action Control

- START (System Initialization)
- Input:
- Webcam video stream
- MediaPipe Hand Landmark Model
- Gesture–action mapping list:

Right Hand Gestures:

- Gesture 1: Volume Control

Left Hand Gestures:

- Gesture 2: Brightness Control

Universal Gestures (Either Hand):

- Gesture 3: Open Palm → Close Tab
- Gesture 4: Peace Sign (Two Fingers Up) → Screenshot
- Gesture 5: Fist → Play/Pause
- Gesture 6: Index Finger Up → Scroll Up
- Gesture 7: Index Finger Down → Scroll Down [5]

Frame Acquisition:

- Read frame from webcam using cv2.VideoCapture().

Hand Landmark Detection:

- Use MediaPipe to identify hand region and extract 21 landmark points.
- IF no hand is detected, GOTO Step 3.

Gesture Feature Extraction:

- Determine which fingers are raised or folded
- Identify hand type (Left or Righ)
- Compute distances/angles between key landmarks
- Detect dynamic changes (e.g., finger movement for volume/brightness control)

Gesture Classification:

Match extracted features to predefined gesture patterns: **Right Hand → Volume Control:**

- Move hand up → Increase Volume
- Move hand down → Decrease Volume

Left Hand → Brightness Control:

- Move hand up → Increase Brightness
- Move hand down → Decrease Brightness
- Open Palm: Close active tab
- Peace Sign: Capture screenshot
- Fist: Play/Pause media
- Index Finger Up: Scroll Up
- Index Finger Down: Scroll Down [6]

Action Execution:

IF detected gesture is stable for N consecutive frames:

- Trigger corresponding system command through OS-level automation libraries (e.g., PyAutoGUI).

Visual Feedback:

- Overlay gesture label, control state, and detection bounding box on the output frame.
- Return to Step 3 for continuous real-time gesture recognition Shown in Figure 1.

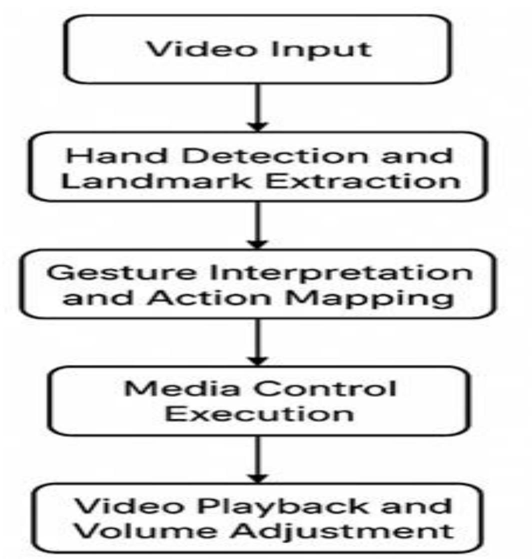

Figure 1 System Flow Diagram

3. Results and Discussion

3.1. Results

Table 1 outlines the primary objective The proposed Hand Gesture Based Video Playback and Volume

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 159-162
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0022

Adjustment System was tested under different lighting conditions, backgrounds, and hand orientations to evaluate its real-time performance, gesture accuracy, and reliability during continuous operation. The system achieved stable recognition using the MediaPipe hand-landmark detector along with custom gesture-mapping logic [7].

3.1.1. Gesture Recognition Accuracy

The system was evaluated using seven predefined gestures. Table 1 summarizes the average recognition accuracy recorded during testing with 20 participants.

Table 1 Gesture Detection Performance

Gesture	Function	Accuracy
Right hand index finger up	Scroll up	93.4%
Left hand index and thumb	Scroll down	91.8%
Right hand index and thumb	Volume Control	92.6%
Peace sign	Screenshot	95.6%
Right hand open Palm	Pause / Play	93.1%

The fist gesture exhibited the highest accuracy because of its distinct landmark pattern, whereas brightness and volume control achieved slightly lower accuracy due to variable lighting affecting continuous tracking [8].

3.1.2. Real-Time Performance

The system maintained an average frame processing rate of 25–30 FPS on a standard laptop (Intel i5/8GB RAM), ensuring smooth gesture recognition without noticeable latency.

- Latency between gesture recognition and action execution was observed to be below 200 ms, making the system responsive for real-time interaction.
- CPU usage remained within 40–55%, indicating that the system is lightweight and suitable for real-time environments.

3.1.3. Robustness Under Varying Conditions

Testing was conducted under different illumination levels and backgrounds:

- **Bright lighting:** Achieved the highest detection stability.
- **Low lighting:** Accuracy dropped by approximately 8–10%, mainly for dynamic gestures such as volume and brightness control.
- **Cluttered background:** Did not significantly affect recognition due to MediaPipe's robust segmentation model.

Despite environmental changes, the system maintained overall reliability, demonstrating adaptability for general indoor usage [9].

3.1.4. User Experience Evaluation

Participants reported that the system felt:

- Intuitive, since gestures closely resembled natural hand movements.
- Effortless, enabling video control without physical contact.
- Useful, especially for situations requiring touch-free operation (e.g., while cooking, exercising)

However, some users faced difficulty maintaining stable gestures for continuous controls (brightness/volume), suggesting a need for enhanced smoothing algorithms.

3.2. Discussion

The results demonstrate that the system successfully recognizes a wide range of gestures and translates them accurately into system-level commands. The combination of MediaPipe landmarks with simple geometric features provides a strong balance between accuracy and computational efficiency. While environmental factors such as lighting influence dynamic gesture performance, the overall usability remains high [10 - 15].

Future improvements may include:

- integrating adaptive brightness/volume sensitivity
- using temporal gesture recognition models such as LSTM/Transformers
- improving low-light detection with histogram equalization or IR-based cameras.

Overall, the system proves to be a reliable touch-free

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 159-162
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0022

interface for media playback and system control.

Conclusion

Hand Gesture-Based Video Playback and Volume Adjustment System using computer vision, offering a natural intuitive and contactless interface for multimedia control. The system successfully integrates a real-time hand detection, gesture recognition, and dynamic volume adjustment, allowing users to perform essential video operations such as play, pause, forward, backward, scroll up and scroll down, screenshot and smooth volume control without relying on physical devices. The proposed system enhances user experience, accessibility and hygiene, making it suitable for applications such as smart classrooms, public displays, and home entertainment environments. By utilizing computer vision techniques, the system enables seamless and interactive multimedia control in a user-friendly manner.

Acknowledgements

Authors acknowledge the support from Poojya Doddappa Appa College of Engineering, Kalaburagi -585102, India for the facilities provided to carry out the research.

References

Journal reference style:

[1]. Xiang Li, Yan Liang, "GesPlayer: Using Augmented Gestures to Empower Video Players", IEEE Transactions on Human-Machine Systems, 2022.
[2]. Vasileios Sideridis et al., "GestureKeeper: Gesture Recognition for Controlling IoT Devices", IEEE Access, 2019.
[3]. Okan Köpüklü et al., "Real-Time Hand Gesture Detection and Classification Using Convolutional Neural Networks", IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2019.
[4]. Pascal Schneider et al., "Gesture Recognition with Body Keypoints and DTW for Human-Computer Interaction", IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2019.
[5]. Kumar, R., & Thakur, D. (2021). Gesture Controlled System using Computer Vision. International Research Journal of Engineering and Technology (IRJET), 8(5), 1234-1238.
[6]. Szeliski, R. (2010). Computer Vision: Algorithms and Applications. Springer.
[7]. Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.
[8]. Molchanov, P., Gupta, S., Kim, K., & Kautz, J. (2015). Hand Gesture Recognition with 3D Convolutional Neural Networks. CVPR.
[9]. M. Bhuiyan R. Picking "Gesture Control User Inter-Face What Have We Done and What's Next?" ,2009.
[10]. R. GONZALEZ R WOODS Digital Image Processing New Jersy: Pearson Education. Inc.
[11]. S. NAJI R. ZAINUDDIN H.A. JALAB "Skin segmentation based on multi pixel color clustering models,2012.
[12]. D DIAS R MADEO T. ROCHA H. BISCARO S. PERES "2009. Hand movement recognition for Brazilian sign language: a study using distance-based neural networks. Neural Networks,2009.
[13]. G. GOMEZ "On selecting colour components for skin detection. Pattern Recognition 2002".
[14]. S. SINGH D. CHAUHAN M. VATSA R. SINGH "A robust skin color based face detection algorithm", 2003.
[15]. S. UMBAUGH Computer Vision and Image Processing: A Practical Approach Using Cviptools with Cdrom Prentice Hall PTR 1997.