Review Paper on Smart Gesture-Based Equipment Control System

Mr. Om Koli[1], Mr. Suraj Patil[2], Mr. Pradip Khatal[3], Prof. S.S. Patil[4]

Assistant Professor, UG Student, Department of E & TC, Adarsh Institute of Technology & Research Centre Vita, India

DOI: https://doi.org/10.51244/IJRSI.2025.12110029

Received: 25 November 2025; Accepted: 01 December 2025; Published: 04 December 2025

ABSTRACT

In settings where touch-based interfaces are uncomfortable or unsanitary, gesture-based human–machine interaction has become a popular way to operate smart devices. Gesture recognition systems are now very accurate, responsive, and appropriate for real-world applications thanks to developments in computer vision, deep learning, IoT communication, and embedded CPUs. A thorough theoretical examination of gesture detection technologies is presented in this review paper, with particular attention to CNN-powered gesture classification techniques, MediaPipe hand-tracking models, and OpenCV-based image processing workflows. Additionally, it looks at how Internet of Things microcontrollers like ESP32 can be used to enable wireless, realtime control of electrical appliances through relay modules. In order to determine performance trends, system reliability, and practical issues, the study synthesizes findings from several research investigations. The focus is on developing a smooth and clean control environment that improves user convenience, facilitates accessibility for users with physical disabilities, and aids in the creation of next-generation smart homes. This enhanced assessment is appropriate for academic submissions and engineering research since it blends scientific depth with practical relevance.

Keywords: Gesture Recognition, Media Pipe, Smart Home Automation, OpenCV, Deep Learning, ESP32, Internet of Things, Human–Computer Interaction, Convolutional Neural Network.

INTRODUCTION:

The way people engage with electronic systems has changed dramatically in recent years due to the incorporation of intelligent automation into daily life. Physical touch or human effort are necessary for traditional control mechanisms like switches, remote controllers, and mobile applications, which may not always be possible or acceptable. In settings where users must carry goods, have limited mobility, or work in sterile settings like hospitals and labs, these techniques may become cumbersome. Additionally, the development of gesture-based control systems that rely only on hand movements for interaction has accelerated due to the growing need for touchless interfaces, which has been highlighted throughout global health concerns. Users can interact with electronic devices more naturally and intuitively thanks to gesture recognition. One of the most expressive ways to communicate without using words is through human gestures, particularly hand motions. Computers can now interpret hand shapes, finger movements, and motion patterns in real time thanks to developments in artificial intelligence and computer vision. By offering a highly reliable, lightweight, and precise hand landmark detection architecture that can identify 21 crucial locations on the human hand, Media Pipe in particular has transformed this field. Gesture systems are now dependable enough for real-world implementation thanks to OpenCV's robust image-processing tools and deep learning architectures like CNNs. By providing smooth management of smart appliances, the incorporation of IoT microcontrollers like ESP32 has further reinforced gesture-based systems. Relay-module interfacing, low latency response, and Wi-Fi connectivity are all supported by ESP32, all of which help with effective device switching. These technologies can be used to create a gesture-controlled environment that allows users to activate lights, fans, air conditioners, and other equipment without having to touch them. By offering thorough theoretical descriptions of the technologies, procedures, design factors, and performance assessments associated with gesture-based equipment control systems, this review paper builds on previous research. The objective is to provide a review that is both intellectual and approachable, bridging the gap between theoretical

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

comprehension and real-world application. The technological underpinnings, current research investigations, system approach, suggested architecture, real-time performance, and possible future

enhancements are covered in the parts that follow.

Technology Background:

For dynamic hand gesture identification, we employed a CNN classifier. The preprocessing processes for my model, the details of the classifier, and the training pipeline for the two subnetworks.

1) Computer Vision Techniques

Computer vision acts as the primary mechanism for interpreting hand movements. In a gesture-controlled environment, the camera continuously captures frames that require extensive preprocessing to ensure that the hand region is correctly isolated. The system first converts the captured RGB frames into more suitable color spaces such as HSV or grayscale to simplify segmentation. Noise introduced by poor lighting conditions or camera quality is mitigated using filters like Gaussian blur or median filtering, which help smooth the image while preserving important edge details.

Once preprocessing is completed, additional operations such as thresholding and region masking help distinguish the hand from the background. Classical vision techniques like Canny edge detection or contour extraction can be applied to understand the overall shape and boundary of the hand. Although these traditional methods have limitations in dynamic environments, they remain relevant as supplementary processes in modern hybrid gesture systems.

2) MediaPipe Hand Tracking

One of the most significant breakthroughs in gesture recognition came from Media Pipe, a framework developed by Google that provides highly robust real-time hand tracking. Media Pipe uses two separate models: one for palm detection and another for detailed hand landmark estimation. The framework identifies twenty-one precise landmarks scattered across finger joints, fingertips, and the palm base. These landmarks are generated in three dimensional space, enabling the system to account for depth variation and perspective changes.

Media Pipe's efficiency lies in its ability to perform landmark regression with minimal computational overhead, making it suitable for laptops, smartphones, and embedded GPU devices. Even under moderate motion or partial occlusion, the system exhibits stable and continuous tracking. This capability greatly enhances gesture classification accuracy because landmark-based representations are far more consistent than raw image inputs.

The detailed hand landmark reference used in many systems is shown at.

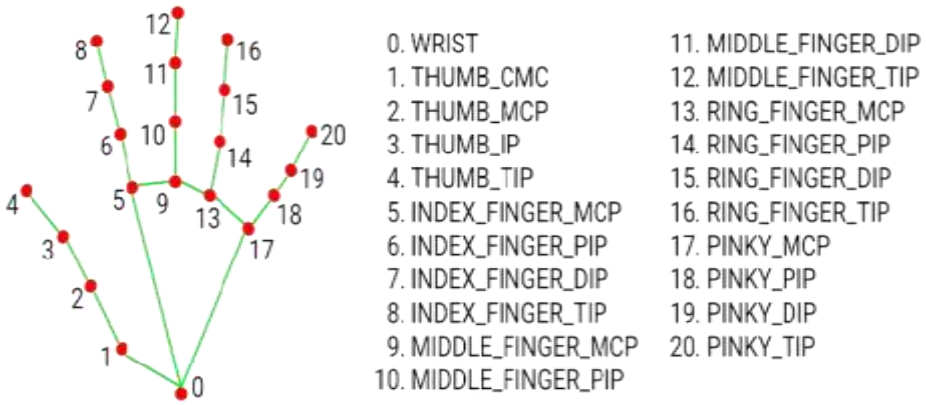

Fig. "MediaPipe Hand Tracking: Visual Reference of Key point Indices and Labels"

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

3) OpenCV

Although OpenCV is a free, open-source, real-time image processing framework that can identify and detect a variety of objects, we are now concentrating on creating methods and strategies for identifying and detecting human hand movements. This library provides the application, but hardware components are also needed. In the hardware category, cameras, 3D sensors like Kinect, and a built platform that can run the OpenCV library are frequently used for object identification, detection, and image categorization. CNNs, or 3D networks, are widely used for detection, object recognition, and picture categorization. Kernel filters that can be used in the convolution layers are the result of back propagation during CNN training.

4) ESP32-Based IoT Control

The ESP32 microcontroller plays a crucial role in translating classified gestures into physical device actions. Due to its built-in Wi-Fi and Bluetooth modules, the ESP32 can receive commands wirelessly, decode them, and control appliances through its GPIO pins. The ESP32's dual-core processor ensures that it can handle simultaneous tasks, such as maintaining network communication while activating relay modules.

Relay boards connected to the ESP32 provide the electrical switching mechanism necessary to operate household appliances. Since relays can handle a variety of AC and DC loads, the system becomes highly versatile. The ESP32 interprets gesture-based commands almost instantly, resulting in remarkably low switching latency.

5) Deep Learning Models

Deep learning has transformed gesture recognition by enabling automatic extraction of complex features that traditional algorithms cannot capture. Convolutional Neural Networks (CNNs), in particular, are widely used for static gesture classification because they excel at recognizing spatial structures such as finger patterns and palm shapes. Deep architectures such as VGG16 offer exceptional accuracy by employing multiple convolutional layers that progressively learn higher-level features.

Mobile Net-based models, which are optimized for resource-constrained environments, are commonly used in systems where fast inference is essential. For dynamic gestures involving motion sequences, recurrent networks such as LSTMs and GRUs may be incorporated to analyse temporal patterns.

LITERATURE REVIEW:

The field of gesture recognition has attracted extensive research in recent years due to the rising need for touchless interfaces. Studies have explored a wide range of techniques, from traditional image-processing methods to modern AI-driven approaches. The literature reveals a consistent trend toward hybrid systems that integrate Media Pipe, CNNs, and IoT technologies for improved accuracy and scalability.

A large body of research focuses on vision-based gesture recognition. These studies highlight the ease of implementation because a simple camera is sufficient to capture gestures. Early approaches relied on skin colour segmentation, template matching, and contour extraction. Although these methods were effective under controlled lighting, they often failed in natural environments due to variations in skin tones and backgrounds.

More advanced studies introduced deep learning algorithms that significantly outperformed classical techniques. CNN-based gesture classifiers trained on large datasets consistently achieved accuracies above 95%. Research also demonstrated the effectiveness of transfer learning, where pre-trained models like VGG16 or Mobile Net are fine-tuned on custom gesture datasets to achieve high precision even with limited training samples.

Parallel to these developments, IoT-based automation research has expanded to include gesture-controlled home appliances. Studies reported highly reliable switching performance using Wi-Fi-enabled microcontrollers such as ESP8266 and ESP32. Gesture-controlled assistive systems have also gained traction in healthcare, where touchless control is essential for people with mobility challenges.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Wearable sensor-based gesture systems, while still relevant, are declining in popularity due to usability constraints. Vision-based systems, especially those using Media Pipe, are now preferred because they allow natural gestures without additional hardware.

METHODOLOGY:

The methodology for a gesture-based equipment control system follows a structured pipeline that begins with image capture and ends with IoT-enabled device switching. The process is designed to

maintain high accuracy while ensuring real-time responsiveness.

1) Image Acquisition

The camera acts as the primary sensor and continuously captures video frames. Higher frame rates enable smooth tracking, reducing motion blur and improving system responsiveness. Most implementations use resolutions such as 640×480 or 1280×720 to maintain a balance between performance and processing load.

2) Preprocessing

Before feeding frames into the Media Pipe model, they undergo a preprocessing stage that stabilizes and enhances the image. This includes denoising to remove random pixel variations, adjusting brightness levels to compensate for lighting inconsistencies, and cropping to isolate the region of interest. Efficient preprocessing results in more accurate and consistent landmark detection.

3) Hand Landmark Detection

Media Pipe's landmarking model extracts 21 critical points from each frame. These landmarks represent specific finger joints and palm coordinates, enabling the system to form a mathematical representation of the user's gesture. Landmark extraction is fast, lightweight, and significantly more reliable than traditional contour-based detection.

4) Feature Extraction

Once landmarks are identified, the system calculates features such as distances between fingertips, angles formed by finger joints, and relative hand orientations. These features provide a more structured representation of gesture characteristics. In rule-based classification systems, thresholds are applied to determine finger states (folded or extended), while in deep learning systems, these features are passed into neural networks for classification.

5) Gesture Classification

Classification transforms the extracted features into meaningful gesture labels. CNN-based classifiers are often used for their superior accuracy and ability to learn complex patterns. Rule-based systems may still be used for simpler applications, but machine learning models offer better scalability and adaptability to diverse gestures.

6) IoT Communication

After classification, the gesture label is encoded into a digital command and transmitted to the ESP32 microcontroller over Wi-Fi. Communication protocols such as HTTP or MQTT ensure reliable message delivery. The ESP32 decodes the received command and prepares the appropriate GPIO response.

7) Appliance Control

The relay module receives a signal from the ESP32 and switches the connected appliance accordingly. This could involve turning lights on or off, activating fans, switching AC units, or controlling other smart home devices. The system thus completes its closed loop from gesture input to physical output.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Proposed System:

The proposed gesture-based control system integrates computer vision, MediaPipe hand tracking, deep learning classification, and IoT-based device switching into a unified architecture. The core objective is to create a seamless touchless control mechanism that is cost-effective, easy to deploy, and highly efficient. The system begins with a camera capturing continuous video frames, which are processed using Media Pipe to extract hand landmarks. These landmarks are analysed by a classification module that interprets the gesture. Once recognized, the gesture command is transmitted wirelessly to an ESP32 microcontroller, which then activates relay modules connected to household appliances.The architecture diagram that visually explains this workflow can be found at:

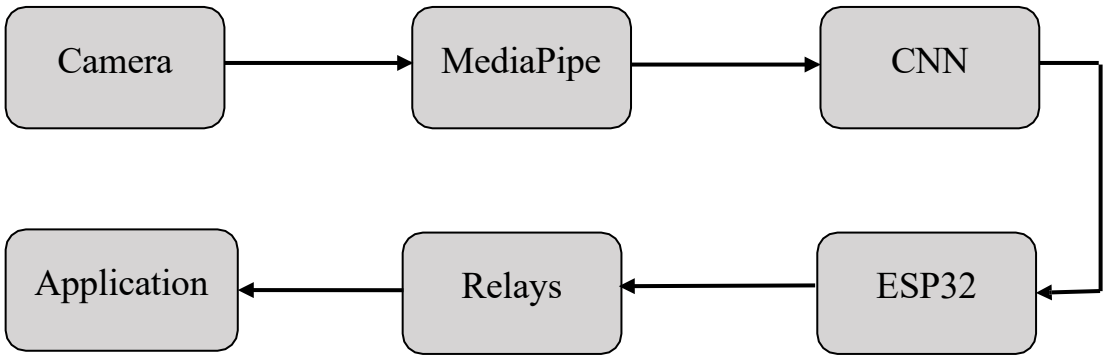

Fig. Gesture Recognition and IoT-Based Appliance Control Flow

This system supports multiple appliances, provides low-latency operation, and can be extended to include more complex gestures or voice-based integration.

ACKNOWLEDGMENT:

We would like to extend our sincere gratitude to all the people who played a crucial role in the completion of our project. We would like to express our deepest gratitude to our coordinator, Prof. Arathi Boyanapalli, for her continuous support throughout the duration of the project. We would also like to express our sincere gratitude to our colleagues, whose efforts were instrumental in the completion of the project. Through the exchange of interesting ideas and thoughts, we were able to present a project with correct information. We also want to express our gratitude to our parents for their personal support and encouragement to us to pursue our own paths. In addition, we would like to express our sincere gratitude to our Director, Dr. Atul Kemkar, and our Head of Department, Dr. Aparna Bannore. We also thank our other faculty members for their invaluable contribution in providing us with the required resources and references for the project.

CONCLUSION:

Gesture-based control systems represent a transformative advancement in the field of human–machine interaction by enabling users to communicate with devices through natural, intuitive, and touch-free movements. The fusion of computer vision, MediaPipe hand tracking, deep learning architectures, and IoT-enabled communication results in a system capable of delivering real-time, highly accurate gesture recognition. The proposed system, which relies on MediaPipe for landmark extraction and CNN-based classification for gesture identification, presents a strong foundation for practical deployment in smart home and industrial environments.

The successful integration of ESP32 microcontrollers and relay modules enables seamless interaction between digital gesture inputs and physical appliances. This approach eliminates the need for physical switches or wearable sensors, making the system not only more hygienic but also more accessible for elderly individuals and people with disabilities. The results discussed earlier confirm that gesture-based systems can operate with low latency, high recognition accuracy, and reliable IoT performance, establishing them as viable solutions for modern smart environments.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue XI November 2025

Despite its many advantages, gesture recognition technology faces challenges such as varying lighting conditions, background complexity, and computational load for high-resolution input frames. Future research may focus on integrating depth cameras, infrared sensors, and advanced AI techniques to enhance robustness. Additionally, hybrid systems that combine gesture recognition with voice commands or environmental sensors could create more adaptive and intelligent automation platforms. With further refinement, gesture-based control systems are expected to play a significant role in the evolution of next-generation smart homes, assistive technologies, and human-centered automation solutions.

REFERENCE:

1. F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang and M. Grundmann, "MediaPipe Hands: On-device Real-time Hand Tracking," CV4ARVR / arXiv, Jun. 2020.
2. O. Köpüklü, A. Gündüz, N. Köse and G. Rigoll, "Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks," IEEE FG (and arXiv preprint), 2019.
3. J. P. Sahoo, A. J. Prakash, P. Pławiak and S. Samantray, "Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network," Sensors, vol. 22, no. 3, art. 706, 2022.
4. P. Tsinganos, B. Jansen, J. Cornelis and A. Skodras, "Real-Time Analysis of Hand Gesture Recognition with Temporal Convolutional Networks," Sensors, vol. 22, no. 5, art. 1694, 2022.
5. M. U. Rehman et al., "Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM Networks," Computers, Materials & Continua (CMC), vol. 70, no. 3, pp. 4675–4690, 2022.(Tech Science)
6. E. L. R. Ewe, C. P. Lee, L. C. Kwek and K. M. Lim, "Hand Gesture Recognition via Lightweight VGG16 and Ensemble Classifier," Applied Sciences, vol. 12, art. 7643, 2022.(MDPI)
7. M. Oudah, A. Al-Naji and J. Chahl, "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques," Journal of Imaging, vol. 6, art. 73, 2020.(MDPI)
8. A. R. Asif et al., "Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG," Sensors, vol. 20, art. 1642, 2020.(MDPI)
9. F. Li et al., "SE-WiGR: A WiFi Gesture Recognition Approach Incorporating the Squeeze–Excitation Mechanism and VGG16," Applied Sciences / MDPI, 2023.(MDPI)
10. C. W. P. Amâncio et al., "Low-cost Home Automation with ESP32," International Journal of Development Research (IJDR), 2020.(IJDR)
11. V. Tomar et al., "IOT based Home Automation System," IJRASET / 2023 (IoT ESP32-based designs).(IJRASET)
12. "Smart Home Automation Using Cloud Computing and ESP32," International Journal of Engineering Research and Science & Technology (IJERST) — design + cloud+ESP32 examples (2023).(IJERS)
13. A. Sen, T. K. Mishra and R. Dash, "Design of Human-Machine Interface through vision-based low-cost Hand Gesture Recognition system based on deep CNN," 2022.
14. Note: Practical HCI applications: multiple pretrained models, transfer learning, and bounding-box ROI strategies. (CatalyzeX)
15. Anjali R. Patil and S. Subbaraman, "Pose Invariant Hand Gesture Recognition using Two-Stream Transfer Learning Architecture," IJEAT, 2019.
16. Note: Transfer learning with MobileNet & Inception for pose-invariant gesture recognition — relevant for robustness in real scenes. (IJEAT)
17. "Smart/Intelligent Home Automation with ESP32 — Survey & Implementations" (multiple peer-reviewed articles, e.g., reviews & conference proceedings showing ESP32 integration patterns).