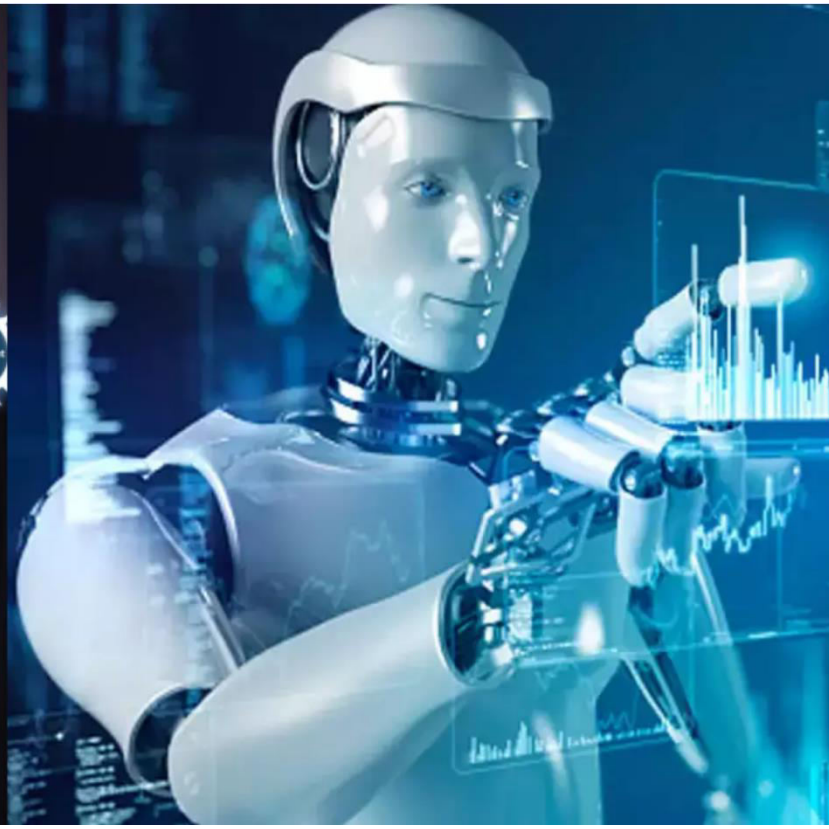
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 8, August 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408097|

Vision-Based Media Controls using MediaPipe and OpenCV

Nishchitha D, Hemanth Kumar

P.G. Student, Department of MCA, JNN College of Engineering, Shivamogga, Karnataka, India

Associate Professor, Department of MCA, JNN College of Engineering, Shivamogga, Karnataka, India

ABSTRACT: Nowadays controlling media by physical devices like keyboards and mouse has become less efficient and less practical. This work introduces a contactless, real- time media controlling system that allows users to interact with a media player through static, dynamic hand gestures. This work utilizes MediaPipe framework for live hand tracking, landmark detection, and OpenCV library handles video pipeline. The system reads dynamic gestures, like hand movements from left - right or right - left, used to control media player navigation (forward and backward) whereas static gestures such as fist, open palm, thumbs up and thumbs down are used to manage introductory media functions like play, pause, volume increase, volume decrease. By utilizing Flask as a web interface, system enables smooth interaction between browser- grounded media playback and gesture recognition segments. This proposed system minimizes the necessity of physical contact, thereby technology enhance availability, accessibility and ease in human-computer interaction.

KEYWORDS: Gesture-Based Media Control, Static, Dynamic, OpenCV.

I. INTRODUCTION

From a various year, the human-computer interaction has mostly depended on physical input devices like keyboards, mouse, or touchscreen boundaries. Even these input techniques work well in variety of settings, they may not be suitable in appropriate condition where is a lack of free hand to operate an input device, such as in giving a presentation, cooking, hygienic settings like labs or hospitals and in hand free environment etc. Controlling system by hand gesture have become a feasible substitute in recent years, giving the consumers the opportunity of a compatible and normal way to interact with digital systems.

This work introduces a gesture-based media player that allow us to execute media actions using a combination of static hand gestures (e.g. fist, palm, thumbs-up, down) and dynamic hand gestures (such as the left, right swipe). Objective is that the users should be able to manage playback actions like play, pause, volume up, and volume down, and video navigation actions like forward and rewind, without using physical input devices.

This work capture video from webcam, perform frame processing BGR-RGB for natural interaction, draw overly and JPEG encoding for streaming with the help of OpenCV library of computer vision. 21 landmark relevant points like wrist, knuckles, fingertips are detected from MediaPipe. Based on distance, angles, short-term motion of landmarks, rule-based module labels gestures like fist, palm, thumbs up, thumbs down and swipes left or right.

The whole gesture recognition process is incorporated within a flask browser application to make this technology accessible and to use easily. Flask is integrated python web framework that may use to create full-stack apps and APIs. In our system flask provides browser with a live webcam feed, it gives the most recent gesture that was detected as a JSON response, and provides a web interface (.html) that permits users to manage video player by hand gestures, monitor the camera stream. Using JavaScript frontend polls the gesture API at every second and applies the corresponding actions like play, pause, control volume etc. to the HTML5 video elements based on gesture received.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408097|

II. LITERATURE SURVEY

In their comprehensive evaluation of hand gesture recognition techniques, Munir Oudah et al. primarily absorbed on computer vision methods [1]. Their research pushes the limits of contactless commerce systems by dividing recognition techniques into two categories: vision-based and glove-based. They went over several methods that are comparable, including stir shadowing, appearance-grounded recognition, skin colour recognition, shell-grounded models, and deep literacy infrastructures. K Harini has proposed a hand gesture identification system capable of recognizing static and dynamic gestures using DCNN and Self-Organizing Map (SOM) [2]. This model successfully detects the gesture area by applying preprocessing methods involving skin color filtering, Hessian-based multiscale filtering, and a transformed marker-controlled watershed algorithm. Yuting Meng developed a live hand gesture tracking system using the registerable technology of MediaPipe [3]. The method utilizes Finger Net, which is an upgraded ResNet, and a new-built dataset named RGDS to process and fuse gesture information. Variability and personalization in gesture recognition are achieved through the system's ability to track user gestures. Through applying Triple Loss and CrossEntropy Loss for better accuracy and integration while training, the paper proposes a new methodology. On experimental results, the model successfully classified gestures with different postures and orientations, especially with a very well-curated dataset. The authors in [4] proposed a multimodal gesture recognition system based on CNN models and HCM to combine temporal and pose information. This architecture employs RGB-D and skeleton information obtained from Kinect sensor, combined with CNNs for pose structure extraction and HCM for motion outline extraction. Abdullah Mujahid et al. With the use of the YOLOv3 deep learning model, which is implemented based on the DarkNet-53 architecture, a real-time hand gesture detection system was created [5]. The system is able to efficiently handle both static images and video inputs, without requiring any complex preprocessing steps. Yet, the real-time capability of the system can still be affected by the capabilities of the used hardware. It was tested with a small dataset containing 216 images that were self-collected and in YOLO format. The TinyYOLOv3 deep learning network is utilized by Abhilash Dayanandan et al. to detect concurrent hand gestures in their gesture-controlled media player system [6]. The system records hand motion through a webcam, which are pre-processed and classified using a trained YOLO model. It is able to identify hand gestures to control media controls including play, pause, volume, fast forward, rewind, and mute. Anklesh G. A created leverages real-time webcam input to control a video player through the recognition of gestures [7]. The system was implemented with Python and the OpenCV, Mediapipe, and Pyautogui packages to perform play, pause, rewind, and fast-forward functions. The model provides a hands-free media playback interface by interpreting hand landmarks to recognize movements and translate them into keyboard inputs. Ji Hwan Kim et al. used a modified KHU 1 data glove, incorporating three tri-axial accelerometers and Bluetooth wireless transmission, to implement a 3D hand motion capture and gesture recognition system [8]. Toward mimic finger joints and phalanges, they constructed a kinematic digital hand model that consisted of chain-linked ellipsoids. Then, they utilized rule-based algorithms to transform the input accelerometer data into joint rotations. Granit Luzhnica et al. used data glove protocol with motion, pressure, and bending sensor was employed to build a system that can identify natural hand gestures [9]. They have derived features from time and frequency domains and used a sliding window approach to handle real-time sensor data. K Harini has proposed a hand gesture identification system capable of recognizing static and dynamic gestures using DCNN and Self-Organizing Map (SOM) [9]. This model successfully detects the gesture area by applying preprocessing methods involving skin color filtering, Hessian-based multiscale filtering, and a transformed marker-controlled watershed algorithm. Moreover, Sakshi Shinde designed a media playback controllable gesture controller that can control VLC Media Player [10]. In order to avoid dependency on input devices such as mice and keyboards, the system uses OpenCV, MediaPipe, and PyAutoGUI algorithms for recognizing hand gestures. Certain functions such as play, pause, volume, and navigation over tracks are governed by pre-defined gestures. Shruti Tibhe created live gesture-based recognition system with computer vision methodology, for managing media player attributes [11]. By OpenCV, MediaPipe, and PyAutoGUI, the system allows users to control functions such as volume, play, pause, and video navigation. Touchless interaction way suitable for physical limitation people, this project promotes accessibility. To communicate with media players without extra hand-worn devices, Maryam Saleem demonstrated gesture recognitions [12]. Traditional camera and vision-based image processing algorithm used to detect, perform hand motions. Modules for finger counting, skin detection, contour extraction, convex hull drawing to identify motions. OpenCV and C# are used in the program's construction within Visual Studio. Priyadarshini Kannan created system, controls media players with hand gestures without use of physical input devices, [13]. To identify and categorize real-time gestures by SqueezeNet CNN. Two essential tools are PyAutoGUI and OpenCV. The emphasis of this work is on intuitive interaction and accessibility, which is particularly advantageous for disabled users and public spaces. Raimundo F. Pinto et al. [14]. proposed A CNN-based model for identifying static gestures. This method employs a preprocessing pipeline that includes division, polygonal estimation, contour detection,

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408097|

morphological filtering to enhance feature quality prior to classification. CNN is then trained using corrected images, and cross-validation used to assess its performance. The system could only move in static motion, though. In their research of HGR techniques, Yashas J and Shivakumar G [15] contrasted camera-based techniques, which permit natural interaction but have drawbacks like dim lighting, background noise, and ROI selection, with sensor-based techniques, which limit hand movement. Some of these problems are lessened by devices like Microsoft Kinect. CNN, RNN, and deep learning models are highlighted in the study, with an emphasis on Adapting CNNs (ADCNN), which use data augmentation to increase classification accuracy. The authors highlight data augmentation as a promising but little-studied method for improving model robustness.

III. METHODOLOGY

This section describes the method used for gesture-based media controllers. The system records live video from webcam utilizing OpenCV, detects 21 landmarks for each frame with MediaPipe Hands, figures out the gesture using straightforward geometry, fast motion checks, and deployment through a web interface built with Flask.

3.1 Live capture and Feedback

The webcam records frames in real-time, to keep the latency low so user can view themselves almost instantly on web stream. Frames are processed and returned quickly as an MJPEG image, creating a seamless video experience. Along with the stream, a light API conveys the latest detected action so that the user interface can react (e.g., Play) regardless of the frame rate. This separation video for visual and JSON for control keeps the interface responsive and predictable.

3.2 Landmark detection and normalization

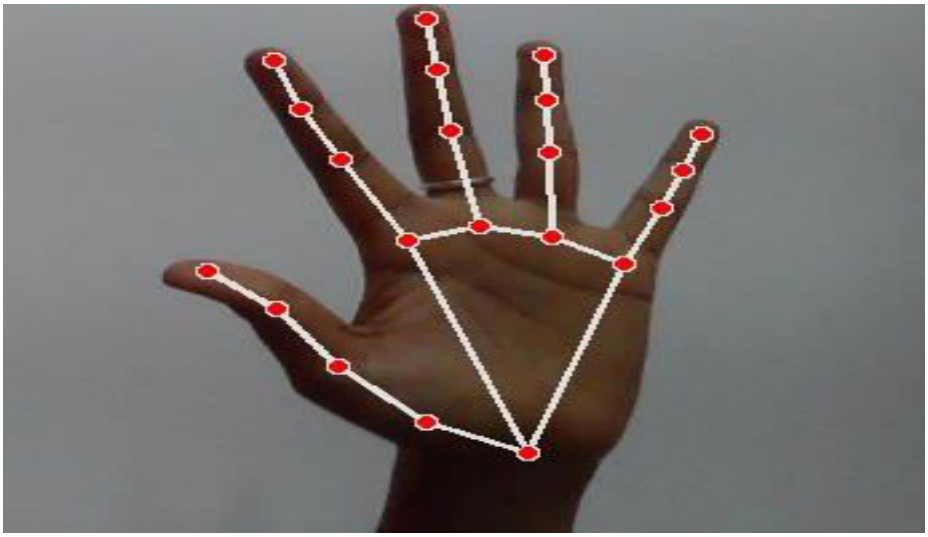


Figure 1: Hand Landmarks Detection

Figure 1, shows Every frame is processed with MediaPipe Hands to get 21 key points that represents positions of fingers, joints, wrist. Raw pixel coordinates are subsequently normalized to eliminate effects of distance, camera angle using relative distances, angles, ratios are referenced to the hand's consistent bone length. This process transforms real images into uniform geometric representation. The same rules will apply whether the hand is far or close to the camera, thus maintaining the robustness of system across many users and lighting setups.

3.3 Rule-based gesture decision

Static gestures are detected by applying number of geometric constraints to normalized landmarks e.g., thresholds on finger curl, thumb direction, palm open. Instead of firing on one frame, the system requires a short cooldown time to prevent successive firings. These time-based protections prevent jitter and false positives. The approach is transparent, every decision can be traced back particular, understandable criteria.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408097|

3.4 Action mapping

When a gesture is confirmed, it is translated into command like play, pause, volume_up, or volume_down. The backend delivers the command through a small endpoint the frontend checks periodically, and it is reset immediately after delivery to guarantee one action equals one trigger. This trigger pattern avoids chance of repetition and ensures deterministic media control. Different modes (static vs. dynamic) can be mapped to different sets of commands without requiring changes to the user interface.

3.5 Visual feedback

Figure 2: Visual Feedback

Figure 2 shows the processed frame is enhanced with indicators like hand skeleton lines, followed landmarks, and current label, allowing for instant user feedback and adjustment. The live intersection helps user to understand how to pose and makes threshold tuning during testing easier, without any additional instrumentation and reduces learning, users quickly modify their hand position as interface demonstrates, in real time, how their movements are being interpreted.

IV. EXPERIMENTAL RESULTS

The proposed work effectively outfits a simultaneous recognizing gesture by OpenCV and MediaPipe. Static and dynamic gestures as custom logic.

Static Gestures: fist, palm, thumbs_up, thumbs_down are static gesture used to control media player. The system worked smoothly once hand was stabilized and centred. The short-lived hold and handle logic successfully removed almost all lag, thumb gestures were smooth after distinct up, down direction was established.

Dynamic Gestures: Swipe Left and Swipe Right are dynamic gesture used to control media player. The short motion window ignored slight lag slanted, avoided continuous triggers.

Mode Switching: switching between static and dynamic is done in this proposed work. The system reset its short history on switching, thereby avoiding any cross-mode misfires. The on-screen markers and captions helped people correctly place there-own and understand every action.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408097|

Static Hand Gesture

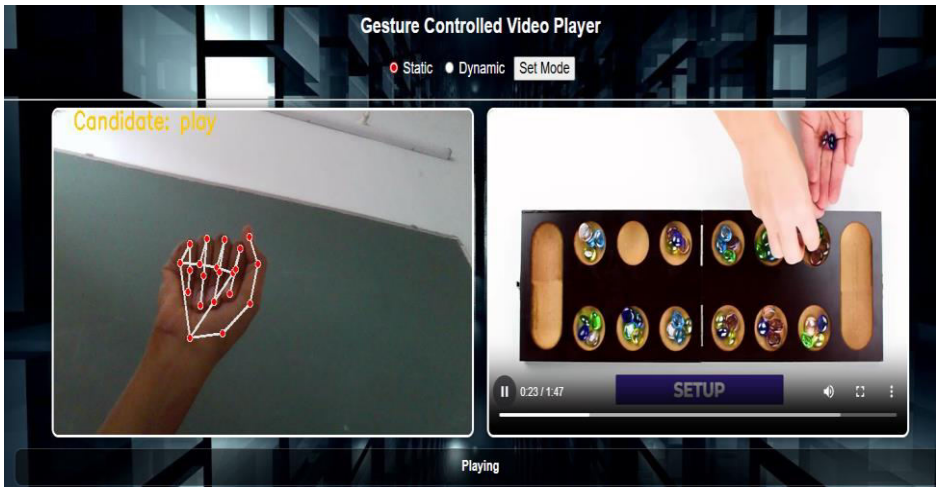


Figure 3: Play Video

Figure 3, shows two sections, on the left side, a live webcam feed illustrates hand landmarks marked with red points and lines, accompanied by play gesture label. On the right side, an integrated video player is operational with standard controls displayed. Static mode is chosen and can be activated through Set Mode, while the footer status verifies the current condition with Playing.

Dynamic Hand Gesture

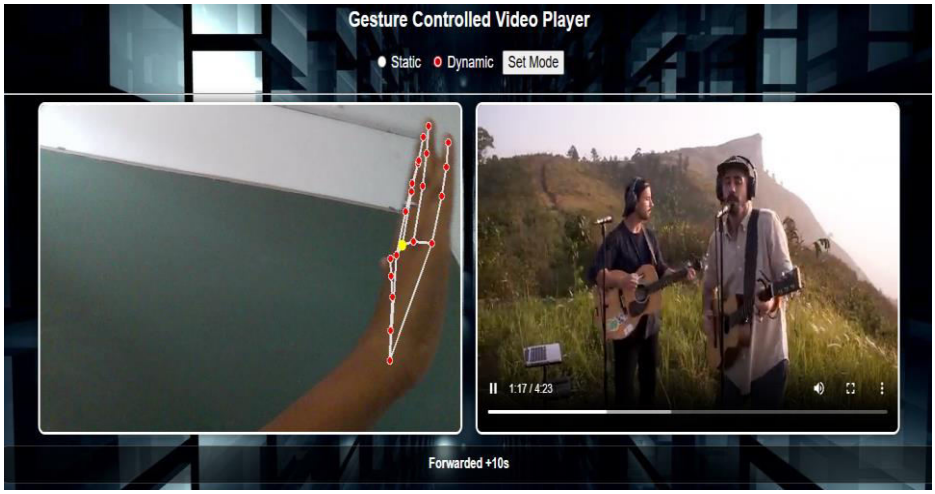


Figure 4: Forward Video

Figure 4, shows dynamic mode. On the left, the webcam stream overlays red hand landmarks, the angled skeleton indicates a motion gesture was tracked. On the right, the video continues playing. The footer message forwarded +10s confirms a right-swipe gesture triggered a 10-second seek forward.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 8, August 2025

|DOI: 10.15680/IJIRSET.2025.1408097|

V. CONCLUSION

This work successfully determines a contactless control of media player by hand gestures recognition system. Combining MediaPipe for real-time hand tracking, OpenCV for image processing, the system exactly knows combination of static and dynamic gestures to control media playback features like play, pause, volume adjustment, and navigation. The convention of Flask and JavaScript permits seamless browser-based interaction and visual feedback connections increase usability and user assurance. By rejecting the dependence on physical input devices like keyboard and mouse, this is the method which enhances both accessibility and convenience mostly in situations where hand usage is limited or unusable. This work bridges the gap between natural human gestures and digital media control, provided that a scalable platform intended for future advances in gesture-driven interfaces and smart surroundings.

REFERENCES

1. P. Maloo, D. Deepa, "Dynamic Hand Gesture Recognition Using 3DCNN and LSTM with FSM Context-Aware Model", Computer Modeling in Engineering & Sciences (CMES), Vol. 138, No. 3, pp. 2629–2652, 2024
2. Harini K., "A novel static and dynamic hand gesture recognition using self-organizing map with deep convolutional neural network", Automatika, Vol. 64, No. 4, pp. 1128–1140, 2023.
3. Meng Y., "Real-Time Hand Gesture Monitoring Model Based on MediaPipe's Registerable System", Sensors, Vol. 24, No. 19, pp. 6262, 2024.
4. Escobedo Cardenas E.J., Camara Chavez G., "Multimodal hand gesture recognition combining temporal and pose information based on CNN descriptors and histogram of cumulative magnitudes", Journal of Visual Communication and Image Representation, Vol. 71, No. –, pp. 102772, 2020
5. Mujahid A., Awan M.J., Yasin A., Mohammed M.A., Damaševičius R., Maskeliūnas R., Abdulkareem K.H., "Real-time hand gesture recognition based on deep learning YOLOv3 model", Applied Sciences, Vol. 11, No. 9, pp. 4164, 2021
6. Dayanandan A., Chakkungal A., Kommeri A., Koppuliparambil D., Nitnaware P., "Gesture controlled media player using TinyYoloV3", International Research Journal of Engineering and Technology (IRJET), Vol. 7, No. 5, pp. 3006–3011, 2020
7. Anklesh G., "Hand Gesture Recognition for Video Player", International Research Journal on Advanced Engineering Hub (IRJAEH), Vol. 2, No. 4, pp. 801–805, 2024.
8. Kim, J. H., Thang, N. D., & Kim, T. S. (2009). 3 D hand motion tracking and gesture recognition using a data glove. In Proceedings of the IEEE International Symposium on Industrial Electronics (ISlE 2009) (pp. 1013–1018)
9. Granit Luzhnica, Jörg Simon, Elisabeth Lex, Viktoria Pammer," A Sliding Window Approach to Natural Hand Gesture Recognition using a Custom Data Glove", Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA), pp. 1–8, 2016
10. Shinde S., "Gesture Based Media Player Controller", International Journal of Research Publication and Reviews, Vol. 3, No. 5, pp. 2289–2294, 2022.
11. Tibhe S., "Media Controlling Using Hand Gestures", International Journal of Creative Research Thoughts (IJCRT), Vol. 11, No. 12, pp. 139–144, 2023.
12. Saleem M., "Interaction with Media Players through Hand Gesture Recognition", Proceedings of the International Conference on Communication Technologies (ComTech), Vol. –, No. –, pp. 1–6, 2023.
13. Kannan P., "Automated Media Player using Hand Gesture", International Research Journal of Engineering and Technology (IRJET), Vol. 10, No. 4, pp. 1466–1472, 2023.
14. Pinto R.F., Borges C.D.B., Almeida A.M.A., Paula I.C., "Static hand gesture recognition based on convolutional neural networks", Anais do XIX Workshop de Visão Computacional (WVC), Vol. –, No. –, pp. 58–63, 2019.
15. Yashas J., Shivakumar G., "A literature survey on hand gesture recognition techniques", International Journal of Engineering Research & Technology (IJERT), Vol. 9, No. 13, pp. 1–4, 2021.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details