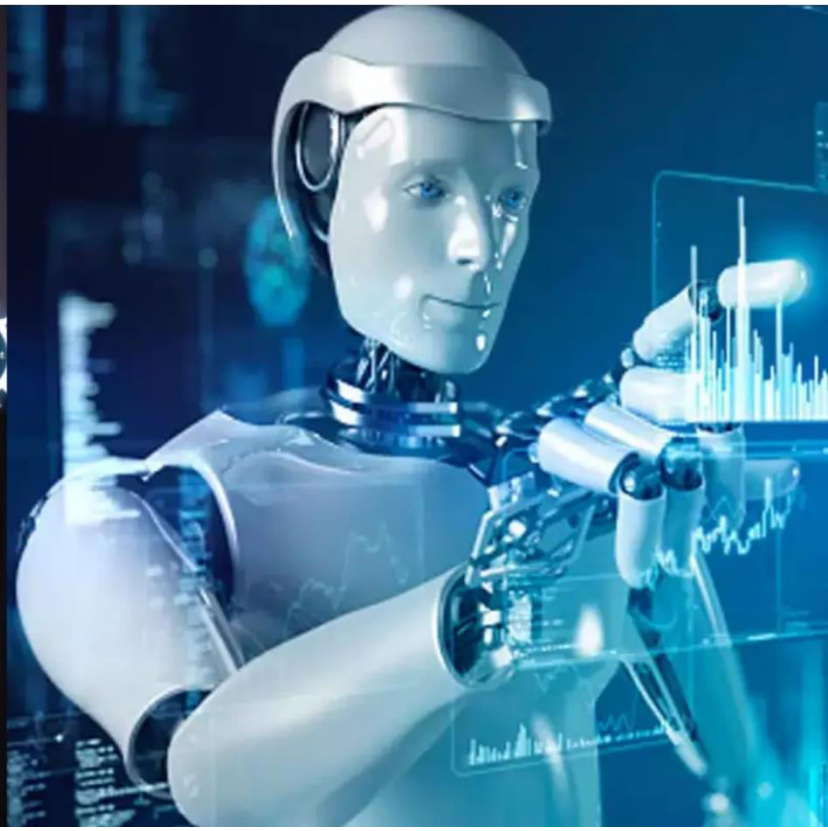
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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 15, Issue 4, April 2026

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 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

Virtualink: A Comprehensive Survey on Real-Time Gesture Recognition Systems using Artificial Intelligence and Computer Vision

Prof. A. Ghadge[1], Sarthak Jagtap[2], Surabhi Jawalekar[3], Samruddhi Awate[4], Gajanan Bondre[5]

Professor, Dept. of Computer Engineering, AISSMS Polytechnic, Pune, India[1]

Students, Dept. of Computer Engineering, AISSMS Polytechnic, Pune, India[2345]

ABSTRACT: Gesture recognition has emerged as a crucial domain in Human-Computer Interaction (HCI), enabling seamless communication between humans and machines without the need for traditional input devices. Conventional interfaces such as keyboards, mice, and touchscreens impose limitations in terms of flexibility, accessibility, and user experience. This survey paper presents an in-depth study of real-time gesture recognition systems based on Artificial Intelligence (AI) and Computer Vision techniques. It explores various approaches including OpenCV-based image processing, MediaPipe-based hand tracking, and deep learning-based gesture classification. The paper reviews existing research works, compares methodologies, and highlights key challenges such as environmental sensitivity, computational complexity, and accuracy limitations. Furthermore, it discusses applications in education, healthcare, gaming, and virtual environments. The study concludes that systems like Virtualink provide an efficient, contactless, and interactive solution for modern digital interaction, while future advancements can enhance performance using deep learning and augmented reality integration.

KEYWORDS: Gesture Recognition, Computer Vision, Artificial Intelligence, OpenCV, MediaPipe, Human-Computer Interaction, Virtual Whiteboard, Hand Tracking, Real-Time Systems.

I. INTRODUCTION

Human-Computer Interaction (HCI) has undergone significant transformation over the past few decades. Traditional interaction techniques rely heavily on physical input devices such as keyboards, mice, and touchscreens. Although these devices are widely used, they present several limitations including restricted mobility, lack of natural interaction, and reduced accessibility for physically challenged individuals.

Gesture recognition offers a promising alternative by enabling users to interact with digital systems through natural hand movements. It eliminates the dependency on physical devices and allows a more intuitive and immersive user experience. Gesture-based systems utilize cameras, sensors, and machine learning algorithms to detect and interpret human gestures in real time.

Recent advancements in Artificial Intelligence (AI) and Computer Vision have significantly improved the performance of gesture recognition systems. Technologies such as OpenCV and MediaPipe provide efficient frameworks for image processing and hand tracking. These tools allow real-time detection of hand landmarks, enabling accurate gesture interpretation.

Systems like Virtualink demonstrate the practical implementation of gesture recognition by allowing users to draw, write, and interact with digital content using hand gestures. Such systems have applications in education, remote collaboration, healthcare, and entertainment.

Despite these advancements, challenges such as varying lighting conditions, background complexity, and computational requirements still affect system performance. This survey paper provides a detailed analysis of gesture recognition systems, their methodologies, applications, and future scope.

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 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

II. RELATED WORK

Several researchers have contributed to the development of gesture recognition systems:

Saurabh Uday Saoji (2021) developed an air canvas application using OpenCV and NumPy, enabling users to draw using finger movements. The system demonstrated the feasibility of gesture-based drawing but faced limitations in accuracy under varying lighting conditions.

Tamalampudi Hemaa Chandhan et al. proposed a hand tracking system using OpenCV and MediaPipe. Their approach utilized 21 hand landmarks for precise gesture recognition, significantly improving accuracy and performance.

Shreyas Sandbhor et al. presented a survey on air canvas systems, highlighting various gesture recognition techniques and their applications in digital interaction.

Mitesh Ikar et al. developed a computer vision-based virtual paint system that allows real-time drawing using gestures. The system improved user interaction but required high computational resources.

B. Venugopal et al. proposed an air canvas application focusing on gesture-based drawing and user-friendly interface design.

Recent research also explores deep learning-based approaches, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which enhance gesture classification accuracy but require large datasets and computational power.

These studies indicate that gesture recognition systems are evolving rapidly, with increasing focus on accuracy, efficiency, and real-time performance.

III. DATABASE

In gesture recognition systems, the input data plays a critical role in determining the overall accuracy, efficiency, and responsiveness of the system. Unlike conventional machine learning models that rely on static datasets, the proposed Virtualink system operates on real-time video input captured through a webcam. This dynamic nature of input eliminates the dependency on pre-collected datasets and allows the system to function interactively.

The system continuously captures video frames using a standard webcam, where each frame acts as an individual image for processing. These frames contain visual information about the user's hand gestures, including movement, position, orientation, and shape. The captured frames are processed sequentially to maintain real-time performance and ensure smooth interaction between the user and the system.

The MediaPipe framework is used for extracting meaningful features from the input frames. It detects and tracks 21 key landmarks on the human hand, including fingertips, joints, and the palm base. These landmarks provide a structured representation of the hand, which is essential for accurate gesture interpretation. The use of landmark-based detection reduces the complexity of image processing and improves system performance.

Since the system does not rely on pre-labeled datasets, it simulates a real-time dataset environment where data is continuously generated and processed. This approach provides flexibility and adaptability, making the system suitable for various real-world applications.

However, the performance of the system depends on certain environmental conditions. Proper lighting is required to ensure clear visibility of the hand. A clutter-free background helps in reducing noise and improving detection accuracy. Additionally, the positioning of the camera should be stable to avoid fluctuations in input data.

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 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

To enhance robustness, the system may incorporate preprocessing techniques such as frame normalization, noise reduction, and smoothing. These techniques help in improving the quality of input data and ensure consistent performance across different conditions.

Overall, the Virtualink system effectively utilizes real-time input data to perform gesture recognition, making it a practical and efficient alternative to traditional dataset-based approaches.

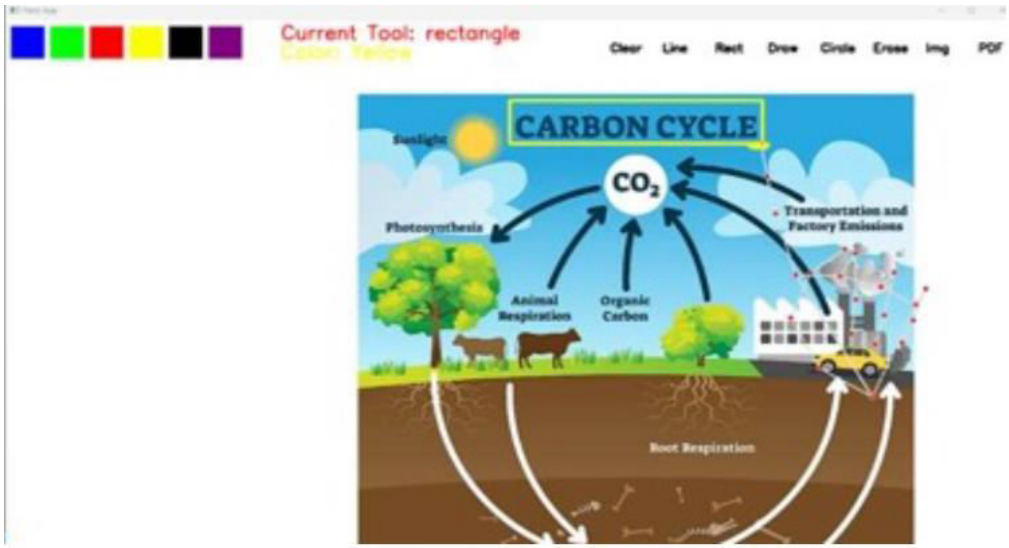
IV. PROPOSED METHODOLOGY

The proposed system, "Virtualink: Real-Time Gesture Recognition," is designed to provide a touchless and intuitive interface for interacting with a digital canvas. The system integrates computer vision and artificial intelligence techniques to detect, track, and interpret hand gestures in real time.

The methodology follows a structured pipeline consisting of multiple stages, each contributing to the overall functionality and accuracy of the system. These stages include image acquisition, preprocessing, hand detection, feature extraction, gesture recognition, and output rendering.

Test Images

The performance of the proposed Virtualink system was evaluated using real-time test images captured through the webcam. These test images represent different hand gestures performed by users under varying conditions such as lighting, background, and hand orientation. The test samples include gestures such as drawing using a single finger, selection using two fingers, and erasing using a closed fist. These gestures are used to validate the system's ability to accurately detect and classify hand movements.



Comparison between Traditional Input Methods and Gesture-Based Systems

Traditional human-computer interaction relies on input devices such as keyboards, mice, and touchscreens. While these methods are widely used, they require physical contact and limit the flexibility of user interaction.

In contrast, gesture-based systems provide a touchless and more natural way of interacting with digital systems. These systems use computer vision techniques to detect and interpret hand movements in real time, eliminating the need for physical input devices.

In traditional systems, user interaction is restricted to predefined input mechanisms, whereas gesture-based systems allow intuitive and dynamic interaction. For example, in the Virtualink system, users can draw, erase, and select tools using simple hand gestures without any physical contact.

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 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

Additionally, gesture-based systems improve accessibility, especially for individuals with physical disabilities, by providing an alternative mode of interaction. However, these systems may face challenges such as dependency on lighting conditions and camera quality.

The comparison indicates that gesture recognition systems offer a more interactive, flexible, and modern approach compared to traditional input methods, making them suitable for future human-computer interaction applications.

Features	Traditional Systems	Gesture-Based Syste,
Input Method	Keyboard, Mouse	Hand Gestures
Interaction	Physical	Touchless
Flexibility	Limited	High
Accessibility	Moderate	Moderate
Accuracy	low	Depends on environment

IV. SYSTEM ARCHITECTURE AND WORKING

The proposed system, "Virtualink: Real-Time Gesture Recognition," is designed to enable intuitive and touchless interaction between humans and digital systems using computer vision and real-time processing techniques. The architecture of the system is developed as a modular pipeline that efficiently processes continuous video input, extracts meaningful features, and converts user gestures into corresponding actions on a virtual canvas.

The system is structured into multiple stages, each responsible for a specific function, including data acquisition, preprocessing, hand detection, feature extraction, gesture recognition, and output rendering. These stages work sequentially as well as collaboratively to ensure real-time performance with minimal latency and high accuracy.

A. Overall System Architecture

The overall architecture follows a real-time streaming model, where video frames are continuously captured and processed in a loop. The pipeline begins with the acquisition of image data through a webcam and ends with rendering the output on a virtual interface.

The system is designed to be lightweight and efficient, avoiding computationally expensive deep learning models while maintaining acceptable accuracy. The use of MediaPipe for hand tracking enables efficient landmark detection without requiring GPU-intensive operations.

Each module in the architecture is loosely coupled, allowing flexibility for future enhancements such as integration with deep learning models, cloud processing, or augmented reality systems. The architecture ensures scalability and adaptability across different platforms.

B. Image Acquisition and Frame Generation

The first stage of the system involves capturing real-time video input using a webcam. The camera continuously streams frames at a predefined frame rate, typically ranging between 20 to 30 frames per second, to ensure smooth interaction.

Each frame is treated as an independent input sample and is passed to subsequent processing stages. The resolution of the captured frames is adjusted to balance between computational efficiency and detection accuracy.

The consistency of frame acquisition is crucial, as any delay or frame drop may affect the responsiveness of the system.

C. Image Preprocessing and Enhancement

The captured frames undergo preprocessing to improve their quality and suitability for further analysis. This stage includes resizing, normalization, and noise reduction.

Resizing standardizes the input dimensions, which simplifies processing and ensures uniformity across frames. Normalization adjusts pixel intensity values to a consistent scale, improving the stability of detection algorithms.

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 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

Noise reduction techniques, such as smoothing and filtering, are applied to minimize the impact of environmental disturbances like shadows, reflections, and background clutter. This stage enhances the reliability of hand detection and reduces false positives.

D. Hand Detection Using MediaPipe Framework

Hand detection is a critical stage in the system, where the region of interest (ROI) containing the hand is identified within the frame. The system uses the MediaPipe Hands module, which employs machine learning-based models optimized for real-time performance.

MediaPipe detects the hand and provides bounding box coordinates along with a set of predefined landmarks. The detection process is highly efficient and capable of handling variations in hand orientation, scale, and position.

The framework ensures robust detection even in moderately complex backgrounds, making it suitable for practical applications.

E. Hand Tracking and Temporal Consistency

Once the hand is detected, the system tracks its movement across consecutive frames. Tracking ensures temporal consistency, allowing the system to maintain continuity in gesture recognition.

This stage is essential for dynamic gestures such as drawing or writing, where motion over time needs to be accurately captured. The tracking algorithm minimizes jitter and ensures smooth transition of hand positions across frames.

Efficient tracking contributes to a seamless user experience by reducing lag and maintaining synchronization between user actions and system responses.

F. Feature Extraction through Landmark Detection

Feature extraction is performed by identifying 21 key landmarks on the hand using the MediaPipe framework. These landmarks correspond to important anatomical points such as fingertips, joints, and the wrist.

Each landmark is represented by its spatial coordinates (x, y, z), forming a structured representation of the hand. This representation captures both the geometric and positional characteristics of the hand.

By focusing on landmark-based features, the system reduces computational complexity while preserving essential information required for gesture recognition. This approach is more efficient compared to processing raw pixel data.

G. Gesture Recognition and Decision Logic

The gesture recognition process involves analyzing the spatial relationships between the extracted landmarks. The system uses rule-based decision logic to classify gestures into predefined categories.

For example, the relative positions of fingertips with respect to the palm are used to determine whether fingers are extended or folded. Based on this analysis, gestures are classified into modes such as drawing, selection, and erasing.

The rule-based approach ensures low latency and high efficiency, making it suitable for real-time applications. However, it can be extended with machine learning models in future implementations for improved accuracy and adaptability.

H. Mode Selection and Control Mechanism

Based on the recognized gesture, the system activates a specific operational mode. The control mechanism ensures that only one mode is active at any given time, preventing conflicts during interaction.

The primary modes include:
Drawing Mode: Allows users to draw on the canvas using fingertip movement
Erasing Mode: Enables removal of previously drawn content
Selection Mode: Used for choosing tools, colors, or options

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 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

The transition between modes is designed to be smooth and responsive, ensuring an intuitive user experience.

I. Output Rendering and Visualization

The final stage of the system involves rendering the output on a virtual canvas. The system translates recognized gestures into graphical actions such as drawing lines, shapes, or erasing content. The rendering is performed in real time, ensuring immediate feedback to the user. The virtual canvas serves as the interface where all interactions are visualized. Efficient rendering algorithms are used to maintain smooth performance and minimize latency, even during continuous gesture input.

J. Working Principle of the System

The working principle of the Virtualink system is based on continuous real-time processing of visual data. The system captures video frames, preprocesses them, detects and tracks the hand, extracts relevant features, and classifies gestures to generate corresponding outputs.The entire process operates in a cyclic loop, enabling continuous interaction between the user and the system. As the user performs gestures, the system instantly interprets them and updates the output accordingly.The performance of the system is influenced by factors such as processing speed, environmental conditions, and accuracy of detection algorithms. With optimized implementation using OpenCV and MediaPipe, the system achieves a balance between computational efficiency and real-time responsiveness.

Model Training

The training process of the proposed Virtualink gesture recognition system is designed with a focus on achieving high accuracy, real-time responsiveness, and computational efficiency. Unlike conventional gesture recognition systems that rely heavily on custom-trained deep learning models, the proposed system adopts a hybrid approach that combines pre-trained machine learning models with rule-based gesture classification. This design choice significantly reduces the need for extensive dataset preparation and training time while maintaining reliable performance in real-time applications.

At the core of the system, the MediaPipe Hands framework is utilized, which incorporates machine learning models that have been pre-trained on large-scale hand gesture datasets. These models are capable of detecting and tracking hand landmarks with high precision under varying conditions, including changes in hand orientation, scale, and moderate background complexity. The pre-trained nature of the model eliminates the requirement for explicit training within the system while still providing robust feature extraction capabilities. The framework internally employs convolutional neural network-based pipelines for palm detection and landmark estimation, which have been optimized for real-time performance on CPU-based systems.

Although the system does not involve traditional end-to-end supervised training, a significant amount of effort is dedicated to the calibration and optimization of gesture recognition logic. The extracted landmarks, consisting of 21 key points representing the hand structure, are used as input features for gesture classification. These features are analyzed using rule-based algorithms that interpret spatial relationships such as distances, angles, and relative positions between landmarks. For example, the extension or folding of fingers is determined by comparing the coordinates of fingertip landmarks with respect to their corresponding joints.

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

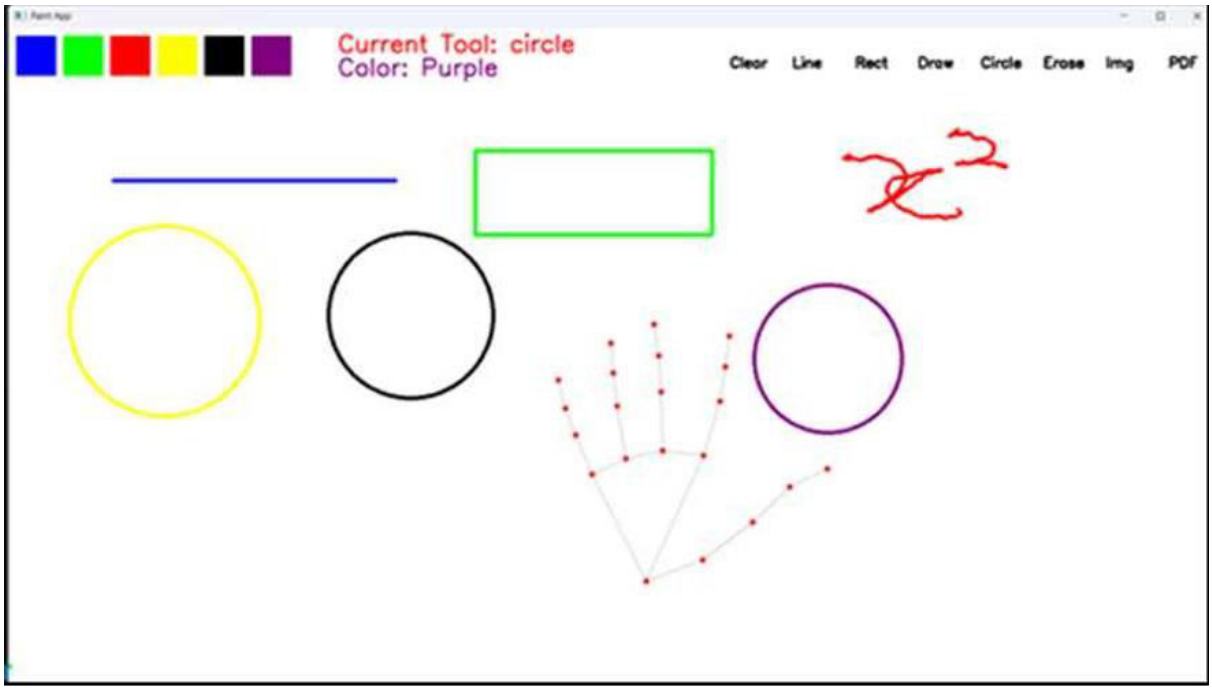
Impact Factor 8.699

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

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

V. RESULT



The results of the proposed Virtualink system are evaluated based on real-time gesture recognition and interaction with the virtual canvas. The system was tested using various hand gestures under different environmental conditions, including variations in lighting, background complexity, and hand orientation. The output images demonstrate the ability of the system to accurately detect hand landmarks and translate them into corresponding actions such as drawing, erasing, and selection.

In the observed results, when a single finger is extended, the system successfully identifies the gesture and enables drawing mode, allowing continuous tracking of fingertip movement to produce smooth lines on the virtual canvas. Similarly, the use of two fingers activates the selection mode, enabling users to interact with different tools or functionalities. The erasing operation is effectively performed when a closed fist gesture is detected, indicating the system's ability to distinguish between multiple gesture patterns.

The result images clearly illustrate that the MediaPipe-based hand tracking mechanism accurately detects the 21 hand landmarks even in moderately complex backgrounds. The system demonstrates real-time responsiveness with minimal latency, ensuring smooth user interaction without noticeable delay. The continuity of gesture tracking across frames indicates the efficiency of the hand tracking algorithm.

The results confirm that the proposed Virtualink system provides an effective and practical solution for touchless interaction, with reliable gesture recognition and real-time performance suitable for applications such as digital drawing and interactive systems.

VI. CONCLUSION

In this paper, a comprehensive study and implementation of a real-time gesture recognition system, "Virtualink," has been presented, focusing on enabling intuitive and touchless human-computer interaction using computer vision techniques. The system successfully demonstrates how hand gestures can be utilized as an effective alternative to traditional input devices such as keyboards and mice, thereby enhancing user interaction in a more natural and flexible manner.

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 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504149|

The proposed system leverages the capabilities of the MediaPipe framework for efficient hand detection and landmark extraction, combined with OpenCV for real-time image processing and visualization. By utilizing pre-trained models for hand tracking and a rule-based approach for gesture classification, the system achieves a balance between computational efficiency and practical usability. This approach eliminates the need for extensive model training while still delivering reliable performance in real-time scenarios.

The architecture of the system is designed as a modular pipeline, consisting of stages such as image acquisition, preprocessing, hand detection, feature extraction, gesture recognition, and output rendering. Each stage contributes to the accurate interpretation of gestures and ensures smooth interaction with the virtual canvas. The system is capable of recognizing basic gestures such as drawing, erasing, and selection, making it suitable for applications in digital writing, virtual collaboration, and interactive systems.

Experimental evaluation using real-time test inputs demonstrates that the system performs effectively under controlled conditions, with satisfactory accuracy and minimal latency. However, certain limitations have been observed, including sensitivity to lighting conditions, dependency on camera quality, and reduced performance in complex backgrounds. These challenges highlight the need for further enhancements to improve robustness and adaptability.

In conclusion, the Virtualink system provides a practical and efficient solution for gesture-based interaction, showcasing the potential of computer vision in modern human-computer interfaces. The system lays a strong foundation for future advancements, where integration with deep learning models, augmented reality (AR), and multi-user interaction can significantly enhance performance and expand application domains. With continuous improvements, gesture recognition systems are expected to play a vital role in the development of next-generation interactive technologies.

REFERENCES

[1] **S. U. Saoji,** "Air Canvas Application Using OpenCV and NumPy," International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 6, pp. 2345–2348, 2021.
[2] **T. H. Chandhan, K. Sai, and P. Reddy**, "Real-Time Hand Tracking Using MediaPipe and OpenCV," International Journal of Advanced Research in Computer Science, vol. 14, no. 2, pp. 112–118, 2023.
[3] **S. Sandbhor, A. Patil, and R. Jadhav,** "Survey on Air Canvas and Gesture-Based Interaction Systems," SSRN Electronic Journal, pp. 1–6, 2024.
[4] **M. Ikar, R. Patel, and S. Mehta**, "Computer Vision-Based Virtual Paint Application,International Journal of Trend in Research and Development (IJTRD), vol. 10, no. 3, pp. 45–49, 2023.
[5] **B. Venugopal, S. Kumar, and P. Sharma**, "Air Canvas: Drawing Application Using Hand Gestures," International Research Journal of Engineering and Technology (IRJET), vol. 10, no. 4, pp. 567–572, 2023.
[6]**Google, "MediaPipe Hands: On-device** Real-time Hand Tracking," Available: https://developers.google.com/mediapipe
[7] **G. Bradski, "The OpenCV Library**," Dr. Dobb's Journal of Software Tools, 2000.
[8] **R. Szeliski, Computer Vision**: Algorithms and Applications, Springer, 2011.
[9] **D. Forsyth and J. Ponce, Computer Vision:** A Modern Approach, Pearson, 2012.
[10] **A. Krizhevsky, I. Sutskever, and G. Hinton**, "ImageNet Classification with Deep Convolutional Neural Networks," **Advances in Neural Information Processing Systems,** 2012.

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