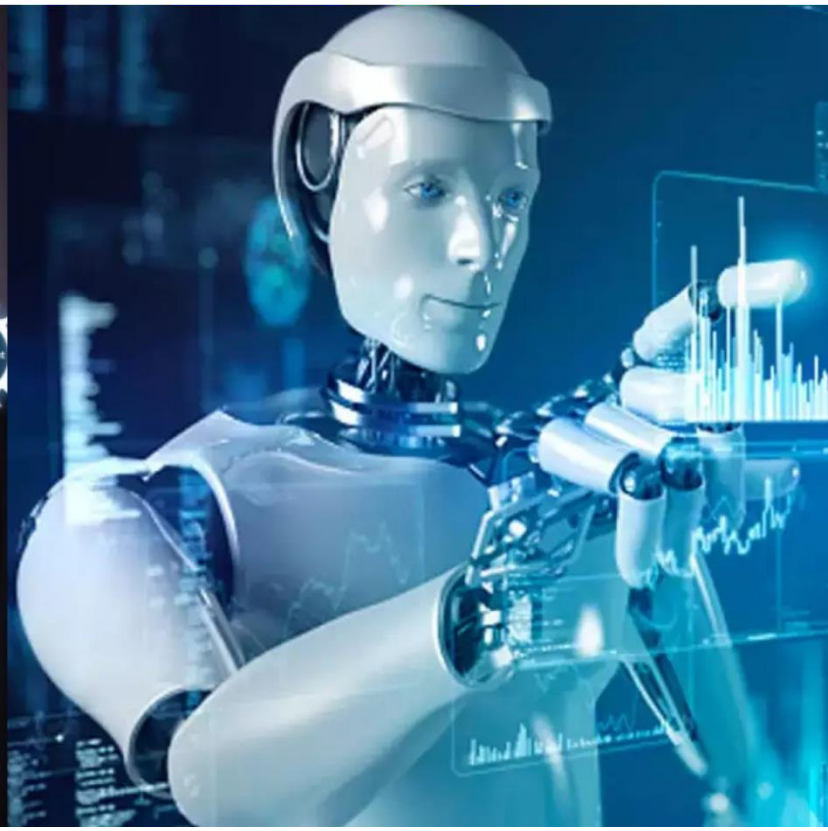
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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.2026.1504133|

Virtualink: A Study on Gesture Recognition System for Touchless Interaction”

Ankush Gadage[1], Sarthak Jagtap[2], Surabhi Jawalekar [3], Samruddhi Awate[4], Gajanan Bondre[5]

Lecturer, Department. of Computer Engineering, AISSMS Polytechnic, Pune, India[1]

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India [2]

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India [3]

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India [4]

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India [5]

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.

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.2026.1504133|

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.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.

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.2026.1504133|

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.

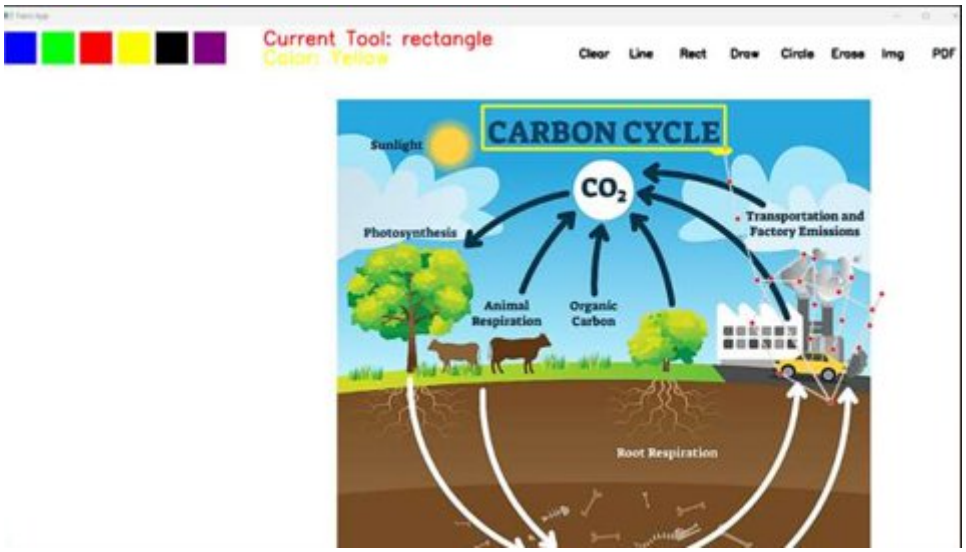
Table I illustrates the number of images available for each skin disease category. The dataset contains a diverse set of dermatological images with variations in lighting conditions, skin tones, and lesion characteristics. This diversity enables the model to learn meaningful visual patterns associated with different skin diseases and improves its ability to generalize when classifying unseen images.

IV. PROPOSED METHODOLOGY

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..



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.

.

System Architecture and Working principle

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

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.2026.1504133|

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.

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.

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.2026.1504133|

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

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

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.2026.1504133|

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.

To ensure reliable gesture recognition, the system undergoes an iterative tuning process, which can be considered as an implicit training phase. During this phase, multiple experimental trials are conducted using different hand gestures under varying environmental conditions. The system parameters, including detection confidence thresholds, tracking stability, and gesture sensitivity levels, are adjusted to optimize performance. This tuning process helps the system generalize across different users, hand sizes, and interaction styles.

Furthermore, temporal consistency is incorporated into the training strategy by analyzing gesture continuity across consecutive frames. Instead of relying solely on single-frame detection, the system evaluates gesture patterns over time to reduce noise and false positives. This temporal smoothing mechanism improves the robustness of gesture recognition, especially for dynamic gestures such as drawing and writing.

To evaluate the effectiveness of the system, various performance metrics such as recognition accuracy, response time, and stability are considered during testing. The system demonstrates high responsiveness with minimal latency, making it suitable for real-time applications. However, performance may vary under extreme lighting conditions or highly cluttered backgrounds, which are common challenges in vision-based systems.

In addition to the current implementation, the system is designed with extensibility in mind, allowing integration of advanced machine learning techniques in future work. For instance, a supervised learning approach can be implemented using labeled gesture datasets, where each gesture is associated with a specific class. A Convolutional Neural Network (CNN) can be trained to automatically learn feature representations from image data, while a Recurrent Neural Network (RNN) or Long Short-Term Memory (LSTM) network can be used to capture temporal dependencies in dynamic gestures.

The training process for such models would involve dataset collection, preprocessing, feature extraction, model training using optimization algorithms such as Adam or SGD, and evaluation using metrics like accuracy, precision, recall, and F1-score. Data augmentation techniques such as rotation, scaling, and flipping can be applied to improve model generalization and reduce overfitting.

Overall, the proposed system achieves an effective balance between performance and computational efficiency by leveraging pre-trained models and lightweight gesture classification techniques. This approach ensures that the system remains scalable, adaptable, and suitable for real-time deployment without requiring extensive computational resources. The training methodology, although simplified, provides a strong foundation for future enhancements and integration

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.2026.1504133|

of advanced deep learning models.

V. 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.

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..

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