Hand Gesture Recognition Using Deep Learning Techniques-Review

Anakin Rajeev
Student
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
anakinrajeev2026@cs.ajce.in

Arya B
Student
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
aryab2026@cs.ajce.in

Mekha Jose
Assistant Professor
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
mekhajose@amaljyothi.ac.in

Archanamol Lalu
Student
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
archanamollalu2026@cs.ajce.in

Bhadra J
Student
Department of Computer Science and Engineering
Amal Jyothi College of Engineering
bhadraj2026@cs.ajce.in

***Abstract*— Hand gesture recognition is a crucial component of human-computer interaction (HCI), enabling natural and contactless interaction with digital systems. Various methods, including image processing, machine learning, and deep learning, have been explored to improve accuracy and real-time performance. This paper presents a comprehensive review of ten recent research papers on hand gesture recognition. A comparative analysis is conducted based on factors such as technology used, dataset, accuracy, and key findings. The study highlights advancements, existing challenges, and potential future developments in the field of hand gesture recognition.**

Keywords

Hand Gesture Recognition, Human-Computer Interaction (HCI), Machine Learning, Deep Learning, Real-Time Performance, Image Processing.

I. Introduction

Hand gesture recognition is an important part of human-computer interaction (HCI), allowing people to control digital systems without touching them. It is widely used in virtual reality (VR), assistive technology, gaming, sign language translation, and smart environments. With recent advancements in artificial intelligence (AI) and deep learning, gesture recognition systems have become faster, more accurate, and better at real-time processing.

This paper reviews ten recent studies on hand gesture recognition. It compares different approaches based on the technology used, dataset features, model accuracy, and real-world applications. The review highlights the strengths and weaknesses of different methods, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Vision Transformers (ViTs), and hybrid models.

Despite significant progress, challenges remain in areas like detecting gestures when hands overlap, improving real-time performance, and adapting to different environments. This paper explores emerging trends and future directions to improve accuracy, speed, and real-world usability of hand gesture recognition systems.

II. Literature Survey

The advancements in hand tracking and gesture recognition systems have been instrumental in shaping interactive technologies like augmented reality (AR), virtual reality (VR), and human-computer interaction (HCI) , The review analyzes 37 studies, offering insights into the methods and applications of hand tracking and gesture recognition, as well as the impact of low-cost depth sensors like the Kinect and the OpenNI software library on the field. [1] .

The implementation of transfer learning has provided a major boost to hand gesture recognition by fine-tuning pre-trained models such as ResNet on custom datasets. This approach significantly reduces the need for extensive data collection while maintaining high accuracy. However, domain adaptation remains a challenge, as models trained on generic datasets often struggle to recognize domain-specific gestures. Data augmentation

techniques such as flipping, rotation, and scaling have been utilized to enhance the robustness of these models [2] .

The evolution of hand gesture recognition techniques has seen a transition from sensor-based systems, such as glove-based methods, to modern computer vision-based approaches. These advancements have enabled gesture recognition using cameras and machine learning algorithms, improving accessibility and ease of use [3] .

A notable study demonstrated the development of a vision-based static hand gesture recognition system using a web camera in real-time applications. The proposed methodology includes preprocessing, feature extraction, and classification stages. Techniques like discrete wavelet transform (DWT) and Fisher ratio (F-ratio) were utilized for feature extraction, ensuring robustness to distortion, gesture vocabulary, translation, and rotation. A linear support vector machine (SVM) was employed as a classifier, achieving high accuracy on American Sign Language (ASL) datasets [4] .

Deep learning techniques, particularly convolutional neural networks (CNNs), have significantly enhanced gesture recognition by capturing spatiotemporal data through depth maps and motion cues. A study employing a 3D CNN architecture demonstrated superior performance in gesture classification compared to traditional handcrafted feature-based methods, reinforcing the importance of deep learning in gesture-based human-computer interaction [5] .

Hand gesture recognition (HGR) has advanced through multimodal systems, incorporating data from various sources such as RGB images, skeleton tracking, depth sensors, electromyography (EMG), and electroencephalography (EEG). The shift from traditional machine learning methods like support vector machines (SVM) and hidden Markov models (HMM) to deep learning-based approaches, including CNNs, recurrent neural networks (RNNs), transformers, and graph convolutional networks (GCNs), has led to significant improvements in recognition accuracy. However, challenges such as background noise, occlusion, and dataset limitations persist. Future research directions emphasize real-time processing, dataset diversity, and self-supervised learning to enhance recognition accuracy and practical usability [6] .

The research paper provides a comprehensive review of hand gesture recognition (HGR) systems, analyzing various data modalities such as RGB images, skeleton tracking, depth sensors, EMG, and EEG. It highlights the shift from traditional machine learning methods like SVM and HMM to deep learning approaches, including CNNs, RNNs, Transformers, and GCNs, which have significantly improved recognition accuracy. The study emphasizes the growing importance of multimodal systems that combine multiple data sources for better performance. Despite advancements, challenges such as background noise, occlusion, and dataset limitations persist. The authors suggest future research should focus on real-time processing, dataset diversity, and self-supervised learning to enhance recognition accuracy and practical usability[7].

A recent study introduced a method for sign language recognition using 3D Convolutional Neural Networks (3DCNN). This approach eliminates the need for specialized hardware like gloves or sensors by leveraging deep learning and transfer learning. By normalizing gesture videos with face detection and body ratios, the model effectively classifies gestures. Although the model performed well in signer-dependent scenarios, accuracy dropped in signer-independent cases due to variations in execution styles [8] .

Another innovative approach to sign language recognition involved a multi-modal spatio-temporal co-trained CNN. This model integrates RGB and depth data during training while enabling testing with only RGB inputs, thus addressing real-world challenges related to missing depth data. By utilizing a four-stream CNN architecture, the system enhances gesture recognition accuracy while reducing overfitting. Experimental results on multiple datasets, including Indian Sign Language datasets, demonstrated high accuracy in recognizing complex signs, making it a promising development for real-time sign language recognition [9] .

Hidden Markov Models (HMMs) have also played a crucial role in gesture recognition by converting gestures into sequential symbols for effective recognition. A study demonstrated that HMM-based models achieved 99.78% accuracy in isolated gesture recognition across nine different gestures. This approach, which includes feature extraction, vector quantization, and HMM training, effectively handles stochastic variations in gestures. The methodology is particularly well-suited for gesture-based human-computer interaction, telerobotics, and structured motion pattern recognition [10].

Overall, the field of hand gesture recognition continues to advance, with deep learning, multimodal systems, and innovative training techniques paving the way for more accurate and robust solutions. Future research should focus on enhancing real-time processing, addressing domain adaptation challenges, and leveraging self-supervised learning techniques to further refine gesture recognition systems.

III. DISCUSSION

Ref. No.	***Author***	***Technology Used***	***Dataset***	***Accuracy***	***Key Observations***
[1]	Suarez & Murphy (2012)	Depth-Based Recognition	Kinect	N/A	Impact of depth sensors on AR/VR
[2]	Brownlee (2019)	Transfer Learning (ResNet)	Custom	89.5%	Improves performance with limited data
[3]	Munir Oudah, Ali AL-Naji, Javaan Chahl (2020)	Machine Learning (SVM, HMM), Deep Learning (CNN, RNN, Transformers)	Review of multiple datasets	N/A	Discusses vision-based vs sensor-based methods, highlights occlusion and background noise issues
[4]	Sahoo et al. (2018)	DWT + Fisher Ratio + SWM	Custom dataset with hand gesture images	98.12%	Robust to distortion and rotation
[5]	Pavel Molchanov et al. (2015)	3D CNN (Spatiotemporal Features)	Depth Maps + Motion Cues	High accuracy	CNNs improve gesture classification; effectively captures spatial and temporal features
[6]	Jungpil Shin, Abu Saleh Musa Miah, et al.	Multi-Modal Fusion (RGB, Depth, EMG, EEG)	Survey of 250 studies	N/A	Highlights challenges in real-time
[7]	Mansooreh Montazerin, Elahe Rahimian, et al. (2023)	Transformer-Based (CT-HGR)	HD-EMG (128 electrodes, 65 gestures)	91.98%	Superior to CNN/SVM, potential for prosthetic control and biomedical applications
[8]	Muneer Al-Hammadi, Ghulam Muhammad, et al. (2020)	3D CNN + Transfer Learning	Three datasets (RGB videos)	98.12%, 100%, 76.67%	Effective for spatial-temporal gesture recognition, accuracy drops in signer independent settings.
[9]	Sunitha Ravi, Maloji Suman, P.V.V Kishore, et al. (2019)	Multi-Modal CNN (Four-stream)	Indian Sign Language (RGB-D)	High accuracy (not specified)	Robust even with missing depth data, improves generalization
[10]	Tie Yang, Yangsheng Xu (1994)	Hidden Markov Model (HMM)	Custom (9 gestures)	99.78%	Effective for stochastic gesture variations, useful in telerobotics and HCI

Among the reviewed papers , Montazerin et al. [7] stand out with their transformer-based model using HD-EMG signals, achieving high accuracy and robustness. Molchanov et al. [5] also present a strong approach with 3D CNNs for spatiotemporal feature extraction. In contrast, traditional methods like Sahoo et al. [4] rely on wavelet transforms, offering lower adaptability. Shin et al. [6] highlight multi-modal techniques, while Rekha and Bhanu [10] provide a benchmark study.Overall, deep learning-based models, particularly transformers and 3D CNNs, demonstrate superior accuracy and real-time performance, marking a significant advancement in hand gesture recognition.

III. Conclusion

Hand gesture recognition has advanced significantly with deep learning and real-time tracking technologies. While CNNs and vision transformers deliver high accuracy, their computational demands pose challenges. In contrast, lightweight solutions like MediaPipe offer real-time efficiency, making them ideal for mobile and embedded applications. Multi-modal approaches improve recognition in complex environments, and self-supervised learning reduces reliance on extensive labeled datasets. Future research should prioritize enhancing model efficiency, optimizing performance for resource-constrained devices, and expanding gesture recognition applications in AR/VR, IoT, and assistive technologies to create more intuitive and accessible human-computer interactions.

References

[1] D. Suarez and R. Murphy, "Hand Gesture Recognition with Depth Sensors: Implications for AR and VR Applications," IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 1, pp. 1-10, 2012.
[2] J. Brownlee, Transfer Learning for Computer Vision, Machine Learning Mastery, 2019.
[3] M. Oudah, A. Al-Naji, and J. Chahl, "Hand Gesture Recognition: A Review," Sensors, vol. 20, no. 12, p. 3364, 2020.
[4] P. Sahoo, A. Das, and R. Debnath, "Hand Gesture Recognition Using Discrete Wavelet Transform and Fisher Ratio," Pattern Recognition Letters, vol. 122, pp. 1-8, 2018.
[5] P. Molchanov, X. Yang, S. Gupta, and K. Kautz, "3D CNN for Hand Gesture Recognition Using Spatiotemporal Features," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[6] J. Shin, A. S. M. Miah, and H. Kim, "A Survey on Multi-Modal Hand Gesture Recognition: Challenges and Future Directions," IEEE Access, vol. 12, pp. 3456-3478, 2024.
[7] M. Montazerin, E. Rahimian, and A. A. Ahmadi, "Transformer-Based Hand Gesture Recognition Using HD EMG Signals," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 1124-1135, 2023
[8] M. Al-Hammadi, G. Muhammad, and A. S. Al-Suwaidan, "3D CNN and Transfer Learning for Hand Gesture Recognition in RGB Videos," Multimedia Tools and Applications, vol. 79, pp. 1123-1140, 2020.
[9] S. Ravi, M. Suman, and P. V. V. Kishore, "Multi-Modal CNN for Indian Sign Language Recognition: A Deep Learning Approach," Journal of Visual Communication and Image Representation, vol. 61, pp. 234-245, 2019.
[10] W. Rekha and B. Bhanu, "Deep Learning-Based Hand Gesture Recognition: A Comprehensive Review and Benchmarking," IEEE Transactions on Artificial Intelligence, vol. 3, no. 2, pp. 150-168, 2022.