© May 2024| IJIRT | Volume 10 Issue 12 | ISSN: 2349-6002

Mapping the Evolution of Hand Gesture Recognition: Techniques, Challenges and Innovations

Rameesa A B[1], Bismin V Sherif[2]
[1,2]Dept. of Computer Applications, MES College Marampally, Kerala, India

Abstract- Hand gesture recognition (HGR) has emerged as a vital component in human-computer interaction (HCI) systems, enabling natural and intuitive communication between humans and machines. It has many applications in virtual environment control and sign language translation, robot control, or music creation. This survey paper provides a comprehensive overview of the various techniques, methodologies, and advancements in the field of HGR. This research focuses on the development of an efficient hand gesture recognition system leveraging machine learning algorithms. The primary objective is to interpret and classify hand movements captured by sensors, transforming them into actionable commands.

Keywords: Hand Gesture Recognition, Convolution Neural Networks, Machine Learning, MediaPipe.

1. INTRODUCTION

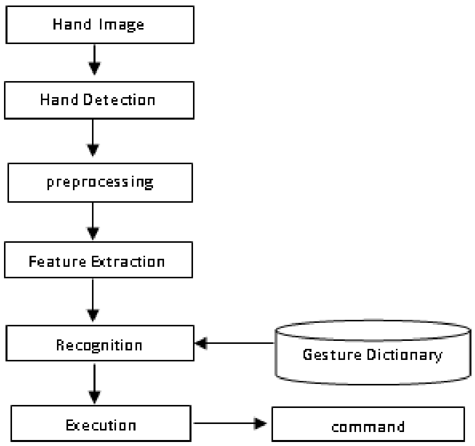
Gesture recognition is technology that uses sensors to read and interpret hand movements as commands. In the automotive industry, this capability allows drivers and passengers to interact with the vehicle — usually to control the infotainment system without touching any buttons or screens. Hand gesture recognition is one of the active research areas in the field of human-computer interface due to its flexibility and user friendliness. The gesture recognition technique is used to develop a system that can be used to convey information among disabled people or for controlling a device. Major challenges for the development of an efficient hand gesture recognition technique are illumination variation, nonuniform backgrounds, diversities in the size and shape of a user's hand, and high interclass similarities between hand gesture poses.

The ability to perceive the shape and motion of hands can be a vital component in improving the user experience across a variety of technological domains and platforms. For example, it can form the basis for sign language understanding and hand gesture control, and can also enable the overlay of digital content and information on top of the physical world in augmented reality. While coming naturally to people, robust real-time hand perception is a decidedly challenging computer vision task, as hands often occlude themselves or each other (e.g. finger/palm occlusions and hand shakes) and lack high contrast patterns.

MediaPipe Hands is a high-fidelity hand and finger tracking solution. It employs machine learning (ML) to infer 21 3D landmarks of a hand from just a single frame. Whereas current state-of-the-art approaches rely primarily on powerful desktop environments for inference, our method achieves real-time performance on a mobile phone, and even scales to multiple hands. We hope that providing this hand perception functionality to the wider research and development community will result in an emergence of creative use cases, stimulating new applications and new research avenues. The dataset employed in this research it is a sample program that recognises hand signs and finger gestures with a simple MLP using the detected key points. Handpose is estimated using MediaPipe.

The diverse studies presented underscore the continuous advancements in HGR techniques, addressing challenges, and expanding the range of applications. From real-world scenarios with complex backgrounds to wearable technology, drone control, and human-robot collaboration, these studies contribute to the growing body of knowledge in the field of hand gesture recognition, paving the way for future innovations and applications.

2. HAND GESTURE RECOGNITION SYSTEM



1.Capture Hand Image: This step involves capturing hand images or data using input devices such as cameras or depth sensors.
2.Hand Detection: The process of identifying and localizing the regions of an image that contain human hands. It involves techniques such as contour detection, skin color segmentation, or machine learning algorithms to distinguish hands from the background.
3.Preprocessing:The initial stage of image processing where the input hand image is modified or enhanced to improve the quality of subsequent processing steps. This may involve operations such as converting to grayscale, noise reduction, and image normalization.
4.Feature Extraction: The process of identifying and extracting relevant information or features from the hand image that can be used for recognition. Features may include shape descriptors, texture patterns, or motion characteristics of the hand.
5.Recognition: The process of identifying or classifying the hand gesture based on the extracted features. This often involves training a machine learning model on a dataset of labeled hand gestures and then using this model to predict the gesture class for new input images.
6.Execution: In the context of hand gesture recognition, execution refers to the action or command associated with a recognized gesture. For example, if a hand gesture is recognized as a "thumbs up," the execution might involve triggering a specific action such as sending a command to a device or performing a particular function. .
7.Command:A specific action or instruction triggered by a recognized hand gesture. Commands can range from simple operations like turning a device on or off to more complex tasks such as navigating through a menu or controlling a robot.
8.Gesture Dictionary:A collection of predefined hand gestures along with their corresponding meanings or commands. The gesture dictionary serves as a reference for the recognition system to map detected gestures to their intended actions or interpretations.

3. HAND GESTURE RECOGNITION TECHNIQUES

Weina Zhou and Kun Chen[1] Conducted a research on “ A light weight hand gesture recognition in a complex background.” introduced a two-stage Hand Gesture Recognition (HGR) system designed to address challenges in real-world scenarios, particularly the recognition of hand gestures in complex backgrounds. The proposed system focuses on achieving a balance between lightweight design and high recognition accuracy. In the initial stage, the authors employed an accurate segmentation process to separate the hand from the complex background. This segmentation network utilized a combination of a dilated residual network (DRN), an atrous spatial pyramid pooling module (ASPP), and a simplified decoder. The DRN and ASPP were utilized for precise segmentation, while the simplified decoder refined the segmented hand regions. The second stage of the HGR system introduced double-channel Convolutional Neural Networks (CNNs) to enhance recognition performance. This involved taking both the segmented hand image and the original RGB image as input to the CNNs, with the extracted features from both images fused to produce the final HGR results. The authors highlighted the significance of HGR in machine vision, emphasizing its applications in intelligent driving, machine control, and virtual reality. The proposed vision-based HGR system aimed to overcome challenges related to hand recognition in complex backgrounds. The two-stage method consisted of hand segmentation and hand gesture recognition, utilizing an encoder-decoder framework in the first stage, including DRN, ASPP module, and a simplified decoder. The study utilized the OUHANDS dataset, known for its complex backgrounds, featuring ten different gesture classes from 23 subjects with background disturbances. The proposed model demonstrated notable advancements, achieving a duration accuracy of 91.17% with a model size of 1.8 MB. These results surpassed other state-of-the-art models in hand gesture recognition, highlighting the effectiveness of the two-stage HGR system in complex background scenarios.

S.Tiwari et al. [2] propose a method on the development of a" hand gesture-based volume controller." utilizing the OpenCV module for gesture detection and control. The system captures images and videos from the webcam and adjusts the volume based on user gestures, providing a hands-free approach to volume control. This technology is particularly beneficial in scenarios where direct device access is challenging or when users prefer discreet volume adjustments. A key advantage highlighted by the authors is the accessibility of the hand gesture volume controller, enabling individuals with physical disabilities to manage device volume without relying on physical buttons or remote controls. They emphasize the potential revolutionary impact of this technology on human-device interactions. The research involves the use of a camera or sensor to capture user hand gestures, with a focus on addressing challenges related to background images or videos that may affect gesture recognition quality. The authors emphasize the use of open-source software and hardware to enhance accessibility and ease of replication for those interested in building similar volume controllers using hand gestures. To achieve their goals, the authors utilized essential packages such as ImUtils, OpenCV-Python, SciPy, TensorFlow, NumPy, and MediaPipe. The gesture recognition process using an Artificial Neural Network (ANN) typically involves steps such as data collection, data preprocessing, feature extraction, ANN training, testing, and evaluation. The authors evaluated their proposed system using a dataset comprising 50 different hand gestures, including actions such as decreasing volume, increasing volume, reaching minimum and maximum volume, and mute. The reported results demonstrated a success rate exceeding 95%, showcasing the effectiveness of the hand gesture volume controller.

In [3] develop a study "Comparing EMG-based hand gesture recognition (HGR) systems employing supervised and reinforcement learning." While many HGR methods rely on supervised machine learning (ML), the application of reinforcement learning (RL) for EMG classification remains underexplored. The performance of HGR systems based on ML and RL methods for user-general HGR on large datasets is an ongoing research challenge. This study compares supervised learning, featuring k-nearest neighbors (K-NN), support vector machine (SVM), artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent neural networks (RNNs) like long-short-term memory (LSTM), with a reinforcement learning approach. The HGR systems consist of pre-processing, feature extraction using CNN, classification, and post-processing stages. Both supervised and RL models were evaluated for six hand gestures. The learning agents in both cases use CNN for feature extraction. The proposed models were tested on a validation set of 306 users for the supervised method and 40 users for the RL method. The supervised learning method achieved the best results with an accuracy of 90.49% (+9.7%).

Yong Soon Tan and Kian Ming Lim, Chin Poo Lee [4] conducted a study on "hand gesture recognition using an enhanced densely connected convolutional neural network." Hand gesture recognition can be broadly categorized into static and dynamic recognition. Static recognition classifies gestures from images, while dynamic recognition processes videos with multiple frames. Two main approaches for static recognition are wearable device-based and vision-based. Wearable devices offer accuracy but are inconvenient, while vision-based approaches are flexible but face challenges like variation in skin tones and background noise. Vision-based recognition can be achieved through hand-crafted and deep learning approaches. Deep learning, particularly convolutional neural networks (CNN), is widely used for image recognition tasks, including sign language recognition. Despite its effectiveness, CNNs have challenges, leading to the proposal of densely connected convolutional networks (DenseNet) for better parameter efficiency. DenseNet improves feature propagation and gradient flow within a dense block.The study utilized three datasets, introducing an Enhanced DenseNet (EDenseNet) for vision-based static gesture recognition. EDenseNet, with its emphasis on dense connectivity, achieved 98.50% average accuracy without augmented data and

99.64% with augmented data. It outperformed other deep learning instances in both settings, showcasing its effectiveness in static gesture recognition.

In [5] develop a study on “Gaze-aware hand gesture recognition for intelligent construction.” proposing an innovative framework serving as a human–robot interface. This framework aims to address limitations in existing hand gesture approaches for robot–worker collaboration, encompassing three key components: visual detection and tracking, machine-of-interest generation, and hand gesture recognition. The study includes a validation test to evaluate precision and recall performance, demonstrating that the proposed framework is effective for facilitating interaction between workers and multiple construction machines. Implemented on a Windows 10 64-bit operating system, the framework utilizes Python 3.6 with support from PyTorch and Tensorflow platforms, incorporating essential algorithms, functions, and tools. Notably, this research marks the first integration of gaze tracking and gesture recognition for collaborative interactions with construction equipment. The gaze-aware hand gesture recognition framework achieved a precision of 93.8% and a recall of 95.0% during the validation test, highlighting its suitability for one-to-many collaboration in construction applications.

JP.Vasconez et al. [6] conducted a study on "Hand Gesture Recognition Using EMG-IMU Signals and Deep Q-Networks." They introduced an HGR system based on the DQN algorithm for classifying 11 distinct hand gestures, encompassing both static and dynamic gestures. The research involved the evaluation and comparison of two sensors, namely the Myo armband and G-force sensors. EMG and IMU signals were utilized to derive feature vectors from these sensors. The proposed models underwent validation on 43 users and testing on an additional 42 users. The Myo armband sensor exhibited superior performance, achieving a classification accuracy of up to 97.50%±1.13% and 88.15%±2.84% for static gestures, and 98.95%±0.62% and 90.47%±4.57% for dynamic gestures in terms of classification and recognition, respectively. The study demonstrated that the DQN effectively learned a policy from online experience to classify and recognize gestures based on EMG and IMU signals, surpassing results obtained by similar methods using only EMG. Furthermore, the Myo armband sensor outperformed the G-force sensor in terms of accuracy and data distribution.

In [7] develop a research on "Hand Gesture Recognition based on a Harris Hawks Optimized Convolution Neural Network." The study employed Convolutional Neural Networks (CNN) for the classification of hand gesture images. To optimize the CNN's hyper-parameters, they utilized the Harris Hawks Optimization (HHO) algorithm, a recently developed metaheuristic approach. Through a thorough comparative analysis, the proposed HHO-CNN hybrid model demonstrated superior performance, achieving a remarkable accuracy of 100%. The research highlights the significance of hyper-parameter tuning in optimizing Convolutional Neural Networks for image classification. The use of nature-inspired metaheuristic algorithms, such as the Harris Hawks Optimizer, proved effective in selecting optimal parameters, leading to enhanced performance and reduced training time for the CNN in hand gesture image classification.

E. Stergiopoulou and N. Papamarkos [8] presented a study on "Hand Gesture Recognition using a Neural Network Shape Fitting Technique." The paper introduces an innovative approach to hand gesture recognition based on extracting hand gesture features and employing a neural network shape fitting method. Initially, a skin color filtering process is applied in the YCbCr color space to swiftly isolate the hand region, ensuring noiseless segmented images despite variations in skin color and lighting conditions. The subsequent stages involve fitting the shape of the hand and recognizing the finger configuration. The developed hand gesture recognition system, implemented in Delphi, underwent testing with hand images from diverse individuals, considering variations in morphology, slope, and size. The system was trained to identify 31 hand gestures, involving combinations of raised and not raised fingers, facilitating human-computer communication without the need for specialized hardware. Extensive testing of the proposed system with a substantial number of input images yielded a highly promising recognition rate. The success of the system is demonstrated through rigorous testing.

Ji-Won Lee and Kee-Ho Yu [9] proposed a study titled "Wearable Drone Controller: Machine Learning-Based Hand Gesture Recognition and Vibrotactile Feedback." The research introduced a wearable drone controller integrating hand gesture recognition and vibrotactile feedback. The control system utilized an inertial measurement unit (IMU) positioned on the back of the hand to detect hand motions for drone navigation. The recorded motions were categorized through machine learning employing the ensemble method, achieving a classification accuracy of 97.9%. Additionally, the controller incorporated vibrotactile feedback by relaying information about the distance to obstacles in the drone's heading direction. This feedback was delivered to the user through a vibration motor attached to the wrist. In simulated experiments with a participant group, the hand gesture control exhibited robust performance. The vibrotactile feedback proved beneficial in enhancing the user's awareness of the drone's operational environment, particularly in scenarios with limited visual information. To evaluate the proposed controller, a subjective assessment was conducted with participants to gauge its convenience and effectiveness. Subsequently, a real drone experiment validated the applicability of the controller as a natural interface for drone operation. The average accuracy in direct mode was approximately 96%, exhibiting a marginal decrease of 2.6% compared to the simulation. Conversely, the average accuracy in gesture mode was around 98%, indicating a slight improvement of 1.4% over the simulation results.

Dandu Amarnatha Reddy et al. [10] conducted a research study titled "Hand Gesture Recognition Using Local Histogram Feature Descriptor." The system focused on hand gesture recognition for enhancing human-computer interaction, employing a vision-based approach with three main stages: preprocessing, feature extraction, and classification. In the preprocessing stage, the goal was to localize the hand region within the image frame. The authors utilized the Laplacian of Gaussian filtering technique in conjunction with a zero-crossing detector to identify the edges of the hand region in hand gesture images. A novel feature extraction technique, the Local Histogram Feature Descriptor (LHFD), was proposed in this paper. This method involves extracting features by computing the local histogram of the grayscale gesture image, utilizing the entire hand region. Importantly, the proposed method demonstrated invariance to scaling and illumination changes. The evaluation of the proposed technique was conducted on two standard datasets: the Massey University Gesture Dataset (MUGD) and Jochen Triesch Static Hand Posture Database. The recognition performance of the proposed technique was reported as 99.5% for the Massey University Gesture Dataset and 95% for the Triesch dataset. The evaluation utilized a multi-class support vector machine (SVM) classifier for classification purposes.In summary, the study presented a comprehensive approach to hand gesture recognition, introducing a novel feature extraction method (LHFD) that exhibited strong performance on standard datasets, showcasing its effectiveness in real-world applications.

No	Title	Techniques used	Accuracy
1.	A light weight hand gesture recognition in complex backgrounds.[1]	• Vision based HGR system. Wearable equipment base gesture recognition.	Vision based HGR have an accuracy of 91.17%.
2.	Volume controller using Hand Gestures.[2]	Artificial Neural Network (ANN).	Accuracy of 95%.
3.	Comparing EMG-based hand gesture recognition (HGR) systems employing supervised and reinforcement learning.[3]	• SupervisedMachine learning. • RL(Reinforcement learning).	Supervised learning have an accuracy of 90.49% (+9.7%).
4.	Hand gesture recognition using an enhanced densely connected convolutional neural network.[4]	• Convolutional Neural Network.	Achieved 98.50% average accuracy without augmented data.
5.	Gaze-aware hand gesture recognition for intelligent construction.[5]	gaze-aware hand gesture recognition framework.	Gaze-aware HGR framework have precision of 93.8%.
6.	Hand Gesture Recognition Using EMG-IMU Signals and Deep Q-Networks.[6]	Research involved the evaluation and comparison of two sensors, namely the Myo armband and G-force sensors.	98.95%±0.62% and 90.47%±4.57% for dynamic gestures in terms of classification and recognition (Myo armbad).
7.	Hand Gesture Recognition based on a Harris Hawks Optimized Convolution Neural Network.[7]	HHO-CNN hybrid model.	Accuracy of 100%.

8.	Hand Gesture Recognition using a Neural Network Shape Fitting Technique.[8]	Neural network shape fitting method.	A success accuracy.
9.	Wearable Drone Controller: Machine Learning-Based Hand Gesture Recognition and Vibrotactile Feedback.[9]	Machine learning (directed mode and gesture mode).	Machine learning classification accuracy of 97.9%.
10.	HGR Using Local Histogram Feature Descriptor.[10]	Local Histogram Feature Descriptor (LHFD).	99.5% for the Massey University Gesture Dataset.

4. CONCLUSION AND FUTURE TRENDS

In the above papers discussed about Hand gesture recognition used on machine learning and reinforcement learning, HGR Using Local Histogram Feature Descriptor, HGR used on neural network, HGR used in volume controller etc. These studies contribute significantly to the advancement of hand gesture recognition technologies, showcasing innovations in methodology, system design, and practical applications across various domains. Each paper have using different methods each have better accuracy result.

As technology continues to advance, several trends are emerging in the field of hand gesture recognition, pointing toward exciting possibilities for the future. Some of the key future trends in hand gesture recognition include: Improved Deep Learning Models, Edge Computing for Real-Time Processing, Fusion of Multiple Sensors, Enhanced Gesture Vocabulary etc.

REFERENCE

[1] Weina Zhou, Kun Chen “A Lightweight Hand Gesture Recognition In Complex Backgrounds,” Shanghai Maritime University, Shanghai 201306, China (2022), doi:https://doi.org/10.1016/j.displa.2022.102226

[2] S. Tiwari, A. Mishra, D. Kukreja and A. L. Yadav, "Volume Controller using Hand Gestures," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10308134

[3] JP. Vasconez, LI. Barona Lopez, AL. Valdivieso Caraguay, ME. Benalcazar, “A Comparison of EMG- based hand gesture recognition systems based on supervised and reinforcement learning,” Engineering Application of Artificial Intelligence (2023), doi: https://doi.org/10.1016/j.engappai.2023.106327

[4] Yong Soon Tan, Kian Ming Lim, Chin Poo Lee, “Hand gesture recognition via enhanced densely connected convolutional neural network,” Multimedia University, Expert Systems with Applications (2021), doi: https:// doi.org/ 10.1016 /j.eswa.2021.114797

[5] X.Wang, D.Veeramani, Z.Zhu, “Gaze-aware hand gesture recognition for intelligent construction,” University of Wisconsin-Madison, Engineering Applications of Artificial Intelligence (2023), doi: https://doi.org/10.1016/j.engappai.2023.106179

[6] 6. JP. Vasconez, LI. Barona Lopez, AL. Valdivieso Caraguay, ME. Benalcazar, “Hand gesture recognition using EMG-IMU Signals and Deep Q-Networks,” Artificial Intelligence and Computer Vision Research Lab, Escuela Politécnica Nacional, Quito 170517, Ecuador (2022), doi:https://doi.org/10.3390/s22249613

[7] 7.TP.Gadekallu, G.Srivastava, M.Liyanage, Iyyapparaja M, C.Lal Chawdhary, S.koppu, PK.Reddy Maddikunta,” Hand Gesture recognition based on a Harris Hawks optimized Convolution Neural Network ,” Computers and Electrical Engineering (2022), doi: https:// doi.org/10.1016/ j. compeleceng.2022.107836

[8] E.Stergiopoulou, N.Papamarkos “Hand gesture recognition using a neural network Shape fitting technique,” Engineering Application of Artificial Intelligence(2009),doi:https://org./10.1016/j.engappai.2009.03.008

[9] Ji-won Lee, Kee-Ho Yu “Wearable Drone Controller: Machine Learning -Based Hnad Gesture Recognition and vibrotactile feedback,” Joenhuk National University (2023), doi:https://doi.org.10.3390/s23052666

[10] Danda Amarnatha Reddy, java prakash Sahoo, Samit Ari” Hand Gesture Recognition using local histogram Feature Descriptor,” Department of EC, NIT (2018), doi:10.1109/ 1C0EI. 2018. 855 3849