

Review

Not peer-reviewed version

---

# Hand Gestures Recognition System

---

[DHANAMMA SHANKAR JAGLI](#)\*, [Nitu Kumari](#)\*, Laxmi Nakirekanti

Posted Date: 26 June 2024

doi: 10.20944/preprints202406.1837.v1

Keywords: Hand gesture recognition; Sign language interpretation; Media Pipe framework; Machine learning algorithms; Real-time gesture detection



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

# Hand Gestures Recognition System

Dhanamma Jagli and Nitu Kumari

Vivekanand Education Society's Institute of Technology

\* Correspondence: dhanamma.jagli@ves.ac.in; 2022.nitu.kumari@ves.ac.in

**Abstract:** Hand gestures are an essential kind of nonverbal communication that let people engage and be expressive with one another. As technology has progressed, hand gesture recognition systems have become an important field of study, providing a link between human activities and artificial intelligence. An overview of the field of hand gesture recognition systems is given in this abstract, along with an explanation of its importance, difficulties, and possible uses. This abstract begins by examining the role that hand gestures have in communication. It then goes on to show the cultural significance and wide range of variations in gestural expressions that occur in many contexts and societies. Artificial intelligence systems face a unique problem in comprehending and interpreting these motions, which calls for the combination of computer vision, machine learning, and pattern recognition approaches. The abstract goes on to address why people develop hand motions. The purpose behind creating systems to recognize hand gestures is also covered in the abstract, with a focus on how these systems could improve human-computer interaction in several fields, including robotics, virtual reality, healthcare, and assistive technologies. These systems have the potential to enhance accessibility, efficiency, and user experience by allowing people to engage with devices and interfaces naturally and intuitively. The abstract also discusses environmental elements including backdrop clutter and lighting, as well as the technological difficulties in identifying and understanding hand gestures, which include variations in hand forms, orientations, and movements. To tackle these obstacles, strong algorithms that can record and interpret complex hand movements in real-time must be developed.

**Keywords:** Hand gesture recognition; Sign language interpretation; Media Pipe framework; Machine learning algorithms; Real-time gesture detection

---

## 1. Introduction

Human-computer interaction (HCI) systems have been the focus of significant advances in the field of computer vision in recent times. Among them, sign language recognition by hand is one of the most promising technologies with a variety of uses which includes interactive gaming and sign language interpretation. Given their ease of use and natural communication style, hand gestures are a perfect input modality for a variety of scenarios involving human-computer interaction.

To shed light on the methods, difficulties, and potential future directions in the field of hand gesture recognition, this paper provides an extensive analysis of the most recent advancements in this subject. We investigate the fundamental ideas behind hand gesture detection, looking at both conventional computer vision techniques and cutting-edge deep learning-based alternatives. Furthermore, we examine the primary determinants of gesture recognition system performance, such as dataset characteristics, gesture representation, feature extraction, and classification techniques.

Additionally, we emphasize the various uses of hand gesture detection in several industries, including augmented reality, robotics, healthcare, and smart surroundings. Through an awareness of the capabilities and constraints of current hand gesture recognition systems, scientists and industry professionals can use this technology to create novel solutions for practical issues.

In general, the goal of this review is to be a useful tool for scholars, programmers, and hobbyists who want to push the boundaries of hand gesture detection technology. By delving further into the

underlying approaches and difficulties, we can open up new avenues for improving human-computer interaction and producing more natural and engaging user interfaces.

### *1.1. Objectives of the Project*

The primary goal of this research is

To develop a computer vision and image processing recognition system for sign language interpretation.

To compile and train a testing dataset.

To anticipate the gesture that the user will add.

### *1.2. Scope*

The following are included in the project's scope:

Construct sign language based on the needs of the user.

Enter the gesture name to remove any sign language.

Users receive appropriate validation and success signals when their tasks are completed successfully.

Makes it possible to build and train the model.

Users have the option to test the project in testing mode, which forecasts indications that they generate for the project.

Easy to use UI that is simple to engage with.

There is no need to touch or press any keys when creating the dataset.

## **2. Literature Survey**

Ravi Kiran [1] proposes a finger detection technique based on boundary-tracing and cusp detection analysis to locate fingertip positions. This method, while simple and efficient, is particularly effective for identifying open-finger hand gestures in American Sign Language. Despite a 5% error rate in gesture recognition, the obtained accuracy is deemed suitable for sign language conversion to text and speech, with the possibility of using a dictionary to rectify spelling errors.

Srivastava [2] employs the TensorFlow object detection API to develop a system for real-time detection of Indian Sign Language gestures. While achieving an average confidence rate of 85.45%, the system is limited by the small size of the training dataset.

Chai [3] introduces a sign language recognition and translation system utilizing the Microsoft Kinect sensor. This system offers real-time recognition of sign language gestures and provides translations in text and speech output. Leveraging depth data and skeletal tracking, the system accurately tracks hand and body movements, achieving a recognition rate of 92.1% on a dataset comprising 36 Chinese sign language gestures.

De Smedt [4] presents a dynamic hand gesture dataset and evaluation methodology for 3D hand gesture recognition using depth and skeletal data. Their skeleton-based approach demonstrates promising results, with accuracy rates of 88.24% and 81.90% for 14 and 28 different gestures, respectively.

Padmaja [5] outlines a real-time sign language detection methodology using Faster R-CNN and ResNet-50 for deep learning-based sign detection. Their vision-based approach includes word-level and finger-spelling sign language detection using a dataset comprising 200 photos, including letters V, L, and U, as well as the phrase "I love you" for each sign, captured at varying resolutions.

### *2.2. Limitation Existing System or Research Gap*

Limitations of [1] it works with only American sign language, can't work if all the fingers are closed.[2] not proper GUI. Limited with only a small number of datasets [3]. Kinect is a hardware device that provides depth and color data for motion tracking and recognition. Kinect has limitations compared to other motion-tracking systems, such as limited range, accuracy issues in certain lighting conditions, and difficulty tracking fine-grained movements [4]. Another limitation is the reliance on

skeletal information, which may be unavailable in some scenarios or subject to noise or errors. The authors do not provide a detailed analysis of the robustness of their method to these issues [5]. One limitation is the use of a very small dataset that may limit the result of the given method to other datasets and real-world scenarios.

3. Research Methodology

The system utilizes a machine learning algorithm known as the Random Forest Classifier to both train and predict gestures. Random Forest is a well-established supervised learning method that can address classification and regression tasks in machine learning. Its core principle lies in ensemble learning, which involves combining multiple classifiers to improve model performance. In essence, the Random Forest classifier comprises numerous decision trees built on different subsets of the dataset, and it aggregates their predictions to enhance predictive accuracy. Unlike relying on a single decision tree, the Random Forest algorithm makes predictions based on the majority vote from each tree. This ensemble approach not only boosts accuracy but also mitigates the risk of overfitting. The Random Forest algorithm is explained in the image below:

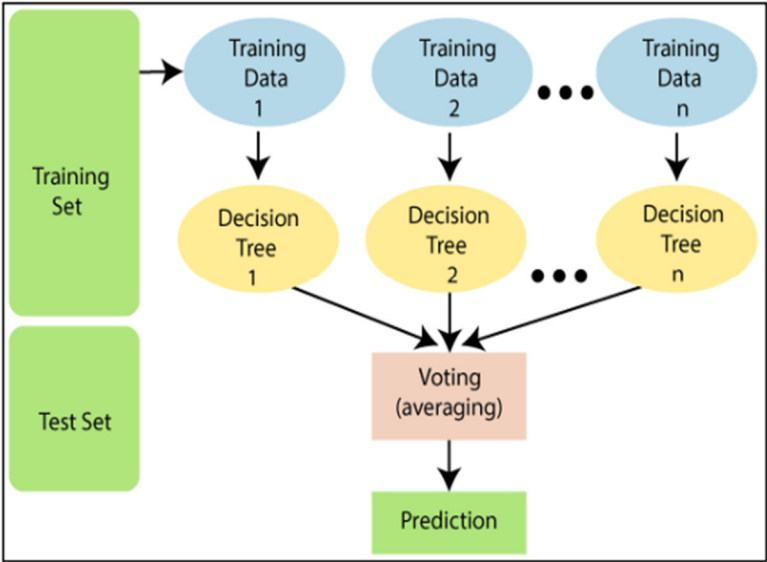
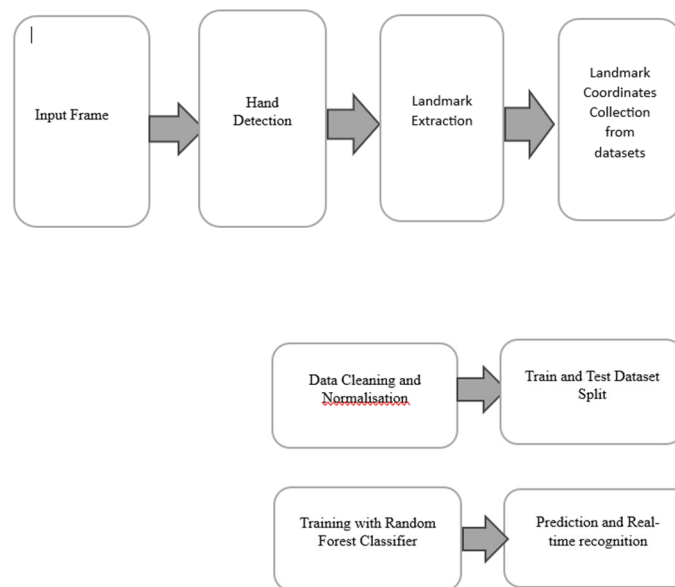


Figure 1. Flowchart of method.

3.3  
3.3.1 Architecture/ Framework Design of the method



**Figure 2.** Steps of hand gesture recognition.

#### Stage 1 - Pre-Processing of the Images to find hand Landmarks using Media Pipe

Media Pipe serves as a versatile framework empowering developers to construct cross-platform machine learning pipelines for multi-modal data, encompassing video, audio, and various time series data. Within its hand-tracking solution, a sophisticated ML pipeline operates at the backend, comprising two interdependent models: the Palm Detection Model and the Hand Landmark Model. The Palm Detection Model accurately isolates palm images, which are subsequently passed on to the Landmark Model. This model meticulously identifies 21 3D hand-knuckle coordinates within the detected hand regions, demonstrating robustness even in cases of partially visible hands. Upon successful implementation of the Palm and Hand detection model, it is applied to diverse language datasets, including the American Sign Language alphabet. Each alphabet folder undergoes hand detection, resulting in the extraction of 21 landmark points per image. Below figure shows the 21 landmark points detected by the Hand Landmark model.

2. The extracted landmark points are saved in a CSV file, with a concurrent elimination process to refine the data. Specifically, only the x and y coordinates identified by the Hand Landmark model are retained for training the ML model. The time taken for landmark extraction varies depending on the dataset size, typically ranging from 10 to 15 minutes.

#### Stage 2 - Datasets Cleaning and training:

In the initial stage, we focus solely on the x and y coordinates provided by the detector. Each image in the dataset undergoes this process to consolidate all data points into a single file. Subsequently, this file is examined using the panda's library to identify and address any null entries, which may result from blurry images that impede hand detection. To ensure the integrity of the predictive model, rows containing null entries are identified and removed from the dataset. Following this data cleaning step, the x and y coordinates are normalized to suit our system's requirements. The dataset is then partitioned into training and validation sets, with 80% of the data allocated for training purposes, allowing for the application of various optimization and loss functions. The remaining 20% is reserved for validating the model's performance.

#### Stage 3 - Testing:

Machine learning algorithms are employed to conduct predictive analysis across various sign languages. Users demonstrate signs using hand gestures captured by a camera, and the models then predict the corresponding sign. Achieving an accuracy rate of 98.9%, the model demonstrates high performance in sign recognition.

#### 4. Result Analysis

In this research, we present a comprehensive framework leveraging Media Pipe for real-time hand gesture recognition across various sign languages. Our approach utilizes a two-stage pipeline consisting of Palm Detection and Hand Landmark models to accurately localize and extract 21 3D hand-knuckle coordinates. We applied this framework to the In the American Sign Language dataset, collecting landmark points for each alphabet gesture. Subsequently, data cleaning and normalization were conducted to prepare the dataset for training, ensuring robustness against null entries and image inconsistencies. Through meticulous training and validation processes, our machine learning model achieved an impressive accuracy of 98.9% in predicting sign gestures from live camera feeds. This work underscores the efficacy of our approach in real-world scenarios, promising advancements in accessibility and communication for individuals with hearing impairments.

#### 5. Conclusion

**Our suggested approach demonstrates the effective** utilization of the Media Pipe framework for precise detection of intricate hand gestures. The training time for the model is efficiently managed, ensuring optimal efficiency. By integrating machine learning and image processing principles, we have successfully developed a real-time system capable of identifying hand gestures. Additionally, our model extends beyond merely predicting American Sign Language gestures; it accommodates custom signs provided by the users. Through extensive training and testing with diverse sign language datasets, our framework exhibits adaptability and achieves high levels of accuracy specifically tailored for the American Sign Language dataset.

##### Future Work

The envisioned sign language recognition system, initially designed for letter recognition, holds the potential for expansion into gesture and facial expression recognition realms. A shift towards displaying sentences rather than individual letter labels would enhance language translation and readability. Moreover, the scope can be broadened to encompass various sign languages, facilitating inclusivity and accessibility. Future iterations could also incorporate speech output, converting signs into spoken language. Extending the system to accommodate Indian Sign Language, with its inherent complexity involving two-hand gestures, presents an intriguing challenge. Furthermore, enhancing system response time can be achieved through improved camera and graphics support, ensuring smoother user interactions.

#### References

1. Ravikiran, J., et al. "Finger detection for sign language recognition." *Proceedings of the International Multiconference of Engineers and Computer Scientists*. Vol. 1. No. 1. 2009.
2. Srivastava, Sharvani, et al. "Sign language recognition system using TensorFlow object detection API." *Advanced Network Technologies and Intelligent Computing: First International Conference, ANTIC 2021, Varanasi, India, December 17–18, 2021, Proceedings*. Cham: Springer International Publishing, 2022.
3. Chai, Xiujuan, et al. "Sign language recognition and translation with kinect." *IEEE conf. on AFGR*. Vol. 655. 2013.
4. De Smedt, Quentin, et al. "3d hand gesture recognition using a depth and skeletal dataset: Shrec'17 track." *Proceedings of the Workshop on 3D Object Retrieval*. 2017.
5. Padmaja, N., B. Nikhil Sai Raja, and B. Pavan Kumar. "Real time sign language detection system using deep learning techniques." *Journal of Pharmaceutical Negative Results* (2022): 1052-1059.
6. Li, Y., & Quek, C. (2005). Hand gesture recognition using color and depth images. In *Proceedings of the IEEE International Conference on Multimedia and Expo* (Vol. 2, pp. 5-8). IEEE. doi:10.1109/ICME.2005.1521255
7. Starner, T., Weaver, J., & Pentland, A. (1998). Real-time American sign language recognition using desk and wearable computer based video. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 634-639). IEEE. doi:10.1109/CVPR.1998.698602
8. Soomro, T. R., Chong, A. K., & Senanayake, S. M. (2018). A survey on hand gesture recognition systems for human-machine interaction. *Artificial Intelligence Review*, 49(1), 1-40. doi:10.1007/s10462-016-9541-4
9. Mittal, S., & Zisserman, A. (2012). Hand detection using multiple proposals. In *European Conference on Computer Vision* (pp. 346-359). Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-33712-3\_25

10. Zhang, C., & Zhang, Z. (2002). Vision-based hand gesture recognition for human-computer interaction: A review. *Pattern Recognition*, 35(11), 1377-1391. doi:10.1016/S0031-3203(02)00065-5

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.