



Hand Gesture Recognition Using Deep Learning

Anusha S¹, Basavaraj², Divya Sundar S³, C R Sudhanva⁴, G N Samrudda⁵

¹Assistant Professor, Department of Computer Science and Engineering, MVJ College of Engineering, Bangalore, Karnataka, India.

^{2,3,4}Undergraduate Scholar, Department of Computer Science and Engineering, MVJ College of Engineering, Bangalore, Karnataka, India.

¹anushas@mjce.edu.in, ²basavarajsalibasavarajsal@gmail.com, ³mailtosundar01@gmail.com, ⁴crs.holla@gmail.com, ⁵samrudda13@gmail.com.

ABSTRACT

Nowadays technology is developing at a great pace where humans are getting closer and closer to interacting with computers every day. Human Computer Interaction will reach a stage where the users will have to instruct the system in a more dynamic way than the current mouse and keyboard methods. Computer programs like Gesture Recognition can greatly support such kind of dynamic control with the help of Deep Learning.

Keywords: Hand Gesture recognition, Hand Gesture Using Deep Learning, Image Identification, Image Filtering using CNN.

1. Introduction

Gesture Recognition has been an active field of research for more than two decades in human computer interaction. Recent developments in computer software and related hardware technology have provided a value-added service to the users. In everyday life, physical gestures are a powerful means of communication. They can economically convey a rich set of facts and feelings. The main goal of Human Computer Interaction is to improve the user experience between the Human and the Computer. There have been many great strides that has bought users much closer to interacting with the system, in those gesture recognition is about to one of them. For example, waving one's hand from side to side can mean anything from a "happy goodbye" to "caution". Use of the full potential of physical gesture is also something that most human computer dialogues lack. The task of hand gesture recognition is one the important and elemental problem in computer vision. With recent advances in information technology and media, automated human interactions systems are built which involve hand processing task like hand detection, hand recognition and hand tracking. Gesture recognition technology that is vision based uses a camera and motion sensor to track user movements and translate them in real time. Newer cameras and programs allow for tracking of depth data as well, which can help improve gesture tracking. Through the use of real-time image processing, users can interact with the program immediately to achieve the desired results. For example, the Xbox Kinect relied on a camera to translate players movements as part of different games. There have also been experiments performed around using a camera to track an individual's gait and then utilizing deep learning algorithms in order to assess their chance of falling, and to make recommendations on how to lower those chances. As the technology matures, gesture recognition will move beyond infotainment and will allow drivers to control other systems within the vehicle, such as heating and cooling, and to connect with smart home systems. The task of hand gesture recognition is one the important and elemental problem in computer vision. With recent advances in information technology and media, automated human interactions systems are built which involve hand processing task like hand detection, hand recognition and hand tracking. The hand detection task was however challenging because of variability in the pose, orientation, location and scale.

2. Literature Review

This paper, titled "A Sliding Window Approach to Natural Hand Gesture Recognition using a Custom Data Glove," presents a unique method for recognizing natural hand gestures through the utilization of a specially designed data glove. The objective of this project was to detect specific hand gestures by manipulating our hands in predefined patterns. The data glove employed in this study incorporated sensors capable of capturing and measuring hand movements and directions. These captured movements were then used to generate a three-dimensional representation of the hand. The main focus of this research is on the recognition of hand gestures using a data glove that is equipped with motion, bending, and pressure sensors. The authors carefully selected 31 hand gestures that are both natural and interaction-oriented, enabling them to serve as versatile controls for various computing systems. By employing the data glove and its sensors, the study aims to accurately recognize and interpret these hand gestures, enabling effective communication and control with computing systems. Through their investigation, the authors explore the potential of the custom data glove in capturing and analyzing hand movements, allowing for the recognition of the predefined hand gestures. The findings of this study contribute to the development of gesture-based interfaces, offering new possibilities for intuitive and natural interaction with computing systems. [1]

The research paper titled "3-D Hand Motion Tracking and Gesture Recognition Using a Data Glove" introduces an innovative approach for tracking hand motion and recognizing gestures by employing a data glove. This study builds upon the principles outlined in the Sliding Window Project, exploring

similar concepts and techniques. The authors of the paper have developed a comprehensive system for 3-D hand motion tracking and gesture recognition, utilizing a specifically designed data glove known as the KHU-1. The KHU-1 data glove incorporates three tri-axis accelerometer sensors, one controller, and a Bluetooth module, enabling the transmission of hand motion signals wirelessly to a connected PC. Furthermore, the researchers have implemented a sophisticated 3-D digital hand model, which is employed for accurate hand motion tracking and recognition. This digital hand model is constructed based on the principles of kinematic chain theory, utilizing ellipsoids and joints to represent the various parts of the hand. By analyzing the signals received from the data glove and mapping them onto the 3-D digital hand model, the system is able to accurately track hand movements and recognize gestures. The findings of this study contribute to the advancement of hand motion tracking and gesture recognition technologies, providing a framework for more intuitive and immersive human-computer interaction. The developed system showcases the potential of data gloves in capturing and interpreting hand movements, opening up possibilities for various applications in fields such as virtual reality, gaming, and robotics. [2]

This paper titled "Home outlet and LED array lamp controlled by a smartphone with hand gesture recognition" presents a system that allows users to control home outlets and LED array lamps using a smartphone through hand gesture recognition. This project was used to control home automation devices with the help of a smartphone. It measured the readings from the in-built sensor of phone like gyroscope and accelerometer to determine gestures and activate them accordingly. In this paper, they designed a home outlet and a LED array lamp controlled by hand gesture recognition with a smartphone that has a system composed of two parts: a smartphone's application and a wireless remote-control unit (WRCU). The application can read the accelerometer and gyroscope in a smartphone by means of hand gesture recognition and send a control command to the wireless remote-control unit. The wireless remote-control unit can decode the received message and echo the corresponding action. This design incorporates a LED array lamp as the home outlet used to control both the on/off function and the dimming function.[3]

This paper titled "Home appliance control by a hand gesture recognition belt in LED array lamp case" presents a novel approach to control home appliances using a hand gesture recognition belt. This project had an accelerometer and a gyroscope to discern hand gestures and a filter to remove jitter. The receiver unit decodes the info and echoes the action. In this paper, they designed a wearable appliance gesture remote controller that is composed of two parts: a hand gesture recognition belt and a receiver unit. The controller uses both an accelerometer and a gyroscope to discern hand gestures and a Kalman filter to reduce some jitter noise such as an individual's hands that are trembling slightly. The receiver unit can decode the received information and echo the corresponding action. This design incorporates an LED array lamp as the home appliance used to control the on/off function and the dimming function.[4]

This paper titled "A real-time hand gesture recognition approach based on motion features of feature points" presents an approach for real-time hand gesture recognition. This project uses depth sensors like Microsoft Kinect and Leap Motion to detect gestures. This application tried to detect and pinpoint the exact location of each joint in fingers, etc. This only tracks your fingers and doesn't perform any other functions. This paper presents a precise tracing of feature points including palm center, fingertips and joints by using Kinect. A novel recognition method based on precise motion features of these feature points is also presented. Having been tested with a series of applications, their method is proved to be robust and suitable for further application in real-time HCI systems.[5]

This paper titled "Technologies of Hand Gesture Recognition Based on Vision" provides an overview of various technologies used for hand gesture recognition based on vision. This paper explores the advancements and potential applications of hand gesture recognition. The authors discuss various computer vision techniques that can be utilized to recognize and interpret hand gestures, enabling more natural and intuitive human computer interaction. The paper highlights the importance of this technology in fields like healthcare, virtual reality, and robotics, showcasing its potential to revolutionize these domains. Overall, the authors provide valuable insights into the capabilities and future scope of hand gesture recognition technology based on vision.[6]

The paper titled "Visual interpretation of hand gestures for human-computer interaction: A review" provides a comprehensive review of visual interpretation techniques for hand gestures in the context of human-computer interaction. The authors analyze various approaches used for hand gesture recognition, including appearance-based methods, model-based methods, and data-driven approaches. They discuss the challenges and opportunities in hand gesture recognition and highlight the importance of robust and real-time solutions. The review paper serves as a valuable resource for researchers and practitioners in the field of gesture-based human-computer interaction. It was published in the IEEE Transactions on Pattern Analysis and Machine Intelligence in July 1997.[7]

This paper titled "Review of Gestures Recognition Based on Vision" provides a comprehensive review of gesture recognition techniques based on vision. It presents a comprehensive overview of gesture recognition technology based on vision. The authors provide an extensive review of various computer vision techniques and algorithms employed in gesture recognition systems. They discuss the challenges and advancements in gesture recognition, including feature extraction, gesture classification, and real-time tracking. The paper also explores the applications of gesture recognition in areas such as human-computer interaction, sign language interpretation, and virtual reality. The authors' review offers valuable insights into the state-of-the-art techniques and future directions in vision-based gesture recognition, making it a useful resource for researchers and practitioners in the field.[8]

This paper titled "Research on Static Gesture Recognition Algorithm Based on Neural Network" presents a study focused on static gesture recognition using neural network algorithms. This research focuses on the development of a static gesture recognition algorithm based on neural networks. The study investigates the application of neural network techniques to accurately recognize and classify static hand gestures. By analyzing various neural network architectures and training algorithms, the research proposes an efficient approach for static gesture recognition. The algorithm's effectiveness is evaluated through experiments and comparisons with existing methods. The findings contribute to the field of gesture recognition by providing insights into the application of neural networks for static gesture analysis, showcasing its potential for real-world applications.[9]

The paper titled "Sign language recognition and translation with Kinect" presents a study on utilizing Kinect, a depth-sensing camera system, for sign language recognition and translation. The authors explore the use of Kinect to capture and analyze hand gestures and movements in sign language. They propose a framework that includes gesture segmentation, feature extraction, and classification stages to recognize sign language gestures. Experimental results demonstrate the effectiveness of the proposed approach in accurately recognizing and translating sign language. The research contributes to the field of sign language technology by providing a practical solution for real-time recognition and translation using Kinect. The paper was presented at the IEEE Conference on Automatic Face and Gesture Recognition (AFGR) in 2013.[10]

3. Working Principle

The Gesture Recognition system is a carefully designed program that prioritizes functionality, aiming to accurately identify and interpret gestures. It consists of four interconnected parts, each playing a vital role in the overall system performance. The system's effectiveness heavily relies on the presence and cooperation of all its components, as the absence of even one part can diminish its functionality. These components work in unison to ensure seamless operation and dependable results. Each part of the system contributes its unique function, with the absence of even one part compromising the system's capability to deliver accurate gesture recognition. The Gesture Recognition system represents an integrated framework where every component harmoniously interacts to fulfill the overarching objective. Its comprehensive design incorporates the essential elements required for accurate gesture identification and interpretation. At the core of Gesture Recognition using Deep Learning is the detection system, responsible for capturing and processing input data such as images or video streams to identify gestures. It utilizes advanced computer vision techniques and deep learning algorithms to precisely locate and recognize gestures in real-time or near real-time scenarios. Complementing the detection system is the training system, which is responsible for training the deep learning models used in gesture recognition. Through supervised learning, the training system prepares a diverse dataset of labeled gesture samples, allowing the models to accurately identify and classify various gestures. These interdependent detection and training systems work together to deliver robust gesture recognition capabilities. The trained models produced by the training system empower the detection system to accurately recognize and interpret detected gestures. By continuously refining and improving the models through additional data, the system achieves enhanced accuracy and adaptability. The detection system is further divided into two distinct parts, catering to different use cases: one focusing on single-hand gestures for specific actions, and the other capable of detecting gestures from both hands simultaneously, providing real-time feedback and engagement. Overall, the Gesture Recognition system demonstrates a comprehensive approach to gesture interpretation, expanding its applications and improving user experiences.

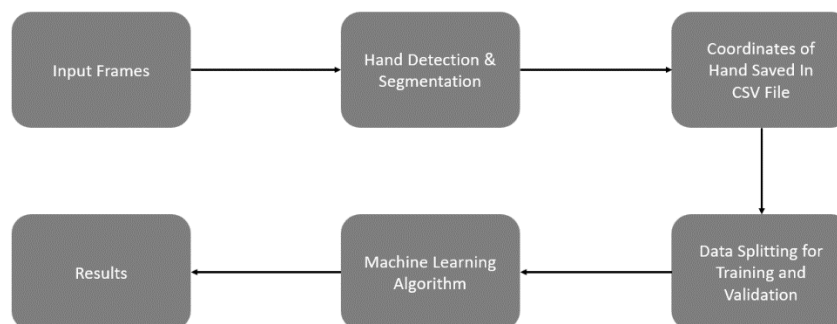


Figure 1: Architecture Diagram

4. IMPLEMENTATION

The implementation of the Gesture Recognition system involves several key steps. Firstly, a diverse dataset of labeled gesture samples needs to be collected. This dataset should encompass a wide range of hand gestures performed by users, considering various angles, lighting conditions, and hand positions to ensure robust training. Once the dataset is gathered, preprocessing techniques are applied to enhance the quality of the input data. Image resizing, normalization, and noise reduction methods can be employed to ensure consistent input for the detection and training systems. The next step is to implement the detection system, which is responsible for capturing and processing the input data to detect the presence of gestures. Advanced computer vision techniques, such as hand detection algorithms, are leveraged to identify and locate hands within the input. Hand landmarks or key points can be extracted using methods like hand tracking or pose estimation. These landmarks are then stored for further analysis.

Simultaneously, the training system is developed to train deep learning models for gesture recognition. This involves utilizing a supervised learning approach where the labelled gesture samples from the dataset are used to train the models. Deep learning frameworks such as TensorFlow or PyTorch can be implemented, and architectures like convolutional neural networks (CNNs) or recurrent neural networks (RNNs) can be utilized to optimize the models' performance. The trained models are continuously refined and improved through a process of retraining with additional data. This iterative process helps enhance the models' accuracy, adaptability, and generalization capabilities. Techniques such as data augmentation, regularization, and hyperparameter tuning can be employed to optimize the models' performance further. Once the detection and training systems are developed, they are integrated to work in tandem. The detection system relies on the trained models to recognize and interpret the detected gestures. Realtime or near real-

time processing is implemented to enable quick and seamless gesture recognition. A user interface is designed to allow users to interact with the system using gestures, and visual feedback is provided to indicate the recognized gestures and corresponding actions.

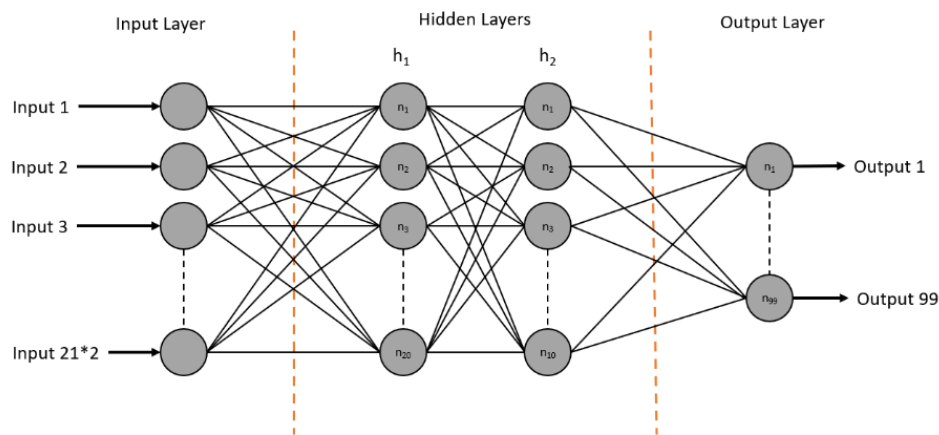


Figure 2: Neural Network Model for Static Gestures.

Step By Step Implementation:

1. Training System: The program can train the model on user defined gestures.
2. Detection System: the program can detect gestures on user trained gestures.
3. Data Collection: The program collects the data of each gesture.
4. Preprocessing: The collected gesture is processed for training or detection.
5. Model Optimization: The neural network model is optimized to process the data of gestures.
6. Integration and Real-time Processing: The model is trained upon the processed data.
7. User Interface and Feedback: The result of the gesture is model is displayed on a user interface.
8. Testing and Evaluation: It is an additional feature, if the user wishes they can see the spread of data and accuracy of data.

The program has been meticulously developed with a focus on providing users with a seamless experience in detecting and utilizing specific gestures for computerized actions. It eliminates the need for specialized hardware typically required to perform actions like mouse cursor control and keyboard input. With this program, users have the freedom to program each gesture to a specific control set, allowing for a high level of customization. For example, imagine a user who wants to execute the action of pressing the right arrow button on the keyboard. They can simply demonstrate a gesture where their index finger points towards the right, and configure the program to mimic the corresponding keyboard input. Consequently, whenever the user performs this particular gesture, the program responds by virtually triggering the action of pressing the right arrow key. This intelligent mapping of gestures to actions empowers users to interact with their computer systems in a more natural and intuitive manner. Moreover, the program offers users the flexibility to train the model according to their unique needs and preferences. By associating each gesture with a specific action, users can create a highly personalized gesture recognition system. This capability opens up a world of possibilities, allowing individuals to define their own set of gestures and assign them to perform actions that align with their workflow or specific requirements. The program's versatility and adaptability make it an invaluable tool for enhancing user experience and expanding the capabilities of gesture-based control. By leveraging the power of hand gestures, users can unleash a new level of interaction with their computer systems, enabling more efficient and intuitive ways of navigating, controlling, and manipulating digital content. The program's user-centric design philosophy ensures that individuals can harness the potential of gesture recognition technology without the limitations of dedicated hardware, providing a seamless and empowering user experience.

Due to the module's low compatibility and lack of refinement, we are faced with the need to run the program exclusively on the CPU, rather than utilizing the GPU. This unfortunate limitation significantly impacts the program's performance, especially considering that graphical data, such as images and videos, typically benefit from the specialized processing capabilities of GPUs. GPUs are specifically designed to handle the intensive computations involved in graphical data processing. However, due to the inherent constraints of the mediapipe module, it cannot take advantage of the GPU's processing power. As a result, the module is limited to relying solely on the CPU for its computations. This poses a challenge because CPUs are generally not as optimized for graphics processing as GPUs are. Consequently, the program may experience reduced performance and efficiency when dealing with complex graphical data. Despite this limitation, we must adapt our approach and optimize the program as best as possible to maximize its performance within the CPU's capabilities. While it may not match the speed and efficiency of GPU processing, with careful optimization and fine-tuning, we can still strive to achieve satisfactory results in gesture recognition using the available CPU resources.

Through these implementation steps, the Gesture Recognition system can recognize and interpret gestures, providing users with intuitive and efficient control over computer functions. This implementation opens up possibilities for enhanced human-computer interaction, virtual reality experiences, gaming applications, and more.

5. RESULTS

This gesture recognition using landmark detection and deep learning is a robust system that can effectively detect and classify hand gestures based on landmark information. The system is capable of real-time or near real-time recognition of various gestures performed by users. By leveraging landmark detection algorithms, the system can identify and extract specific key points or landmarks on the hand, representing its position and shape. These landmarks serve as essential features for gesture recognition. Through training with deep learning techniques, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), the system learns to recognize patterns and correlations between the extracted landmarks and the corresponding gestures. It optimizes its internal parameters and weights to accurately classify and associate the given landmarks with specific gestures.

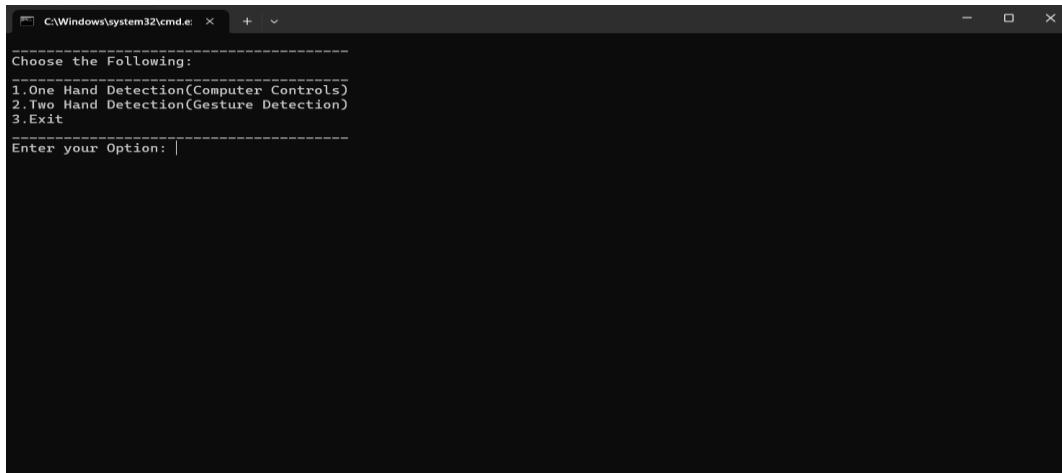


Figure 3: Screenshot of the Main Menu

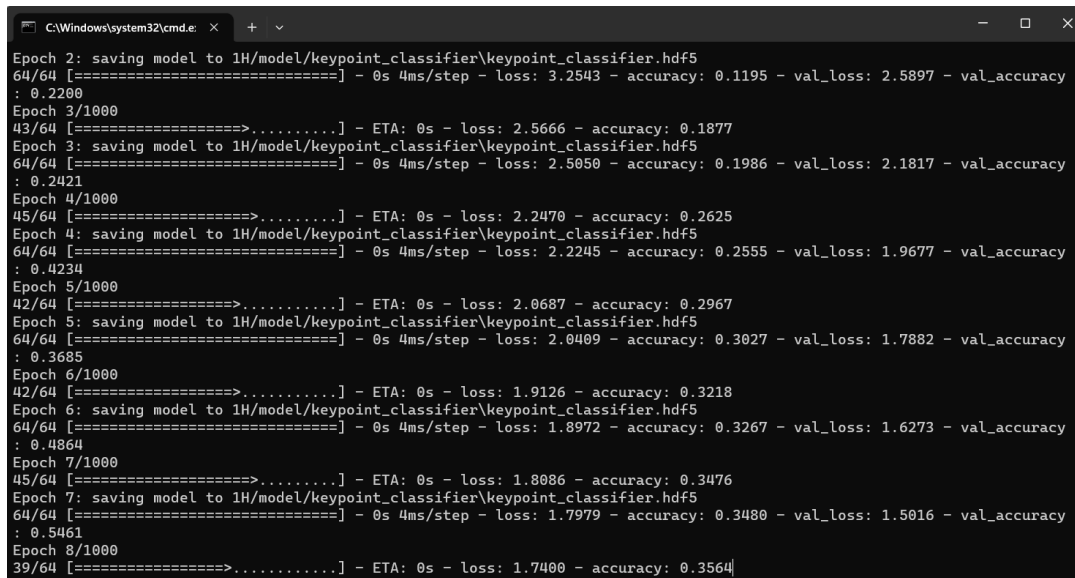


Figure 4: Screenshot of the program during the Training

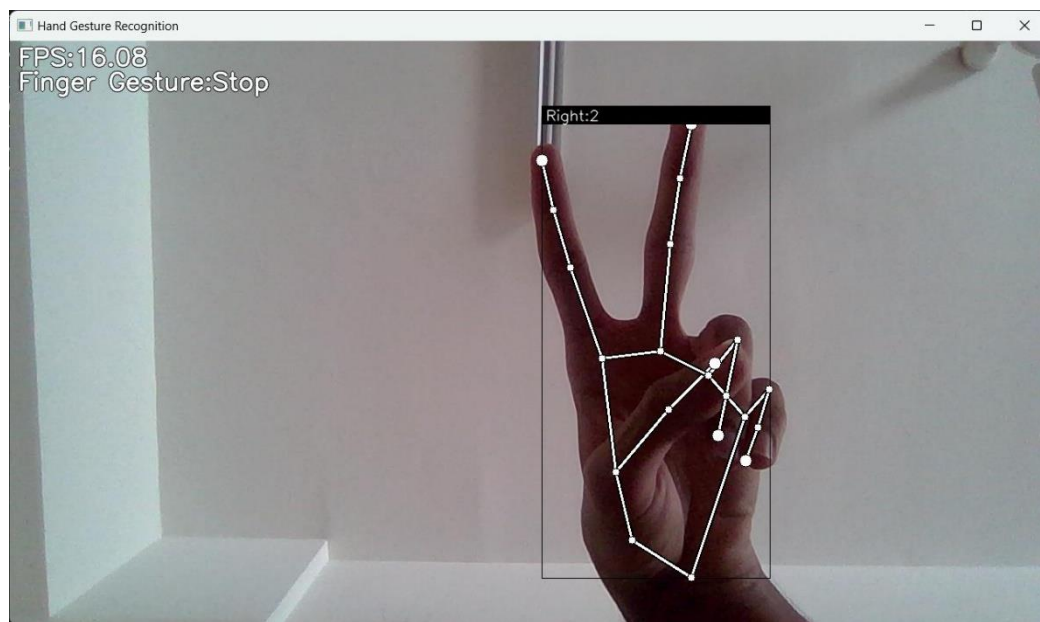


Figure 5: Screenshot of the Program During the One Hand Gesture Recognition

Finally, the program can recognize the gestures in real time and associate certain hand poses to certain gestures, thereby informing the user regarding the certain hand gestures. This identification of gestures can be used in several scenarios in real life. We have also given the user to train their own gestures without needing to stick to the pre-defined set of gestures which is available in the internet. Thereby it allows the user to customize the program as per their requirement and modify it accordingly. Along with this the user can also associate each gesture with their own computer functions like mouse control which includes cursor movement with mouse button click, link certain gestures with keyboard key control which can also allow the user to use the program in various scenarios like PowerPoint presentation, Microsoft Word scrolling and other various functions which uses. keyboard and mouse. This system has the ability to process real-time input data, extract landmarks from hand movements, and predict the corresponding gesture label or class. This output can be utilized for a wide range of applications, including virtual reality interactions, human-computer interfaces, robotics, and more. The success of the final result is measured by the system's accuracy, robustness, and real time performance in correctly identifying and classifying gestures across various scenarios and user inputs. Continuous evaluation, refinement, and training of the system with diverse and representative data can further enhance its performance and generalization capabilities.

6. ADVANTAGES

- It increases the Ease of Use.
- Provides hassle free work environment.
- Better interaction while working.
- Features like PowerPoint slide control using gestures reduce the time to change the slides when users are away from keyboards.
- It can detect certain simple sign language gestures which can help the user in understanding the gestures.
- It works with a combination of both Static and Dynamic gestures.
- It doesn't need a lot of computational resources to perform calculations and to detect gestures.
- Since we are training the data, it doesn't a lot of input data as we can train the same data repeatedly to improve accuracy

7. APPLICATION

- Human-Computer Interaction: Gesture recognition can be used to enable more intuitive and natural interaction between humans and computers. It can be applied in applications such as controlling video games, navigating virtual reality environments, and operating smart devices using hand gestures.
- Sign Language Recognition: Deep learning-based gesture recognition can assist in interpreting and translating sign language into text or speech. This technology can facilitate communication between the hearing-impaired and the general population, bridging the gap in understanding and accessibility.

- **Surveillance and Security:** Gesture recognition can be employed in surveillance systems to detect and recognize specific gestures that indicate suspicious or threatening behavior. It can help in identifying potential security threats and alerting authorities in real-time.
- **Robotics:** Deep learning-based gesture recognition can be used in robotic systems to interpret human gestures and commands. This allows for more seamless human-robot interaction, enabling robots to respond to hand signals and gestures to perform specific tasks or provide assistance.
- **Automotive Industry:** Gesture recognition can enhance the user experience in vehicles by allowing drivers to control various functions without taking their hands off the steering wheel or eyes off the road. It can be utilized for tasks like adjusting music volume, answering calls, or activating specific features.
- **Healthcare:** Gesture recognition using deep learning can be employed in rehabilitation and physical therapy settings. It can assist in tracking and analyzing patients' movements and providing real-time feedback, helping them to perform exercises correctly and monitor their progress.

8. FUTURE SCOPE

The future scope of hand gesture recognition holds immense potential in various domains, ranging from human-computer interaction to healthcare and robotics. With advancements in computer vision and machine learning, hand gesture recognition technology can enable more natural and intuitive ways of interacting with devices and machines, eliminating the need for physical interfaces. It can enhance virtual reality experiences, facilitate remote communication, improve accessibility for individuals with disabilities, and revolutionize industries like healthcare by enabling touchless control and precise gesture-based surgical procedures. Furthermore, as the technology evolves, the integration of hand gesture recognition with artificial intelligence and Internet of Things (IoT) devices opens up new possibilities for smart homes, autonomous vehicles, and intelligent robotic systems that can interpret and respond to human gestures accurately and seamlessly.

9. CONCLUSION

In conclusion, hand gesture recognition is a powerful technology that enables intuitive and efficient human-computer interaction. By accurately recognizing and interpreting hand gestures, this technology offers users a natural and intuitive way to control computer functions, virtual reality experiences, and gaming applications. Through the implementation of advanced computer vision techniques and deep learning algorithms, coupled with a modular system design, hand gesture recognition systems can achieve high accuracy and versatility. The continuous refinement and improvement of deep learning models through retraining and the integration of detection and training systems further enhance the system's performance and adaptability. With its wide range of applications and potential for enhancing user experiences, hand gesture recognition technology holds promise for revolutionizing various industries and transforming the way we interact with digital devices.

References

- [1] Granit Luzhnica, Elizabeth Lex, Viktoria Pammer. A Sliding Window Approach to Natural Hand Gesture Recognition using a Custom Data Glove. In: 3D User Interfaces (3DUI); 2016 IEEE Symposium on; 2016 Mar 19; New York: IEEE; 2016; p.81-90.
- [2] Ji-Hwan Kim, Nguyen Duc Thang, Tae-Seong Kim. 3-D hand Motion Tracking and Gesture Recognition Using a Data Glove. In: Industrial Electronics; 2009 IEEE International Symposium on; 2009 July 5; New York: IEEE;2009; p.1013-1018.
- [3] Hung CH, Bai YW, Wu HY. Home outlet and LED array lamp controlled by a smartphone with a hand gesture recognition. In: Consumer Electronics (ICCE); 2016 IEEE International Conference on; 2016 Jan 7; New York: IEEE;2016; p.5-6.
- [4] Hung CH, Bai YW, Wu HY. Home appliance control by a hand gesture recognition belt in LED array lamp case. In: Consumer Electronics (GCCE); 2015 IEEE 4th Global Conference on; 2015 Oct 27; New York: IEEE;2015; p. 599-60.
- [5] She Y, Wang Q, Jia Y, Gu T, He Q, Yang B. A real-time hand gesture recognition approach based on motion features of feature points. In: Computational Science and Engineering (CSE); 2014 IEEE 17th International Conference on; 2014 Dec 19; New York: IEEE;2014; p.1096- 1102.
- [6] Li-juan Sun, Li-cai Zhang and Cai-long Guo, "Technologies of Hand Gesture Recognition Based on Vision [J]", Computer Technology and Development, vol. 18, no. 10, pp. 214-216, 2008.
- [7] Pavlovic, V., Sharma, R. & Huang, T. (1997), "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Transactions on Pattern Analysis and Machine Intelligence, July, 1997. Vol. 19(7), pp. 677 -695.
- [8] Jing-guo Yi, Jiang-hua Cheng and Xi-shu Ku, "Review of Gestures Recognition Based on Vision [J]", Computer Science, vol. 43, no. 6A, pp. 103-108, 2016.
- [9] Gao Long, Research on Static Gesture Recognition Algorithm Based on Neural Network [D], Ning Xia University, 2017.
- [10] Chai, Xiujuan, et al. "Sign language recognition and translation with Kinect Archived 2021-01-10 at the Wayback Machine." IEEE Conf. on AFGR. Vol. 655. 2013.