IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HAND GESTURE RECOGNITION USING MEDIAPIPE AND OPENCV

[1]Mrs.M.Stella Inba Mary, [2] M.Sakthi Anand, [3] A.Sakthi Manikandan, [4] M.Senthamilmuthu

[1] Assistant Professor, [2] B.E student, [3] B.E student, [4] B.E student
Department of computer Science and Engineering
P.S.R Engineering College, Sivakasi.

ABSTRACT:
Gesture recognition is considered important for the development of business 4.0 technologies in the field of human-computer interaction (HCI); It allows computers to capture and interpret gestures to execute commands without the need for a physical touch device. Media Pipe provides pre-trained hand models for extracting important symbols and hand gesture features in live video streams. These core features provide access to a proprietary machine learning model. OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning library. It provides a variety of tools, functions, and algorithms for a variety of tasks related to computer vision, image processing, and video analysis in real time. The integration of Media Pipe and OpenCV paves the way for significant advances in human-computer interaction through gesture recognition. This project explores the use of this new technology to enable rapid translation and response to human movements. Gesture recognition has a wide range of exciting applications, from improving accessibility to transforming gaming experiences and interactive experiences.

1. INTRODUCTION

In the ever-evolving landscape of human-computer interaction, the ability to comprehend and respond to hand gestures stands as a remarkable frontier. Hand gesture recognition offers a seamless and intuitive means of communication between humans and technology. It not only enhances accessibility but also opens doors to a spectrum of applications, spanning from virtual mouse control to interactive gaming experiences.

Hand gesture recognition is a fascinating and rapidly advancing field of computer vision and human-computer interaction (HCI) that focuses on the interpretation and understanding of hand movements and by computer systems. Computers can able to recognize and respond to the gestures made by users, which is used to interact with the systems. Hand gesture recognition is a technology that enables computers to interpret and understand human hand movements and gestures, converting them into actionable commands or meaningful interactions. This field of computer vision and human-computer interaction has gained significant attention and importance due to its potential applications in various domains. The primary goal of hand gesture recognition is to bridge the gap between human communication and machine interaction, making it more intuitive and natural. By capturing and analyzing the spatial and temporal information of hand movements, as well as the shape and orientation of the hand and fingers, systems can recognize specific gestures and associate them with particular functions or responses. This technology has the potential to enhance user experiences, improve accessibility for individuals with disabilities, and revolutionize the way we interact with digital devices and environments. At the core of this technological leap lies the integration of two powerful tools: Media Pipe and OpenCV. Media Pipe, a versatile framework for machine learning based solutions, provides an arsenal of pre-trained models and features designed for the analysis of human hand movements. OpenCV, a renowned open-source computer vision library, complements this with a rich toolkit for real-time

image processing, enabling the detection and tracking of hands in video streams or images. Hand gesture recognition systems can be used in a wide range of applications, from controlling electronic devices with a wave of your hand to improving accessibility for individuals with physical disabilities.

2. LITERATURE SURVEY

Indriani [1] developed a simple user guide application using the MediaPipe framework. A user manual is often referred to as a document for communication or a manual for helping people with a system. User guides provide a step-by-step description of how to deal with a particular system and assist users in relieving user frustrations by providing them with ways to identify, access, and resolve their own problems as they arise. In his experiments, he used Kinect to capture live images, then learned various motion data, recognized all gestures, and recognized gestures to display language, as described in the app's user manual. Users can store user information based on gesture recognition.

Uzma Akram [2] proposed a multidimensional 2D CNN (two-dimensional convolutional neural network) model to recognize gestures. The proposed model consists of various steps and methods for detecting hand poses from images obtained from depth data. The hyperparameters of the proposed model were modified by experimental observations. Our public proofs: Kaggle, First Person and Dexter are used independently to train and test the scheme. The accuracy of this method is 99.99%, 99.48%, and 98% when using the Kaggle gesture dataset, First Person gesture dataset, and Dexter dataset, respectively. In addition, the results obtained for F1 and AUC scores are close to positive. Compared with the state-of-the-art methods, it can be seen that the proposed model is better than previous methods.

Jannatun Meghna [3] proposed a method for behavior-based character recognition using OpenCV. Deep learning is an exciting way to use computer vision and machine learning techniques to leverage language. These tools can be used in many areas such as language recognition, gesture-based user interface, and even educational applications. Fast, accurate and user-friendly human-computer interaction (HCI) requirements and skills. Computers can understand signs and symbols, but recognizing signs that people draw in front of a camera is still a new concept. To achieve this goal, time-of-flight (ToF) cameras, Kinect sensors, etc. Many experiments have been conducted using different sensors such as special measures or special algorithms. Our research recommends using the video camera that comes with almost every computer for this type of work.

Ahmad Puad Ismail [4] project focuses on how the system sees, recognizes and acts on information through computer vision and related issues that affect the body, orientation, position and change. In order for this project to be developed correctly, different types of gestures such as numbers and sign language need to be developed in this system. The images obtained from the live video are analyzed by the Haar cascade classifier before the image is processed to detect hand movements, in other words, the hand visible in the frame. This project will use ROI theory to complete applied Python programs. Since different hardware is used to read the real-time video input code, the interpretation of the results will focus on the simulation part.

3. PROPOSED SYSTEM

Hand Gesture Recognition using MediaPipe and OpenCV refers to the process of identifying and localizing the presence of one or more human hands in an image or video frame using MediaPipe, an open-source framework developed by Google. MediaPipe offers a pre-trained model specifically designed for hand detection, making it a convenient choice for applications that require identifying and tracking hands in real-time. MediaPipe provides a pre-trained hand detection model that uses machine learning techniques, particularly convolutional neural networks (CNNs), to identify the position of hands within an image or video frame. After detecting a hand, the model also estimates a set of landmarks, which are key points on the hand. These landmarks include points on the fingertips, the palm center, and the base of the hand, among others. The hand detection model also defines a bounding box around the detected hand, which encapsulates the hand's approximate position and size within the frame. Then, The Hand feature extraction which will use the Media Pipe Framework refers to the process of identifying and extracting relevant characteristics or landmarks from a detected hand in an image or video frame using MediaPipe, an open-source framework developed by Google. These extracted features can include the positions and movements of key points on the hand, such as fingertips, palm center, and other landmarks. The extracted features are typically represented as numerical values or vectors, making them suitable for further processing and analysis. These

representations can be in the form of 2D or 3D coordinates, angles, distances, or other metrics. Using MediaPipe simplifies the initial steps of hand detection and tracking, making it easier to develop a hand gesture recognition system. The hand tracking and pose estimation models provided by MediaPipe are trained on large datasets and work well in a variety of scenarios, making them a valuable resource for developers interested in building hand gesture recognition applications. MediaPipe provides a pre-trained hand tracking model that can detect and track hands in real-time within video or image frames. This model identifies the hand's location, orientation, and landmarks, which are the key points on the hand, including the fingertips, knuckles, and palm center. Once the hand is tracked, MediaPipe can estimate the pose of the hand by localizing the 3D positions of the hand landmarks in the 2D image frame. This provides information about the hand's position and orientation.After obtaining the hand pose, you can implement a custom algorithm or machine learning model to recognize specific gestures based on the relative positions and movements of the hand landmarks. Common approaches include using rule-based methods, machine learning models, or deep learning networks to interpret the hand pose data and classify it into predefined gesture categories.

Hand gesture classification using involves the process of recognizing and categorizing hand gestures from images or video frames captured by a camera or other image sources. OpenCV (Open Source Computer Vision Library) is a popular open-source computer vision and image processing library that provides a wide range of tools for working with visual data. This model uses OpenCV for hand detection and tracking. Various techniques can be applied, such as skin color detection, background subtraction, and contour analysis to locate and track the hand within the frames and It will extract the relevant features from the segmented hand region. These features can include

color histograms, texture information, or geometric properties of the hand. Then develop a machine learning model or use a pre-trained model to classify the hand gestures. OpenCV is not a machine learning library on its own, but it can be used in combination with libraries like scikit-learn, TensorFlow, or PyTorch to build and train the gesture classification model. Hand gesture classification using OpenCV can be a complex but rewarding task. OpenCV provides essential tools for image processing and computer vision, while machine learning libraries can help build and train effective classifiers for recognizing hand gestures. The system should be tested and optimized for accuracy and real time performance. Gesture operations typically refer to actions or commands performed through physical movements or gestures, often with the help of technology or devices equipped with sensors and cameras to recognize and interpret these movements. These gestures can be used to interact with computers, smartphones, tablets, and other digital devices. Gesture operations are commonly used in user interfaces to control or manipulate software or hardware without the need for physical buttons or touchscreens.

4.SYSTEM IMPLEMENTATION AND MODULES DESCRIPTION:

Hand Detection: Hand detection in machine learning is the process of identifying and locating human hands within an image or a video frame. It is a fundamental step in various applications, such as gesture recognition, sign language translation, virtual reality, and human-computer interaction. Hand detection involves recognizing the presence of hands in an image or video and often includes localizing the position and possibly the orientation of these hands. Hand detection in machine learning can be challenging due to variations in hand shapes, skin tones, lighting conditions, occlusions, and background clutter. Effective hand detection models are often trained on diverse datasets and designed to handle these challenges. The output of a successful hand detection system is a set of bounding boxes or regions of interest that can be used for further analysis or interaction, such as recognizing gestures or controlling applications through hand movements.

Hand Landmark: In the MediaPipe framework, the "Hand Landmarks" refer to a set of key points or landmarks that are detected and tracked on a human hand within an image or video frame. These landmarks provide detailed information about the hand's shape and position and are crucial for applications like hand gesture recognition, sign language translation, and virtual reality interactions. MediaPipe's hand landmarks model detects 21 key points on a hand, offering a rich representation of hand movements and poses.

1. Wrist (Landmark 0): This landmark represents the wrist's position,

serving as the starting point for the other hand landmarks.

2. Thumb (Landmark 1 to 4): The thumb landmarks represent the base of

the thumb, the knuckle, the middle joint, and the tip of the thumb.

3. Index Finger (Landmark 5 to 8): These landmarks correspond to the
index finger's base, knuckle, middle joint, and fingertip.
4. Middle Finger (Landmark 9 to 12): These landmarks represent the base,
knuckle, middle joint, and tip of the middle finger.
5. Ring Finger (Landmark 13 to 16): These landmarks correspond to the
base, knuckle, middle joint, and tip of the ring finger.
6. Little Finger (Landmark 17 to 20): These landmarks represent the base,
knuckle, middle joint, and tip of the little finger.

Landmark Detection:

Landmark detection using MediaPipe refers to the process of identifying and localizing specific key points or landmarks within an object or region of interest in an image or video frame using the MediaPipe framework. These landmarks are often associated with the detection and tracking of human body parts, objects, or features, and they provide crucial information for various computer vision and machine learning tasks. MediaPipe offers pre-trained models and tools for landmark detection in a wide range of applications, such as hand tracking, facial feature detection, and human pose estimation. The concept of landmark detection can vary depending on the specific application. In hand tracking with MediaPipe, landmark detection involves identifying and tracking key points on a person's hand, which can be used for gesture recognition, sign language translation, and interactive applications.

FUNCTIONAL DIAGRAM:

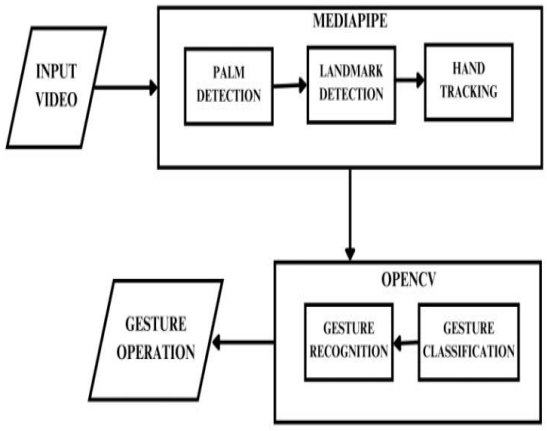

Figure: Flow Diagram

Hand Movement Tracking: Media Pipe tracks the detected hands across video frames, maintaining continuity and understanding hand movement and orientation. This tracking is essential for real-time applications, as it allows the system to follow the hands as they move within the frame. Media Pipe uses a deep learning-based model to detect the presence of hands in a video stream or image. This model identifies the regions in the frame that likely contain one or more hands.

Gesture Recognition: Using MediaPipe simplifies the initial steps of hand detection and tracking, making it easier to develop a hand gesture recognition system. The hand tracking and pose estimation models provided by MediaPipe are trained on large datasets and work well in a variety of scenarios, making them a valuable resource for developers interested in building hand gesture recognition applications. MediaPipe provides a pre-trained hand tracking model that can detect and track hands in real-time within video or image frames. This model identifies the hand's location, orientation, and landmarks, which are the key points on the hand, including the fingertips, knuckles, and palm center.

Gesture Classification: Hand gesture classification using involves the process of recognizing and categorizing hand gestures from images or video frames captured by a camera or other image sources. OpenCV (Open-Source Computer Vision Library) is a popular open-source computer vision and image processing library that provides a wide range of tools for working with visual data. This model uses OpenCV for hand detection and tracking. Various techniques can be applied, such as skin color detection, background subtraction, and contour analysis to locate and track the hand within the frames and It will extract the relevant features from the segmented hand region.

Gesture Operation: Gesture operations typically refer to actions or commands performed through physical movements or gestures, often with the help of technology or devices equipped with sensors and cameras to recognize and interpret these movements. These gestures can be used to interact with computers, smartphones, tablets, and other digital devices. Gesture operations are commonly used in user interfaces to control or manipulate software or hardware without the need for physical buttons or touchscreens.

Hand gestures used to control electronic devices refer to specific hand movements and positions that are employed to interact with and manipulate digital devices or systems. These gestures often utilize sensors, cameras, or motion-tracking technology to recognize and interpret the user's hand movements, allowing them to control various functions and features of electronic devices.

5.SAMPLE OUTPUT:

Figure: Cursor Moving

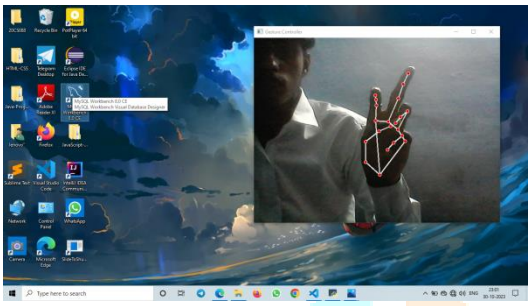

Figure: Left Click

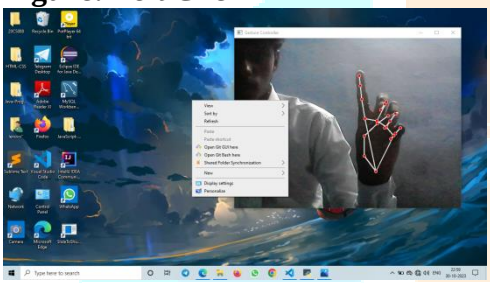

Figure: Right Click

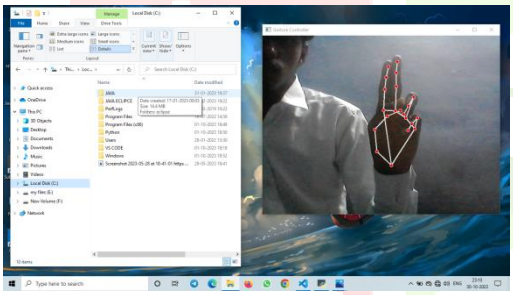

Figure: Double Click

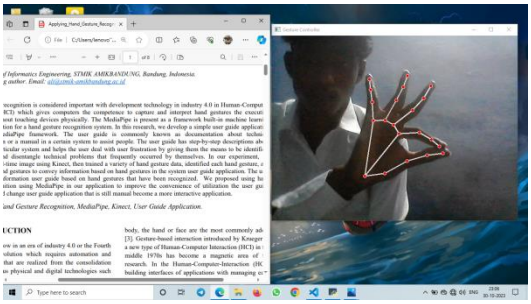

Figure: Scroll

6. CONCLUSION:

Hand gesture recognition using MediaPipe and OpenCV is a promising technology with a wide range of potential applications. MediaPipe and OpenCV are both powerful and open source libraries that provide a variety of tools and resources for developing hand gesture recognition systems. MediaPipe is a cross-platform framework for building real-time computer vision and machine learning pipelines. It provides a number of pre-built models for hand detection, tracking, and gesture recognition. OpenCV is a popular open source computer vision library that provides a wide range of algorithms and functions for image and video processing.

Together, MediaPipe and OpenCV can be used to develop powerful and sophisticated hand gesture recognition systems. For example, a system could be developed to control electronic devices with hand gestures, interact with virtual and augmented reality environments with hand gestures, or translate sign language to spoken language in real time.

While hand gesture recognition using MediaPipe and OpenCV is still under development, it has already made significant progress. There is a growing community of developers who are using MediaPipe and OpenCV to develop new and innovative hand gesture recognition applications.

FUTURE ENHANCEMENT:

1. **Improve accuracy and robustness Using deep learning:** Deep learning algorithms can be used to train models that can learn to recognize hand gestures more accurately and robustly than traditional algorithms. For example, a deep learning model could be trained on a large dataset of hand gesture images and videos to learn to recognize a wide variety of hand gestures, even in challenging conditions.

2. Improve accuracy and robustness using 3D hand tracking: 3D hand tracking algorithms can be used to track the position and orientation of the hand in 3D space. This can improve the accuracy and robustness of hand gesture recognition, especially in challenging conditions such as occlusion.

3. Multi-Hand Gesture Recognition: Extend the system to recognize gestures from multiple hands simultaneously. This is important for scenarios involving interactions between multiple users or complex gestures. Implementing multi- hand gesture recognition can greatly enhance the interactivity of applications involving multiple users, such as virtual reality, gaming, collaborative design, or sign language interpretation. It requires careful design, data collection, and testing to ensure accurate and reliable recognition.

4. Recognize more complex hand gestures using Machine Learning: Machine learning

algorithms can be used to train models that can recognize more complex hand gestures, such as those used in sign language or in virtual reality and augmented reality applications. For example, a machine learning model could be trained on a large dataset of sign language videos to learn to recognize a wide variety of sign language gestures.

5. Real-Time Hand Gesture Recognition Systems: MediaPipe and OpenCV can be used to develop real time hand gesture recognition systems that can be used to control electronic devices, interact with virtual and augmented reality environments, and more. Future research could focus on developing new algorithms and techniques that can improve the performance and efficiency of real-time hand gesture recognition systems.

6. Cross Platform Compatibility: Cross-platform compatibility refers to the ability of software, applications, or technology solutions to work seamlessly and consistently across different operating systems, devices, or platforms. It ensures that users can access and use the software or technology without experiencing significant differences in functionality or user experience, regardless of the specific platform or environment they are using. Cross platform compatibility is essential for developers and organizations that want to reach a broad user base without having to develop and maintain separate versions of their software for each platform. It simplifies development, reduces costs, and provides a consistent user experience, which is crucial for user satisfaction and the success of a product or service.

REFERENCES

[1] Indriani, Moh. Harris, Ali Suryaperdana Agoes : Applying Guide Application Using Media Pipe in Proceedings of the 2nd International Seminar of Science and Applied Technology (ISSAT), Indonesia 2021.

[2] Uzma Akram, Muhammad Hamid, Iram Noreen : Hand Pose Recognition Using Parallel Multi Stream CNN in MDPI, Basel, Switzerland 2021.

[3] Jannatun Meghna, Mohammad Shahria : Hand Gesture-Based Character Recognition Using OpenCV and Deep Learning in International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI) , Rajshahi, Bangladesh 2021

[4] Ahmad Puad Ismail, Mohamat Kasim : Hand Pose Recognition on Python and OpenCV in First International Conference on Electrical Energy and Power Engineering (ICEEPE), Penang, Malaysia 2020

[5] Ashish Sharma, Anmol Mittal, Savitoj Singh, Vasudev Awatramai : Hand Gesture Recognition using Image Processing in International Conference on Smart Sustainable Intelligent Computing and Applications, New Delhi, India 2020

[6] Munir Oudah, Ali Al-Naji, Javaan Chahl : Hand Gesture Recognition Based on Computer Vision in MDPI, Basel, Switzerland 2020.