Hand Gesture Recognition

Shrabanti De, Akash Kumar Singh, Nirmit G Shah

Department of MCA Faculty of IT & Computer Science,

Parul University, Vadodara, India

shrabantide2002@gmail.com

akashsingh777567@gmail.com@gmail.com

nirmitkumar.shah27546@paruluniversity.ac.in

***Abstract*— Sign language is a method for the dumb and deaf society to connect with the exterior world. The dumb and deaf people find it very difficult to express their feelings to the normal people. Thus, the development of a system to fill the communication gap in between mute/deaf and the normal people who are well able to speak and hear would be of great utility to the society. A sign language recognition model involves of accurate and effective tool to translate sign language into text/speech. Gesture recognition identifies significant expressions of a manmade gesture by hands.**

Keywords: **Hand tracking, activity recognition, feature extraction, hardware implementation**

I. Introduction

Gestures originate from any bodily motion or state but commonly originate from the face or hand. In the world of gesture recognition, a gesture is defined as any physical movement, large or small, that can be interpreted by a motion sensor. It may include anything from the pointing of a finger to a roundhouse kick or a nod of the head to a pinch or wave of the hand [16][24]. Three steps make up the entire process of hand gesture recognition. The first step is hand gesture frame acquisition, which uses the webcam to identify the user's hand. The amount of light, the angle, the scale, and the pose make this step difficult. Feature extraction is the secondstep that extracts features of the hand image, such as hand outlines are extracted. The classification process is used in the third and last step to get the result.

One of the challenging tasks in this study is design of the hand gesture recognition system comprises two main issues. The identification of hands comes first. Real-time video capture using a webcam allows for user gesture recognition. The inconsistent brightness, noise, low resolution, and poor contrast would be the issue. To identify the movements, the video's observed hands are identified. Segmentation and edge detection are the steps that the process takes at this point. Making a sign that can be utilized with one hand at a time is another issue. Building the hand gestures and sign languages that will be shown in the system to fulfil the project's name [23]. Feature extraction is a critical component of hand gesture recognition, as it involves identifying and extracting meaningful information from the image or video data captured by sensors or cameras [22]. This process typically involves analyzing various aspects of the hand and fingers, such as shape, size, curvature, and angles, to identify unique features that can be used to distinguish between different gestures. Depending on the specific application, different feature extraction techniques may be employed, such as edge detection, contour analysis, or filtering [22].

Preprocessing steps may also be necessary to enhance the accuracy and reliability of the extracted features. Ultimately, effective feature extraction is essential for enabling machines to interpret and respond to human gestures in a wide range of contexts, from gaming and entertainment to healthcare and industrial automation.

Haar Cascade Classifier is used to detect and extract the features of hand that will be shown in the webcam. Haar Cascade Classifier is an algorithm that detects object in an image regardless of the scale and size of objects [4].

II. Application Areas

A method for a computer to understand human body language is gesture recognition. The need for text interfaces and GUIs (Graphical User Interface) has decreased as a result. Below are a few of the major areas where hand gesture recognition is used.

A. Talking to a computer

Imagine a world in which a person putting together a presentation can add a quote or move an image with a flick of the wrist instead of a click of a mouse. A future in which you can easily interact with the computer with no much touching of mouse.

B. Medical Fields

A surgeon can control the instruments as per the requirements such as controlling the motion of a laparoscope by making appropriate gesture without any touch or voice input.

C. Gesture-based gaming control

In games, computer-vision algorithms must be robust and efficient with no real-time requirement, and where recognition performance is the highest priority. These days, many games are having human interaction through gestures.

D. Hand gesture to control home appliances

Hand gesture-based electronic device control is gaining more importance nowadays. Most electronic devices focus on

the accurate algorithm of hand gesture recognition and the corresponding user interface. Hand Gesture Based Remote is a device to replace all other remotes used in households and perform all their functions. Normally in homes, remotes are used in appliances like TV, CD player, Air Conditioner, DVD Player and Music System. Remotes also used for lights ON/OFF control, Door Opener, etc. One Universal Remote can control all these devices.

E. Gesture control car driving

It is as simple as it sounds: you use your hands to control the functions in your car and no longer have to look away from the road.

F. Gesture control car driving

Gesture used in restaurants for order placing: While you can use proper hand gesture to order any dish. For an example, you're giving a proper hand gesture for number one and the food that has been given number one will be ordered directly with no waiters involved. Human efforts (waiters) are reduced.

G. Security and Surveillance

Hand gesture recognition can aid in security systems by identifying specific hand gestures associated with predefined commands or actions. This can be utilized in access control, authentication, and surveillance applications.

H. Educational and Training

Hand gesture recognition can be employed in educational settings to facilitate interactive learning experiences. It can be used for virtual laboratories, interactive whiteboards, and interactive teaching aids.

I. Public Displays and Presentations

Hand gesture recognition can be employed to interact with public displays, such as interactive kiosks, digital signage, and presentations. Users can navigate menus, select options, or manipulate content through hand gestures[2].

III. Methodologies

A study to detect hand gestures is being done by using the Python programming language and the OpenCV package resides under Python [14]. The system code created by Python is clear and simple to understand. Numpy is another Python package applied here [7]. The image from the webcam will be processed in a region called as the Region of Interest (ROI), which will function as the desired area while ignoring the background. The hand gesture recognition is done in three steps:

A. Gesture frame acquisition

Here, the pixels of an image is changed in a such a way that it is easier to analyse. And this method is known as thresholding [3][21]. In thresholding, we convert an image from colour or grayscale into a binary image, i.e., one that is simply black and white.

Thresholding is a type of image segmentation. Image segmentation involves converting an image into a collection of regions of pixels that are represented by a mask or a labelled image[3]. By dividing an image into segments, you can process only the important segments of the image instead of processing the entire image. A very good segmentation is needed to select an adequate threshold of grey level for extract hand from background for example there is no part of hand should have background and background also should not have any part of hand.

The threshold segmentation process can be regarded as the process of separating foreground from background. Threshold segmentation mainly extracts foreground based on gray value information, so it is especially useful for segmentation of images with strong contrast between foreground objects and background.

K-Means Clustering was tested and found to give good segmentation results for the hand gesture. K-Means clustering is the most popular unsupervised learning algorithm. It is used when we have unlabelled data which is data without defined categories or groups[1].

B. Features Extraction

Features are extracted using an algorithm called as Haar Cascade Classifier. In the 2001 publication, "Rapid Object Detection using a Boosted Cascade of Simple Features," Paul Viola and Michael Jones proposed a effective object recognition method that uses cascade classifiers. A cascade function is trained using a large number of both positive and negative images in this machine learning-based approach. Object detection in other images is then performed using it. Prior to performing any image processing in this project, the hand gesture will be tracked using a machine learning approach. And this is how machine learning comes into the picture [4].

C. Classification and Results

To identify the gesture, the gesture classification is applied. The greater the number of retrieved feature parameters, the more accurate the outcome. In addition to the Finite State Machine (FSM), Learning Vector Quantization, and Principal Component Analysis(PCA) statistical tools used for gesture classification, the Hidden Markov Models(HMM) tool has showed its ability to differentiate between dynamic movements . The neural network has been widely used in the field of hand form extraction and hand gesture recognition.

The results of hand gesture recognition systems vary depending on factors such as the quality and diversity of the dataset, the choice of feature extraction techniques, the complexity of the gestures, and the effectiveness of the machine learning algorithm. High-performing systems can achieve accuracy rates ranging from 90% to over 99%, depending on the specific application and dataset used.

It's worth noting that deep learning approaches, particularly CNNs, have shown remarkable performance in hand gesture recognition tasks, often surpassing traditional machine learning algorithms. They can automatically learn features directly from raw image data, eliminating the need for manual feature engineering and achieving higher accuracy levels in many cases.

IV. Algorithms

There have been several algorithms and techniques used in hand gesture recognition. Here are some of them:

A. Histogram of Oriented Gradients (HOG)

HOG is a feature descriptor used to extract local features of an image. It has been used in hand gesture recognition to extract features from an image of a hand that help identify the gesture. The Histogram of Oriented Gradients (HOG) algorithm is a popular feature extraction method used in computer vision for object detection and recognition tasks, including hand gesture recognition.

B. Scale-Invariant Feature Transform (SIFT)

SIFT is another feature descriptor that is used in hand gesture recognition. It is designed to detect and describe local features in an image that remain invariant to scale, orientation, and affine distortion. The Scale Invariant Feature Transform (SIFT) algorithm is another popular feature extraction method used in computer vision, including hand gesture recognition.

The SIFT algorithm works by identifying keypoints in the image that are invariant to scaling, rotation, and translation. For hand gesture recognition, these keypoints could correspond to important points in the hand, such as the fingertips or the palm center.

Once the keypoints are identified, the SIFT algorithm calculates the local features around each keypoint, which are robust to variations in illumination and contrast. These local features are described using a histogram of gradient orientations and magnitudes, similar to the HOG algorithm.

C. K-nearest neighbour (K-NN)

K-NN is a classification algorithm that has been used in hand gesture recognition papers. It calculates the similarity between the input gesture and the stored training data and assigns the input gesture to the class that has the highest similarity. The k-Nearest Neighbor (k-NN) algorithm is a popular machine learning algorithm that can be used for hand gesture recognition. The k-NN algorithm is a non-parametric method, which means that it does not assume any specific model of the underlying data distribution.

In k-NN, the idea is to classify an input hand gesture image by finding the k training examples that are closest to it in feature space, based on a distance metric such as Euclidean distance. The class label of the input image is then assigned based on the majority class among the k nearest neighbors. To use the k-NN algorithm for hand gesture recognition, the hand images need to be pre-processed and represented as feature vectors. This can be done using techniques such as the Histogram of Oriented Gradients (HOG) or the Scale Invariant Feature Transform (SIFT), as discussed earlier.

Once the feature vectors are obtained, they can be used to train the k-NN classifier. During inference, the k-NN algorithm is applied to the feature vector of the input hand gesture image, and the k nearest neighbors are found. The class label of the input image is then assigned based on the majority class among the k nearest neighbours. One of the advantages of the k-NN algorithm is that it is simple and easy to implement. However, it can be computationally expensive for large training datasets, and the choice of k and the distance metric can have a significant impact on the classification performance. Additionally, the k-NN algorithm may not perform well in cases where the feature space is high-dimensional and the curse of dimensionality becomes a problem.

D. Decision Tree

Decision trees have been used in hand gesture recognition to create a model that can make predictions about the class of a given gesture. The decision tree is created based on the features extracted from the training data set. The decision tree algorithm is a popular machine learning algorithm that can be used for hand gesture recognition. A decision tree is a tree-like model that consists of nodes and branches, where each internal node represents a decision based on one of the input features, and each leaf node represents a class label.

E. Random Forest

Random Forest is an ensemble learning technique that has been used in hand gesture recognition. It combines multiple decision trees to improve classification accuracy [18]. Random Forest is a popular ensemble learning algorithm that can be used for hand gesture recognition. Ensemble learning algorithms combine multiple machine learning models to improve the overall accuracy and robustness of the classifier [20].

F. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of deep learning model that have been particularly successful for hand gesture recognition [19]. CNNs consist of multiple layers of convolutional and pooling operations, which can learn highly abstract features from the input. image. The output of the convolutional layers is typically fed into fully connected layers, which can perform classification based on the learned features [20].

V. Tools & Technologies

There are several tools and technologies used in Hand Gesture Recognition:

A. OpenCV

OpenCV (Open Source Computer Vision Library) is a huge open-source library for computer vision, machine learning, and image processing. OpenCV supports a wide variety of programming languages like Python, C++, Java, etc [6]. It can process images and videos to identify objects, hand gesture, or even the handwriting of a human[15]. When it is integrated with various libraries, such as Numpy which is a highly optimized library for numerical operations, then the number of weapons increases in your Arsenal i.e whatever operations one can do in Numpy can be combined with OpenCV [7]. OpenCV is written in the programming language C++, as is its primary interface, but it still retains a less comprehensive though extensive older C interface [14]. The application programming interface (API) for these interfaces can be found in the online documentation. Wrapper libraries in several languages have been developed to

encourage adoption by a wider audience [23][4].

B. TensorFlow

TensorFlow is an open-source software library for dataflow and differentiable programming across a range of tasks. TensorFlow is an end-to-end open source platform for machine learning [8][11]. It has a comprehensive, flexible ecosystem of tools, libraries, and community resources that lets researchers push the state-of-the-art in ML, and gives developers the ability to easily build and deploy ML-powered applications. It has pre-trained models that can be used for training and testing hand gesture recognition algorithms. TensorFlow was developed by the Google Brain team for internal Google use in research and production. The initial version was released under the Apache License 2.0 in 2015. Google released the updated version of TensorFlow, named TensorFlow 2.0, in September 2019 [8][11][12][13].

C. Keras

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It is useful for rapid prototyping, advanced research, and production.[10][12] Keras has pre-built models that can be fine-tuned for hand gesture recognition tasks. Keras is used to label the data from the dataset directories, to augment the data by shifting, zooming, rotating and mirroring. This augmentations(modification) on the image, help to increase the number of training data and assure that the data are not biased to a particular handedness [5].

D. Intel RealSense

Intel RealSense is another technology that can be used for hand gesture recognition. It uses a combination of 3D cameras, sensors, and software to detect and track hand movements. It can also recognize hand gestures and can be used in applications such as healthcare, gaming, and robotics. Intel RealSense Technology is a product range of depth and tracking technologies designed to give machines and devices depth perception capabilities. The technologies, owned by Intel are used in autonomous drones, robots, AR/VR, smart home devices amongst many others broad market products [9] [25].

VI. Latest R&D in the field

A. Deep Learning Techniques

Researchers are exploring the use of deep learning techniques such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for improving the accuracy of hand gesture recognition. These techniques are known to provide better accuracy than traditional machine learning approaches[17].

B. Convolutional Neural Networks (CNNs)

CNNs are widely used in image recognition tasks, including hand gesture recognition. CNNs consist of multiple convolutional layers that extract local features from input images, followed by fully connected layers for classification. They can learn hierarchical representations of hand gestures, capturing both low-level and high-level features [20].

C. Recurrent Neural Networks (RNNs)

RNNs are designed to handle sequential data, making them suitable for capturing temporal dependencies in hand gesture recognition. By processing the sequence of hand gesture frames over time, RNNs can learn the dynamics and temporal evolution of gestures. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are popular RNN variants[17].

D. 3D Convolutional Neural Networks (3D CNNs)

Unlike traditional 2D CNNs, 3D CNNs can directly process spatiotemporal data, such as video sequences of hand gestures. They extend the concept of 2D convolution to the temporal dimension, allowing the model to capture both spatial and temporal information simultaneously[17].

E. Spatial Transformer Networks (STNs)

STNs enable the network to learn and apply spatial transformations to input images, such as rotations, translations, and scaling. STNs can be incorporated into the architecture to handle hand gesture variations, such as different hand orientations or positions, by allowing the model to learn invariant representations.

VII. Conclusions

It may be concluded that hand gesture recognition is an important technological advancement for human-computer interaction. A hand gesture detection approach is provided that consists of segmenting the palm and fingers, detecting and identifying fingers, removing the hand area from the background, and applying a rule classifier to predict the labels of hand motions. The approach outperforms a state-of-the-art method on an additional set of hand gesture data, demonstrating excellent performance and great efficiency. identification of hand gestures is very useful in various applications, including augmented reality, robot control, sign language interpreters for the impaired, and sign language identification.

The way we interact with our electronics has been completely transformed by hand gesture recognition technology, which provides a more natural and intuitive user interface. It improves the efficiency and smoothness of interactions by doing away with the requirement for physical buttons or touchscreens. There are a wide range of possible uses for this technology, including gaming, entertainment, healthcare, and accessibility. There are obstacles in the way of its general acceptance, though. Obstacles can be greatly increased by variables including changing illumination, user tiredness, and the requirement for precise training data. The future of hand gesture recognition is still promising in spite of these obstacles. More advanced and integrated applications of this technology are being made possible by the ongoing developments in artificial intelligence and machine learning. As hand gesture recognition advances, we may expect it to become a crucial aspect of our everyday existence, improving our entire user experience. The importance and usefulness of this technology in our digital age is further demonstrated by its potential to revolutionize a number of sectors and provide accessibility for those with impairments.

REFERENCES

[1]. https://www.kaggle.com/code/prashant111/k-means-clustering-with-python

[2]. https://how2electronics.com/gesture-recognition-application-machine-learning/

[3]. https://datacarpentry.org/image-processing/07-thresholding/#:~:text=Thresholding%20is%20a%20type%20of,is%20simply%20black%20an d%20white.

[4].https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier

[5]. https://abenezer-g.medium.com/keras-convolutional-neural-network-cnn-implementation-for-hand-gesture-recognition-d7dd11958af6

[6]. https://sourceforge.net/projects/opencvlibrary/

[7]. https://www.geeksforgeeks.org/opencv-python-tutorial/

[8]. https://en.wikipedia.org/wiki/TensorFlow

[9]. https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html

[10]. https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-is-keras#:~:text=Keras%20is%20used%20for%20creating,%2C%20Yelp%2C%20Uber%2C%20etc.

[11]. https://www.tensorflow.org/datasets/catalog/wikipedia

[12]. https://sandipanweb.wordpress.com/2018/01/20/hand-gesture-classification-using-deep-convolution-and-residual-neural-network-with-tensorflow-keras-in-python/

[13]. https://techvidvan.com/tutorials/hand-gesture-recognition-tensorflow-opencv/

[14]. https://opencv.org/

[15]. https://www.tutorialspoint.com/opencv/index.htm

[16].https://www.frontiersin.org/articles/10.3389/fnins.2020.00637/full

[17]. https://en.wikipedia.org/wiki/Deep_learning

[18]. Leo Breiman (2001), 'Random Forests'

[19]. Peng Liu, Xiangxiang Li, Haiting Cui, Shanshan Li, and Yafei Yuan (2019), 'Hand Gesture Recognition Based on Single-Shot Multibox Detector Deep Learning'

[20]. Adam Ahmed Qaid MOHAMMED, Jiancheng Lv,* and MD. Sajjatul Islam (2019), 'A Deep Learning-Based End-to-End Composite System for Hand Detection and Gesture Recognition'

[21].Aashni Haria, Archanasri Subramanian, Nivedhitha Asokkumar, Shristi Poddar, Jyothi S Nayak (2017), 'Hand Gesture Recognition for Human Computer Interaction', Procedia Computer Science Volume 115, 2017, Pages 367-374

[22]. Rafiqul Zaman Khan and Noor Ibraheem (2012), 'Hand Gesture Recognition: A Literature Review'

[23]. Abbosjon Kudratov (2020), Hand Gesture Recognition Implementation using Python, OpenCV

[24]. Bhupesh Kumar Dewangan and Princy Mishr (2022), 'Hand Gesture Recognition Through Depth Sensor'

[25]. Hitesh Kumar Sharma and Tanupriya Choudhury (2022), 'Hand Gesture Recognition: The Road Ahead'

[26]. Farah Farhana Mod Ma'asum[1], Suhana Sulaiman[1] and Azilah Saparon (2015), 'An Overview of Hand Gestures Recognition System Techniques'