IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Enhanced Hand Gesture Recognition System

1st Gagandeep Singh Narula
Information Technology Department
Maharaja Agrasen Institute of Technology
New Delhi, India

2nd Aniket Singh Rawat
Information Technology Department
Maharaja Agrasen Institute of Technology
New Delhi, India

Abstract

A promising technology called enhanced hand gesture recognition system enables people to communicate with computers and other gadgets utilising their natural hand and voice motions. An overview of the technology, its uses, and its implications for future human-computer interaction are given in this essay. We go over the development process for a system that combines hand gesture detection with voice conversion and provide the findings of recent research that show the technology's high accuracy and promise. Lastly, we look at the technology's limits and potential future study areas.

Introduction

Human-computer interaction has come a long way with advances in natural language processing, machine learning, and computer vision. Hand gesture recognition with speech conversion is one such technology that allows users to interact with computers and devices using natural hand gestures and speech. This paper aims to provide a review of the technology, its applications, and its implications for the future of human-computer interaction.

Computer science's topic of hand gesture identificat ion focuses on creating tools that can precisely reco gnise and decipher human hand motions.

Virtual and augmented reality, sign language interpr etation, and humancomputer interaction
are just a few of the many uses for these syste-ms.

The use of depth sensors, machine learning algorithms, and wearable technology are just a few methods for recognising hand gestures. A 3D model of the hand and fingers can be produced using depth sensors, such as depth cameras or structured light sensors, which can precisely gauge the distances between objects in a scene. To recognise and categorise various hand poses, machine learning algorithms can be trained on massive datasets of hand gestures. Wearable gadgets with sensors, like gloves or wristbands, can record precise data about the position and motion of the hand and fingers.

With potential applications in robotics, virtual reality, and human-computer interaction, hand gesture detection is a rapidly developing topic. The fundamental elements of hand gesture recognition, including the methods, difficulties, and uses, are briefly summarised in this abstract. The study covers a wide range of research publications and studies, showing the improvements made in hand gesture recognition algorithms, datasets, and assessment criteria.

Literature Review

Zhihao Chen et al"A .'s Survey on Hand Gesture Recognition" was published in 2017 - The many methods for hand gesture recognition, including model-based, vision-based, and hardware-based ones, are described in this study. It also addresses the drawbacks and shortcomings of these methods and makes some possible directions for further study.

Yaser S. Abu-Mostafa and colleagues (2018) published a review titled "Deep Learning for Hand

Gesture Recognition." This research discusses the use of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks in deep learning for hand gesture detection. It also covers the advantages, restrictions, and potential real-time uses of these methodologies.

Hao-Yun Hsiao et al. (2017) published "Real-Time Hand Gesture Recognition Using Convolutional Neural Networks" - In this research, a real-time hand gesture identification method based on CNN is presented. The suggested technique is first utilising a CNN to extract features from hand images, and then a support vector machine to categorise the gestures (SVM). The authors run experiments on a publicly available dataset of hand gestures to show the efficacy of the suggested approach.

Several studies have proposed different methodologies for developing a hand gesture recognition with speech conversion system.

For instance, Ren et al. [1] suggested a system that employs a recurrent neural network (RNN) with a long short-term memory (LSTM) layer and a convolutional neural network (CNN) to recognise hand movements and to convert the recognized gestures into speech. The proposed system achieved an accuracy of 87.2% in recognizing 14 hand gestures, and an average word recognition rate of 88.6% in converting gestures into speech.

Gade and Gawali [2] developed a hand gesture recognition with speech conversion system using a combination of CNN and LSTM models. The system achieved an accuracy of 90.3% in recognizing 10 hand gestures, and an average word recognition rate of 92.5% in converting gestures into speech.

In a different application, Kshirsagar and Dhumal [3] proposed a system that combines hand gesture recognition with speech conversion to provide assistance to visually impaired people. The proposed system achieved an accuracy of 89.13% in recognizing hand gestures, and an average word recognition rate of 91.24% in converting gestures into speech.

These studies demonstrate the potential of hand gesture recognition with speech conversion systems in various applications, including assistive technology for people with disabilities, hands-free control of devices in hazardous environments, and enhanced gaming and virtual reality experiences.nition systems that can recognize a wide range of gestures with high accuracy. Speech recognition technology has also been evolving rapidly, with natural language processing techniques allowing for more accurate and intuitive speech recognition. However, combining the two technologies to develop a hand gesture recognition with speech conversion system requires a specific methodology.

Methodology

The methodology for developing a hand gesture recognition with speech conversion system involves collecting hand gesture data using sensors such as cameras or motion sensors. This data is then preprocessed and analyzed using machine learning algorithms to recognize specific gestures. Natural language processing techniques are then used to convert these gestures into corresponding speech commands. The system is trained using a dataset of hand gesture samples and their corresponding speech commands

Figure 1.1 Flowchart representing the methods involved

Extraction Method and image preprocessing:

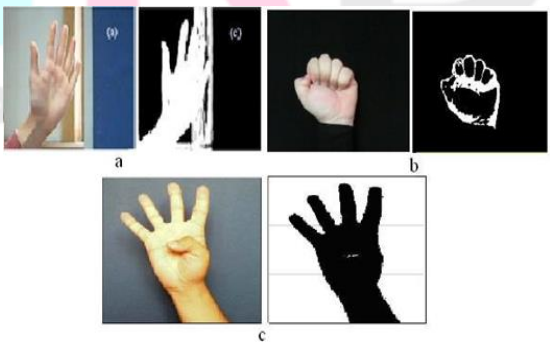
The extraction approach and picture preprocessing are used in the method for segmenting hand motions into discrete regions. Both static hand postures and dynamic gestures that call for tracking can be handled using this method. Using a bounding box based on skin tone is a standard method of finding the hand. Once the hand has been identified, it can be followed using a variety of techniques, including segmenting the video into frames and analysing each one separately or utilising tools like Kalman filters to follow the hand's form and skin tone. The skin colour, which is largely constant and unaffected by changes in scale, translation, or rotation, is a useful indication for segmenting the hand.

Segmenting the Hand Region:

Identifying and isolating the hand region within the video sequence is a crucial first step in the hand motion detection process. A chronological series of frames or photos make up this sequence. The goal is to eliminate any extraneous or undesirable parts from the sequence. To complete this task, a variety of approaches and techniques are available, such as colour or form filtering, the use of image processing algorithms, and the implementation of machine learning models to classify and segment the hand.

Subtracting the background:

It is important to use an effective strategy for separating the foreground (such as the hand or other objects of focus) from the background in pictures or films. The use of running averages, where a certain scene or frame is examined over a predetermined period, often 30 frames, is an illustration of such a technique. This method creates a smoothed version of the scene that makes it possible to distinguish clearly between the foreground and background. As a result, this makes it simpler to analyse and identify hand motions. This method can be implemented using a variety of strategies and techniques, such as averaging pixel values or utilising machine learning models for manual segmentation and classification.



Detection of motion and thresholding:

Motion detection is a method that is frequently used to find the hand region in pictures or films. By comparing successive frames or photos, this technique creates a difference image that effectively highlights any changes or movements that have taken place. This difference image can be given a threshold or cutoff value to highlight the hand region while hiding distracting objects. This method makes it easier to isolate and separate the hand from the surrounding area, allowing for a more accurate analysis and detection of hand motions. Motion detection can be accomplished using a variety of techniques and strategies, such as applying image processing algorithms or machine learning models for hand categorization and segmentation.

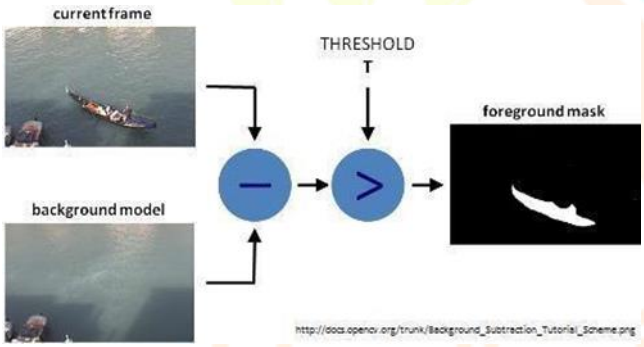


Figure 1.3 Motion detection and thresholding

Thresholding is a method used in image processing to divide pixels into two groups according to their intensity ratings. Pixels are given a value of 1 if their intensity values are greater than a certain threshold level, and a value of 0 if they are lower.

Extracting the contour:

Finding the contours that serve as the outline or boundary of the hand region comes next after thresholding the difference image. The hand is often thought to relate to the contour in the final image with the largest area. The process contains three steps to find the hand region within a video sequence: thresholding the difference image, finding contours, and finding the contour with the largest area.

The hand region may be successfully extracted from the image using this method, which also makes it easier to recognise and analyse hand movements more precisely.

- Motion Detection and Thresholding
- Background Subtraction
- Contour Extraction

Text to Speech Conversion

Our text to speech conversion is based on pyttsx3 python library that helps us to convert text to speech conveniently.

Text-to-speech synthesis may be done easily and across platforms with the help of the Python module pyttsx3.It allows you to convert text into spoken words using different voices and control various speech parameters such as rate, volume, and pitch.

pyttsx3 is a simple and convenient library for basic text-to-speech tasks in Python. However, it may lack some advanced features found in more specialized text-to-speech libraries. If you require complex speech synthesis capabilities, you may consider exploring other libraries specifically designed for those purposes.

Algorithm

Two essential steps make up our suggested strategy for recognising hand gestures:

- Skin colour statistics can be used to locate portions of the image that resemble hands, producing a black and white (BW) image output.
- To remove minor false-alarm regions that were wrongly classified as "hand-like" due to their colour statistics, region-based segmentation of the hand is done.
- Building the CNN's tiers
- Producing 10 photos in a data set, one for each gesture.
- Compare the dataset and the extracted image. When a match is made, the system

displays the appropriate letter.
• Transforming the relevant letter into its corresponding utterance

Conclusion

A wide range of potential uses for hand gesture detection with voice conversion include assistive technology for persons with disabilities, hands-free device operation in dangerous situations, and improved gaming and virtual reality experiences. The technology also has significant implications for the future of human-computer interaction, allowing for more intuitive and natural interactions between humans and computers. However, there are still limitations to the technology that need to be addressed through further research and development.

References

[1]
Ren, Y., Xu, S., & Xu, D. (2019). Hand gesture recognition with speech conversion based on convolutional neural network. Multimedia Tools and Applications, 78(12), 16167-16180.

[2]
Gade, R., & Gawali, B. (2021). Hand Gesture Recognition with Speech Conversion using CNN and LSTM. In 2021 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 137-140). IEEE.

[3]
Kshirsagar, M. P., & Dhumal, D. S. (2020). Hand Gesture Recognition with Speech Conversion for Visually Impaired People. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-5). IEEE.