Hand Gesture Recognition Using TensorFlow and OpenCV

A Project Report

Submitted by

Name of the Candidates:

Mansi Shingate - 201081031
Shruti Tambe - 201081027
Prachi Katela - 201081023
Tanavi Bote - 201081028

For the Academic year 2022-2023
Bachelors of Technology
(Information Technology)

Veermata Jijabai Technological Institute

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our Respected Professor and Course Instructor for the Digital Image Processing Lab Mr. Sunil G Bhirud for providing us with his guidance and knowledge to complete this Project onHand Gesture Recognition Using TensorFlow and OpenCV . We would like to thank him for rendering all possible help and support during the development, implementation and Presentation of the Project.

We, the students of 6th Semester, would like to thank our teaching assistants for helping us and guiding us throughout the process of implementing the project and making the project more efficient and successful.

Table of Contents

ABSTRACT

Sign language is a form of communication used by individuals with speaking and hearing impairments. To facilitate effective communication, there is a need for real-time translation software that can accurately convert hand gestures into spoken languages like English. This proposed system utilizes Python, NumPy, OpenCV, labeling,Mediapipe and TensorFlow to create the software.The software processes images or videos captured from a camera device using convolutional neural networks (CNN). The CNN model is trained with a large dataset, either sourced from open repositories or created specifically for sign language gestures. By analyzing recognition rates and prediction outcomes, the software classifies the provided image or video as a specific alphabet or number from the American Sign Language Set. By implementing this software, individuals with disabilities can easily understand the sign language used by others, enabling better communication and inclusivity.

INTRODUCTION

Sign language serves as a crucial medium for communication for individuals with hearing impairments. However, the challenge arises when conveying sign language to non-sign language users, hindering effective communication and integration. Moreover, individuals with speech disabilities face difficulties expressing their emotions and ideas in a world that heavily relies on spoken language.

To address these issues, a sign language recognition system is proposed. Hand gestures play a vital role in sign language as a form of nonverbal communication. This system aims to enable people with hearing impairments to convey their thoughts and allow non-sign language users to understand them. It leverages deep learning techniques to achieve accurate gesture recognition, considering the variations and richness of sign languages across different countries.

There are two primary approaches for sign language recognition: glove-based systems and vision-based systems. Glove-based systems involve wearing data gloves to capture hand movements accurately. On the other hand, vision-based systems can be categorized into static and dynamic recognition. Static recognition focuses on representing gestures in two dimensions, while dynamic recognition captures real-time gestures.

In this project, static identification of hand gestures is utilized due to its improved accuracy compared to dynamic recognition methods. This approach avoids the need for wearing uncomfortable data gloves and enables real-time recognition without relying on additional hardware. By implementing this system, individuals with hearing or speech impairments can effectively communicate their ideas, emotions, and opinions, bridging the communication gap in a world primarily designed for spoken language users.

LITERATURE SURVEY

The sign language recognition system is designed to interpret hand gestures and enable effective communication between sign language users and non-users. It consists of four main steps: image enhancement and segmentation, orientation detection, feature extraction, and classification.

In the image enhancement and segmentation step, techniques like contrast adjustment and thresholding are applied to improve image quality and separate the hand region from the background. Orientation detection focuses on determining the angle and rotation of the hand gesture, using algorithms like the Hough Transform.

Feature extraction techniques extract relevant information, such as shape, texture, and motion characteristics, from the segmented hand region. These features are then utilized for gesture classification. Machine learning algorithms like support vector machines or deep learning models can be trained on a large dataset to classify and recognize different sign language gestures accurately.

Despite its potential, the sign language recognition system faces limitations. Rapid changes in lighting conditions pose a challenge as it can cause errors or failures in hand region detection. Another limitation is the lack of temporal analysis, which could capture the motion of gestures and improve accuracy. The system may struggle with complex backgrounds or the presence of other objects alongside hand gestures.

Researchers have explored different approaches, including glove-based systems and vision-based systems using cameras or sensors. Each approach has its advantages and limitations in terms of accuracy, comfort, and usability. Data acquisition can be done through various input devices, such as data gloves, markers, or webcam images.

Improvements can be made by incorporating more efficient algorithms like neural networks for edge detection and segmentation. Expanding the dataset to include a wider range of sign gestures would also enhance the system's capabilities.

In conclusion, the sign language recognition system aims to bridge communication gaps by accurately interpreting hand gestures. It involves image enhancement, segmentation, orientation detection, feature extraction, and classification. Overcoming limitations and refining the system through algorithm advancements and dataset expansion will contribute to its effectiveness in facilitating communication for individuals with hearing impairments.

METHODOLOGY

Image Acquisition

A dataset of hand gesture images was collected, ensuring that the images captured the hand region with the gesture clearly visible. A camera or webcam was used to capture real-time video frames for hand gesture recognition. Each frame was read from the camera or webcam using OpenCV's VideoCapture class.

Segmentation

Each frame was converted from the BGR color space to the RGB color space. The Mediapipe library was used to process the RGB frame and obtain hand landmarks. If hand landmarks were detected:

The coordinates of each landmark were extracted.

The landmark coordinates were normalized to a consistent range, typically between 0 and 1. The normalized landmark coordinates were stored.

Classification

The pre-trained hand gesture recognition model was loaded using TensorFlow or Keras. The normalized hand landmarks were passed as input to the model for gesture prediction.

The predicted gesture class probabilities were obtained from the model's output. The class with the highest probability was determined as the predicted gesture class.

The predicted gesture class was mapped to its corresponding class name using the loaded class names.

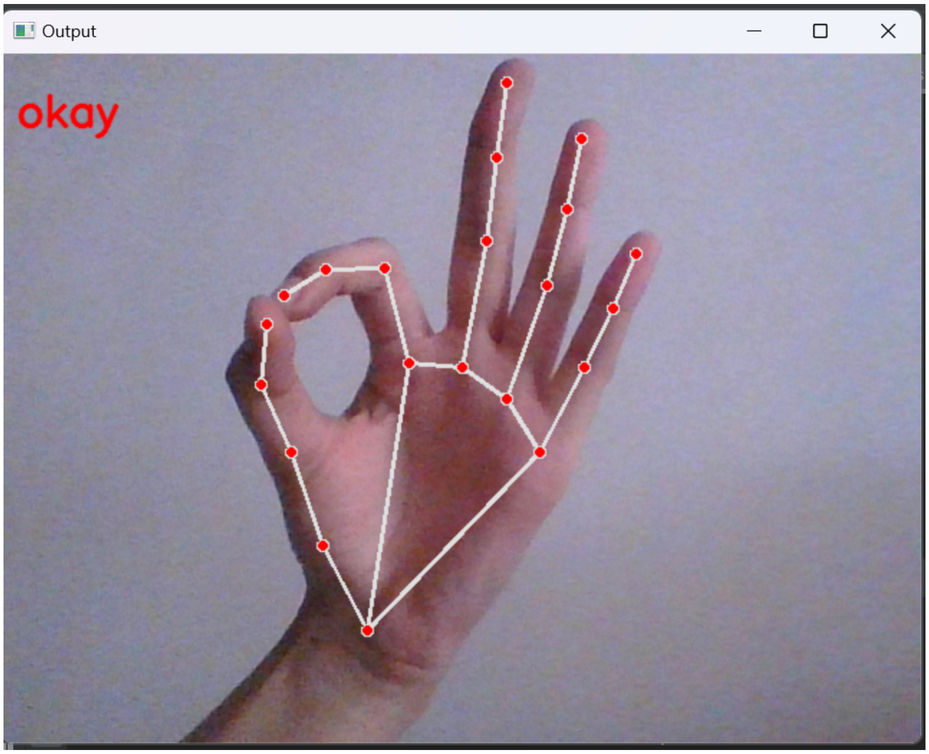
The extracted mask images are trained by a convolution neural network Lenet in KERAS, which is a high-level API written in Python and capable of running on top of the tensor flow. A convolution neural network (CNN, or ConvNet) could be a category of deep neural networks in deep learning, most typically used in the visual representation process analysis, CNNs use a multilayer perceptron variation designed to require the smallest amount of preprocessing. After optimization techniques, the neural network is constructed using the Tensor Flow-based Keras system with the following hyperparameter values: For an individual with limited knowledge of deep learning, this can be daunting. KERAS offers an easy and flexible Network Training API that hides most of the complicated information underneath the hood. After the training dataset, the output predicts the gesture and shows the accuracy of the image.

WORKING ANALYSIS

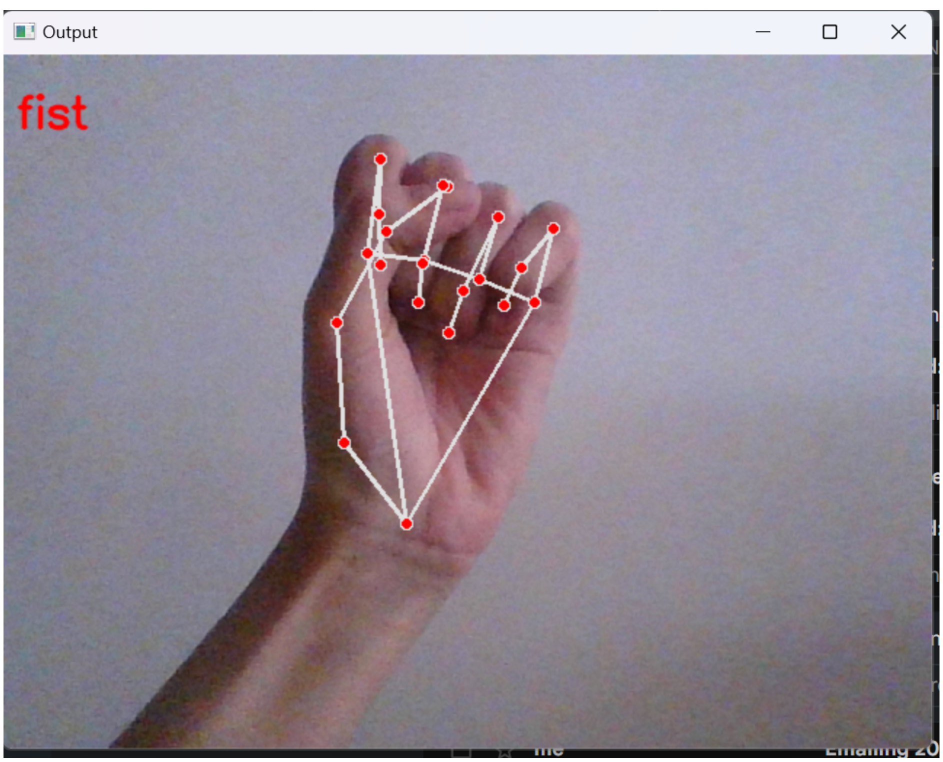
The hand gesture recognition project developed a real-time system that could recognize and classify hand gestures. It involved collecting a dataset of hand gesture images and using a camera or webcam to capture video frames. Dataset taken for this project is large with 10 classes each with over 500 images. The 10 classes differ in the gestures made using a hand. For eg. gestures like calling, peace sign, thumbs up, close fist(rock), etc. are present in the dataset. OpenCV's VideoCapture class was used to read the frames. The project is built using the KERAS library and the model created is sequential. A backbone is first created on which we have built the model. The image dataset is divided into training and testing sets to avoid errors and overfitting. For further processing and classification, the images are converted to arrays. Dense layers are used to build the model. The frames were then processed using the Mediapipe library to extract hand landmarks. The pre-trained hand gesture recognition model, loaded using TensorFlow or Keras, was used to classify the gestures based on the normalized landmark coordinates. The predicted gesture class was displayed on the frames, providing real-time feedback. The system's performance was evaluated based on accuracy and other metrics. Overall, the project successfully implemented a hand gesture recognition system, highlighting its potential and identifying areas for further improvement.

After running the model we get following outputs:

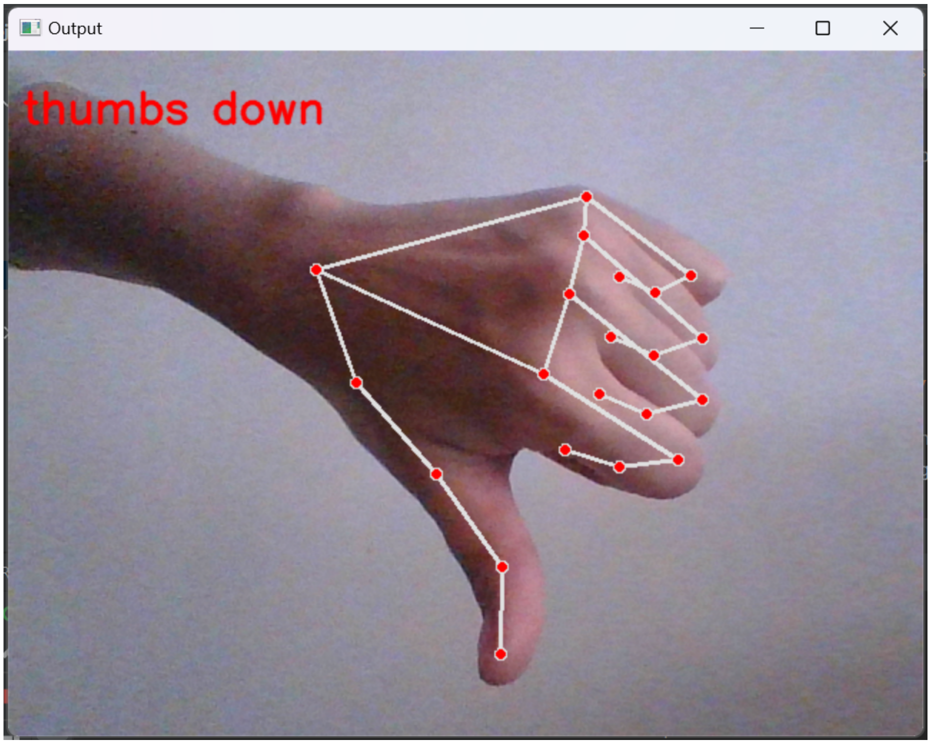
1) For detection of 'okay' gesture:



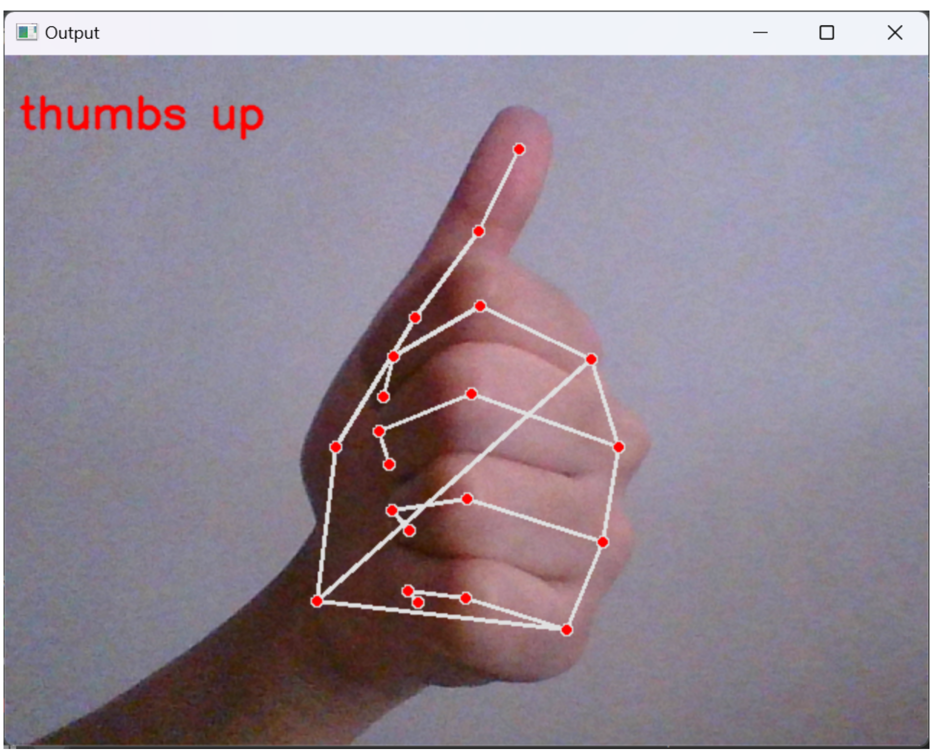
2) For detecting a 'fist' gesture:



3) For detecting a 'thumbs down' gesture:



4) Detection of 'thumbs up' gesture:



CONCLUSION

Our project successfully developed a real-time system for recognizing and classifying hand gestures. By collecting a dataset of hand gesture images and utilizing computer vision techniques and deep learning models, the system was able to accurately identify and classify different hand gestures in real-time video frames. The project demonstrated the effectiveness of the chosen approach and highlighted the potential for practical applications of hand gesture recognition technology. Further improvements could be made to enhance the system's performance and address challenges related to lighting conditions and background interference. Overall, the project contributes to the advancement of hand gesture recognition and its potential for various fields such as human-computer interaction, sign language interpretation, and augmented reality.

REFERENCES

https://www.kaggle.com/datasets/roobansappani/hand-gesture-recognition

https://medium.com/@cmmapada/hand-gesture-recognition-5cdc0e380854

Das, A., Gawde, S., Suratwala, K., Kalbande, D. "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images". 2018 International Conference on Smart City and Emerging Technology (ICSCET), (2018).

Rao, G. A., Syamala, K., Kishore, P. V. V., Sastry, A. S. C. S. "Deep convolutional neural networks for sign language recognition",(2018).

Mahesh Kumar NB. "Conversion of Sign Language into Text". International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9, (2018).