JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Real-Time Hand Gesture Recognition Using TensorFlow and OpenCV

Pratik Thakur, Rupali Borade, Shivani Jadhav, Ronak Gaikwad, Dr. Jayashree Pansare, Shubhangi Khade

Department of Computer Engineering
MES College Of Engineering, Pune

Abstract: Sign language is a way or means of communication used by individuals with speaking and hearing impairments. It is one of the essential means of communication for such individuals to stay connected with the rest of the world and to express their ideas, needs or beliefs. There is a great need for an efficient and cost-effective real-time translation software or tool in the modern-day world to understand what the disabled individual is trying to express with accuracy. The proposed system is a real-time translation software or tool used for the conversion of hand gestures into natural languages such as English used by people for communication. The translated data will interpret the alphabet or number associated with the sign shown on the live camera feed. The software proposed in this project is created using Python, NumPy, OpenCV, labeling, and TensorFlow. The image or video obtained from the camera device will be processed using convolutional neural networks (CNN). The CNN model is pre-trained with a large dataset from open sources or using a custom dataset on sign language gestures. Based on the recognition rate and prediction analysis from the CNN model, the provided image or video will be classified as the respective Alphabet or number from the American Sign Language Set. This helps the individuals to understand the sign language used by disabled individuals with ease.

Keywords: Convolution Neural Network, OpenCV, Computer Vision, Deep Learning, TensorFlow, Sign Gesture.

INTRODUCTION

Language is a medium that human being uses to communicate, express themselves, and understand notions of the real world. Without it, no books, no cell phones, and no words would have any meaning. People take it for granted and don't realize how important it is, because it is so deeply ingrained into their daily lives. Unfortunately, in today's fast-paced world, people with hearing problems are often forgotten and ignored. Also, the major problem faced by persons who are unable to speak is that they cannot express their emotions easily in this world. They must struggle to bring up their ideas, voice out their opinions and express themselves to people who are different from them. Sign language, although being a medium of communication to deaf people, still has no meaning when conveyed to a non-sign language user. Also, they are not able to utilize the applications of Artificial Intelligence (personal Butlers) like google assistance, Apple's SIRI, etc., because all those apps depend on voice commands. To prevent such discomfort, this sign language recognition system is introduced. Hand gestures are one of the nonverbal communication strategies used in sign language. It is mainly used by deaf and dumb people who have hearing or speaking disorders to communicate among themselves or with other people who know sign language. It would be a fantastic medium for people with hearing impairments to share their thoughts, as well as a great way for non-sign language users to understand what the latter is saying. Many countries have their standard of sign gestures and interpretations. For example, an alphabet in Korean sign language is not the same as an alphabet in Indian sign language. Although this emphasizes the richness of sign languages, it also emphasizes their ambiguity. Deep learning must be well-versed in gestures to achieve a reasonable level of accuracy. Either of the two techniques may be used to identify a sign. The first is a glove-based system, in which the singer wears a pair of data gloves while the hand movements are registered. The second approach is focused on vision, which is further divided into static and dynamic recognition. Static is concerned with the representation of gestures in two dimensions, while dynamic is concerned with the capturing of gestures in real-time. Even though gloves have a high accuracy rate of over 90%, they are uncomfortable to wear and cannot be used in rainy conditions. They are not easily carried around since their use needs a computer. This project uses static identification of hand gestures in this case because it improves accuracy as compared to dynamic

hand gestures, such as those for the letters J and Z. The proposed study here tries to increase accuracy by combining Convolution Neural Network (CNN) and open Computer Vision.

LITERATURE REVIEW

The implementation is divided into four main steps: 1. Image Enhancement and Segmentation 2. Orientation Detection 3. Feature Extraction 4. Classification. This work is focused on the above four categories, but the main limitation is a change of color is happening very rapidly by the change in the different lighting conditions, which may cause errors or even failures. For example, due to insufficient light conditions, the existence of the hand area is not detected but the non-skin regions are mistaken for the hand area because of the same color. The system is based on Hand gesture recognition by modeling the hand in the spatial domain. The system uses various 2D and 3D geometric and non-geometric models for modeling. It has used the Fuzzy c- Means clustering algorithm which resulted in an accuracy of 85.83%. The main drawback of the system is it does not consider gesture recognition of temporal space, i.e., the motion of gestures and it is unable to classify images with complex backgrounds i.e., where there are other objects in the scene with the hand objects. This survey focuses on hand gesture recognition using different steps like data acquisition, pre-processing, segmentation, and so on. A suitable input device should be selected for the data acquisition. There are several input devices for data acquisition. Some of them are data gloves, markers, and hand images (from webcam/Kinect 3D Sensor). But the limitation of this work changed in the illumination, rotation and orientation, scaling problem, and special hardware which is considerably costlier.

The system implementation is divided into three phases:

1. Hand gesture recognition using the kinetic camera

2. Algorithms for hand detection recognition

3. Hand gesture recognition.

The limitation here is that the edge detection and segmentation algorithms used here are not very efficient when compared to neural networks. The dataset being considered here is very small and can be used to detect very few sign gestures. The limitations of this system involve

1. The numbers of gestures that are recognized are less

2. The gestures recognized were not used to control any applications.

METHODOLOGY

Image Acquisition

The first phase was to grab the frame from the camera and establish the region of interest in the image. The challenge is to utilize an adequate computer vision method that can detect the slightest discrepancy between similar signals to prevent inaccurate translation in real-time. There are over 150 conversion methods in OpenCV for color-space. But only two of them will be examined here, BGR to Gray and BGR to HSV. In this paper, we convert the colored RGB format into HSV, as it is more convenient to extract the gesture rather than others

Segmentation

Segmentation is also a vital part of the proposed methodology. The process shows a binary object describing the segmentation in the simplistic application. Black pixels are the background pixels and white pixels are the foreground pixels. In simple implementations, a single parameter called the intensity threshold determines the segmentation. Numerous imperfections can occur in binary images.

Classification

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 perception 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.

CONCLUSION

Our project objective is to bridge the gap by introducing an economical computer application in the communication path so that the sign language can be automatically captured, recognized, and translated to text for the benefit of deaf people. The image obtained must be analyzed, processed, and converted to either sign or textual display on the screen for the benefit of the hearing impaired. We have learned and demonstrated that CNN can learn how to identify and predict the text. We have created a model that pre-processes the image to the required nature for it to be fed into the model. The system is an approach to ease the difficulty in communicating with those having speech disabilities. The amount of training and validation loss observed with the proposed CNN architecture was less. We have tried different training and validation loss observed with the proposed CNN architecture was less. We have tried different image processing techniques to find the best one we need for our use. During the live capture testing, the Canny edge algorithm with an accuracy of 98% demonstrated better results than the other techniques.

REFERENCES

[1] 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).

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

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

[4] Pei Xu, Department of Electrical and Computer Engineering, University of Minnesota, A Real-time Hand Gesture Recognition and Human-Computer Interaction System, Research Paper April 2017.

[5] Lewis, M. Paul; Simons, Gary F.; Fennig, Charles D., eds. (2013), "Deaf sign language", Ethnologue: Languages of the World (17th ed.), SIL International, retrieved 2013-12-03.

[6] P.S Rajam and G. Balakrishnan, "Real-time Indian Sign Language Recognition System to aid deaf- dumb people" in IEEE 13th International Conference on Communication Technology.