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

A Review of "Real-Time Hand Gesture Recognition using Deep Learning"

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

Department of Computer Engineering
MES College Of Engineering, Pune

Abstract: The advancements in technology over the past few decades had made possible things that were once deemed impossible. Computer technology has been at the forefront of this revolution. The changes are seen not only in what the technology can do but also in such marvelous ways, that humans can use those technologies. Latest human-computer interface technologies like touch screen operations, voice command recognition, and such have redefined the way we interact with machines. With every step forward in interface technology, we walk a step closer to seamless operations with computers. One such medium of interaction is gesture recognition. While it is at the bleeding edge of research, it is yet a long way from being perfect. This paper aims to study the latest innovations on the subject and to provide insight into the improvements that can be made to the existing theories.

Keywords: Image Recognition, Computer Vision, OpenCV, Machine Learning, Deep Learning, CNN, TensorFlow, Keras, Python.

INTRODUCTION

A gesture is a physical movement of a body (in the scope of this paper, a hand) to express a certain meaning. Gesture recognition is a type of image recognition technique. It can be static or dynamic, depending on the problem being solved. This technique is in several applications in computer technology. The most influential application is sign language translation. Gesture recognition is a frontrunner among the technologies that are making a wide range of information accessible to impaired individuals.

To implement this technique, OpenCV is used widely. OpenCV is a framework that is used to capture, process, and translate information from images. It is implemented using the Python language in this study. Here is a Machine Learning technique by which the ordinary person can grasp what the differently able one tries to convey. There are several similar object recognition systems and algorithms for monitoring. This enables the recognition of gestures to solve the restrictions and limitations prevalent in the previous approaches.

LITERATURE REVIEW

Reference	Year	Accuracy (%)
Malavika Suresh et al.	2019	96
Abhishek B. et al.	2020	-
Snehal Bera et al.		84
Teena Verma et al.	2020	98
Shweta Upadhyay et al.	2018	89.72, 94.87

There have been numerous studies on the subject of Gesture Recognition using Deep Learning. The paper titled "Real-Time Hand Gesture Recognition Using Deep Learning" focuses on one such study by Malavika Suresh et al. The study titled "Real-Time Hand Gesture Recognition Using Deep Learning" does a great job of highlighting the latest advances in applications of this technology, at the same time diving into the depth of the actual implementation.

Along with the aforementioned paper, we will also take a look at "Hand gesture recognition using machine learning algorithms" by Abhishek B et al., and "Real-Time Gesture Detection System Using OpenCV and TensorFlow" by Snehal Bera et al. To compare the results and findings with similar implementations.

Image Processing

Image processing is a method to perform some operations on an image, to get an enhanced image, and or extract some useful information from it. If we talk about the basic definition of image processing then "Image processing is the analysis and manipulation of a digitized image, especially to improve its quality".

An image may be defined as a two-dimensional function f(x, y), where x and y are spatial(plane) coordinates, and the amplitude of fat in any pair of coordinates (x, y) is called the intensity or grey level of the image at that point. In other words, an image is nothing more than a two-dimensional matrix (3-D in the case of colored images) which is defined by the mathematical function f(x, y) at any point is giving the pixel value at that point of an image, the pixel value describes how bright that pixel is, and what color it should be.

Image processing is signal processing in which input is an image and output is an image or characteristics according to the requirement associated with that image.

OpenCV

Computer vision is a process by which we can understand the images and videos how they are stored and how we can manipulate and retrieve data from them. Computer Vision is the base or mostly used for Artificial Intelligence. Computer-Vision is playing a major role in self-driving cars, robotics as well as in photo correction apps.

OpenCV is the huge open-source library for computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even the handwriting of a human. When integrated with various libraries, such as NumPy, python is capable of processing the OpenCV array structure for analysis. To Identify image patterns and their various features we use vector space and perform mathematical operations on these features.

The first OpenCV version was 1.0. OpenCV is released under a BSD license and hence it's free for both academic and commercial use. It has C++, C, Python, and Java interfaces and supports Windows, Linux, Mac OS, iOS, and Android. When OpenCV was designed the focus was real-time applications for computational efficiency. All things are written in optimized C/C++ to take advantage of multi-core processing.

Applications of OpenCV

- Image/video I/O, processing, display (core, improve, highgui)
- Object/feature detection (object, features2d, nonfree)
- Geometry-based monocular or stereo computer vision (calib3d, stitching, video stab)
- Computational photography (photo, video, superheroes)
- Machine learning & clustering (ml, Flann)
- CUDA acceleration (GPU)

CNN

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets can learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlaps to cover the entire visual area. A ConvNet can successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image

dataset due to the reduction in the number of parameters involved and the reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.

The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. ConvNets need not be limited to only one Convolutional Layer. Conventionally, the first ConvLayer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network that has a wholesome understanding of images in the dataset, similar to how we would.

Hand Gesture Recognition

Hand gestures are the foundation of sign language, which is a visual form of communication. In this paper, a deep learning-based convolutional neural network (CNN) model is specifically designed for the recognition of gesture-based sign language. This model has a compact representation that achieves better classification accuracy with a fewer number of model parameters over the other existing architectures of CNN.

The requirements for hand detection involve the input image from the webcam. The image should be fetched with a speed of 20 frames per second. Distance should also be maintained between the hand and the camera. The approximate distance that should be between hand the camera is around 30 to 100 cm. The video input is stored frame by frame into a matrix after pre-processing.

CONCLUSION

From the literature survey and comparative study, we learn the architecture of the hand gesture recognition models and the building blocks of the same. Researchers have used various deep learning algorithms for improving image recognition technology. Deep Learning is at the forefront of this innovation.

There is further scope of research in this study by considering creating a real-time hand gesture recognition model by using the aforementioned technologies.

ACKNOWLEDGMENT

We are grateful to the Head of Department Dr. N. F. Shaikh, Dr. J. R. Pansare, and the entire Computer Engineering Department of MES College Of Engineering, Pune for their constant guidance and support.

REFERENCES

[1] Malavika Suresh et al., Real-Time Hand Gesture Recognition Using Deep Learning, in IJIIE 2019 December Edition Vol. 1

[2] M. Panwar, ―Hand gesture recognition based on shape parameters, in Proceedings of the International Conference on Computing, Communication and Applications (ICCCA '12), pp. 1–6, February 2012.

[3] Kim, Jonghwa & Mastnik, Stephan & Andre, Elisabeth. (2008). EMG-based hand gesture recognition for real-time biosignal interfacing.

[4] Murthy, G.R.S. & Jadon, A review of vision-based hand gesture recognition. International Journal of Information Technology and Knowledge Management

[5] Arpita Ray Sarkar, G. Sanyal, S. Majumder. "Hand Gesture Recognition Systems: A Survey." International Journal of Computer Applications, vol. 71, no.15, pp. 25-37, May 2013.

[6] Ruchi Manish Gurav, Premanand K. Kadbe, Real time finger tracking and contour detection for gesture recognition using OpenCV, IEEE Conference May 2015, Pune India.

[7] Manjinder Kaur, "Hand Posture Detection: Principle, Need and Algorithms", International Journal of Core Engineering and Management, July 2015

[8] Bantupalli, K, Xie, Y. (2018). "American Sign Language Recognition using Deep Learning and Computer Vision". IEEE International Conference on Big Data (Big Data), (2018).

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