DYNAMIC HAND GESTURE RECOGNITION

A PROJECT REPORT

Submitted by

Thevaprakash P (19MIM10003)
Krishna Basak (19MIM10050)
Shinjan Verma (19MIM10061)
Kesavan R (19MIM10086)

in partial fulfillment for the award of the degree of

MASTERS OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Specialization in

Artificial Intelligence and Machine Learning

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

VIT BHOPAL UNIVERSITY

KOTHRIKALAN, SEHORE
MADHYA PRADESH - 466114

OCT 2020

VIT BHOPAL UNIVERSITY,KOTHRIKALAN, SEHORE MADHYA PRADESH – 466114

BONAFIDE CERTIFICATE

Certified that this project report titled **"DYNAMIC HAND GESTURE RECOGNITION"** is the bonafide work of **Thevaprakash P (19MIM10003), Krishna Basak (19MIM10050), Shinjan Verma (19MIM10061)** and **Kesavan R (19MIM10086),** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR
Dr. V Pandimurugan, Asst. Professor
School of AI &ML division
VIT BHOPAL UNIVERSITY

PROJECT GUIDE
Dr. V Pandimurugan, Asst. Professor
School of AI &ML division
VIT BHOPAL UNIVERSITY

The Project Exhibition-I Examination is held on _____.

ACKNOWLEDGEMENT

First and foremost, I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr Manas Kumar Mishra, Dean, School of Computing Science and engineering and Dr Nageswara Guptha, Head, AI & ML Division, for much of their valuable support encouragement in carrying out this work.

I would like to thank my internal guide Dr. V Pandimurugan, for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Computing Science and Engineering, who extended directly or indirectly all support.

Last, but not the least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

LIST OF ABBREVIATIONS

NN – Neural Network

TF – Tensorflow

CNN – Convolutional Neural Network

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

So early human ancestors probably used gestures to Communicate. Once Armstrong said that he thinks gestures involving the hands may well have been the earliest form of complex human communication. As Armstrong said, the beginning stage of human computer Interactions is gesture recognition system and also it is a natural way of communicating with the computers. In this paper, we have designed a Dynamic Hand Gesture Recognition System using Neural Network, which can recognize the gesture using the Computer's or Laptop's Web Camera and do the according tasks.

Table of Contents

INTRODUCTION

Chapter – 1

1.1 Introduction

Gesture Recognition is an important and active field of Computer Vision. Gestures can be Static or Dynamic (Motion), but direct use of hand can be more attractive and natural way of conveying the meaning to the humans and computer. But it also depends on the speed of motion and latency.

Gesture recognition has wide-ranging applications in human-computer interaction, sign-language communication, video surveillance, dance/video annotations and forensic identification.

1.2 Motivation for the work

This project is motivated from the problem that prevailed long ago. The difficulties faced by deaf and mute people are also a motivation for this project. It also eliminates the direct use of hands in public places especially during a pandemic situation like this. So, to develop a system which can detect the gestures and these gestures are widely used for conveying the information or to control the devices. This is the motivation for the work.

1.3 About Introduction to the project including techniques

The introduction to the techniques that we use in this project are as follows. We will use a database of images that represent the gestures and make a Convolutional Neural Network model, then train the model to classify the images based on the gestures from the database. The next step is to use the trained model to predict the gestures from real time live video feed. Finally, we assign some tasks to perform based on the predicted gestures.

1.5 Problem Statement

There are many applications where hand gesture can be used for interaction such as videogames. These hand gestures can also be used by handicapped people to interact with the systems. Classical interactions tools like keyboard, mouse, touchscreen, etc. may limit the way we use the system. All these systems require physical contact, in order to interact with system. Gestures can interpret same functionality without physically interacting with the interfacing devices. The problem lies in understanding these gestures, as for different people, the same gesture may look different for performing the same task. This problem may be solved by the use of Deep Learning approaches. Convolution neural networks (CNN's) are proving to be ultimate tool to process such recognition systems. High computing power is required in order to process gestures with higher accuracy and efficiency.

1.6 Objective of the work

Hand gesture recognition system can be used for interfacing between computer and human. The objective of this project is to develop a model for recognition of hand gestures with reasonable accuracy and also to make the system to perform specific tasks based on the predicted gestures.

1.7 Summary

The detailed summary introduces us to the basic knowledge about hand gesture recognition. It also throws light on the importance of this recognition system and the motivation behind the work. The techniques that were used are also explained along with the problem statement. The final part of introduction contains the main objective of the work.

LITERATURE REVIEW

Chapter – 2

2.1 Introduction

From the existing research paper we found that human computer interaction has a more more important active field, one of the major field of computer vision in the interpretation of human gestures. Hand gestures would be a way of exchanging information with other people in a virtual space, guiding some bots to perform certain tasks in an environment, or interaction with the computers. In that we have two different mode of recognitions like one is the vision-based model but this depends on the environment and another is sensor based with needed background things like lights and isolated environment.

2.2 Core Idea of the Project

Most of the hand gesture recognition systems are uses hardware or some of the equipment for getting gestures but here we have proposed to work the same functionality in our computer systems with the use of laptop or computer's webcam.

2.3 Existing System

The systems that exists now uses some kind of hardware device (other that camera) to get the reading of hand movement as well as the gesture. These hardware devices are mostly some kind of sensors [5][14][15][16][17] or any other kind of Arduino device [6] with some alternative variation of sensors included in it. Thus, the interaction between human and computer becomes less “natural” because of the use of such devices. Also, these devices are not cheap, aesthetic and easy to maintain. For these reasons, it becomes problematic to common

people to access and operate them with ease. Moreover, the devices are to be worn on the persons hand, causing mass spread of micro molecules/microorganisms.

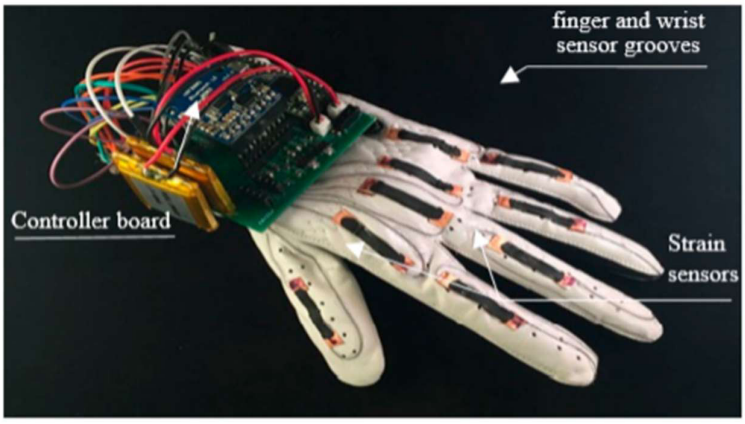

Figure 1 – Sensor-based gesture recognition gloves

2.4 Summary

Gesture recognition is a mathematical interpretation of a human motion to a computer, the vision-based approach deals with the video capturing, image processing and pattern recognition. For this we have to use the CNN classifier to determine the shape of the hand. In the vision-based approach they have avoided the skin color segmentation. The aim is to recognize dynamic hand gestures with maintaining the high accuracy and the time taken to recognized with the use of laptop or computer's webcam.

SYSTEM ANALYSIS

Chapter - 3

3.1 Introduction

Our Hand Gesture recognition system consist Convolutional Neural Network has used to Predicting the Gesture our CNN was created using the Keras which is interface of Tensorflow to detecting the gesture and control the operating system according to the assigned task and also for training the CNN.

3.2 Disadvantages/Limitation in the Existing System

Existing methods for only numbers and some basic gestures so we can't do it for all, meaning tend to be change always we have used these gestures to interact with the computer.

For example, showing your palm to increase the volume, existing methods for only numbers and some basic gestures so we can't do it for all - meaning tend to be change always. We have use these gestures to interact with the computer or they were using the hardware like kinect or leap motion which costs around approximately 100 USD which is much costlier in our country and also electrical equipment are there but that makes more noise so we have come up with this idea.

3.3 Proposed System

We can segment our proposed work in three units/modules –

CNN Model Generation: Here, we will use a database of images and make a CNN model, train it to classify the images and any other same kind of image.

User Feed Taking: Now, as we have made the CNN model, we will use this model to predict the gesture from real time live video feed.

Doing the Task: As the prediction goes on, the system will do the task assigned to each gesture when they are predicted.

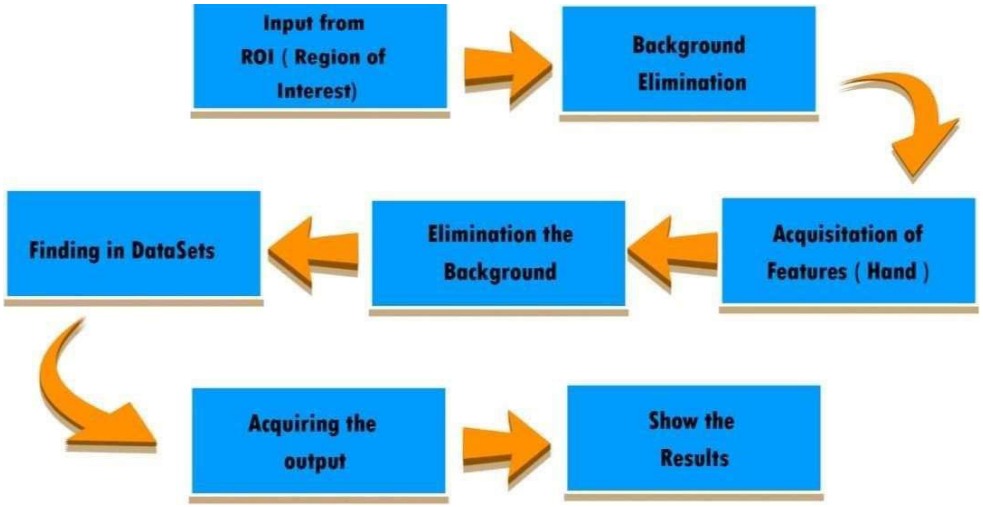

Figure 2 – Process of Hand Gesture Recognition

3.4 Summary

Most of the gesture recognition projects consist of some kind of sensors. Those sensors are used for the prediction of Motion, but this project does not use any of those sensors for prediction. Rather it just takes input from camera and Do the Respective Jobs.

SYSTEM DESIGN AND IMPLEMENTATION

Chapter - 4

1.1 Introduction

We will see the systematic design of the work along with the detailed implementation of various modules which are the building blocks of this project. There are three modules in total and are explained below.

4.2 Module 1

Module 1 is about building a CNN model which we use to predict the gestures. For that we use Keras (interface for TensorFlow) to build and train the model using local database of images. This forms the basis of our work.

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_4 (Conv2D)            (None, 98, 118, 32)       320
_________________________________________________________________
batch_normalization_4 (Batch (None, 98, 118, 32)       128
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 49, 59, 32)        0
_________________________________________________________________
dropout_6 (Dropout)          (None, 49, 59, 32)        0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 47, 57, 64)        18496
_________________________________________________________________
batch_normalization_5 (Batch (None, 47, 57, 64)        256
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 23, 28, 64)        0
_________________________________________________________________
dropout_7 (Dropout)          (None, 23, 28, 64)        0
_________________________________________________________________
flatten_2 (Flatten)          (None, 41216)             0
_________________________________________________________________
dense_4 (Dense)              (None, 128)               5275776
_________________________________________________________________
dropout_8 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_5 (Dense)              (None, 6)                 774
=================================================================
Total params: 5,295,750
Trainable params: 5,295,558
Non-trainable params: 192
```

Figure 3 – Keras Model

4.3 Module 2

Our second module is about predicting the gesture through live video feed. The web camera is used to take real time input from the user. From that input we predict the gestures by the users

by using the trained model. We used OpenCV for the purpose of image and video processing. NumPy is used to make images as array for further computations and lastly TensorFlow keras to predict the gesture.

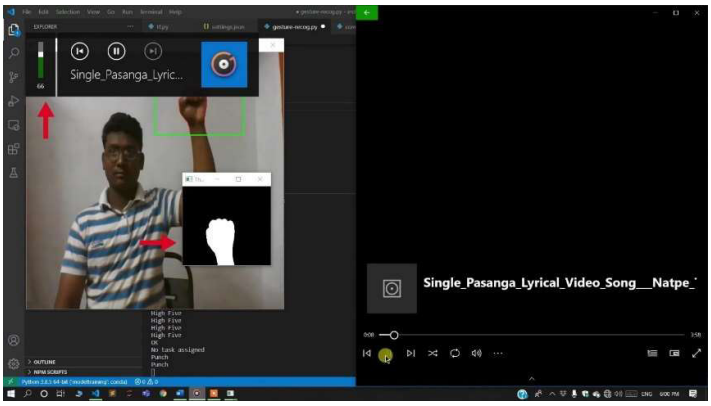

Figure 4 – Get live prediction

4.4 Module 3

This is the final module of the work. This module is about assigning a specified task for a particular predicted gesture. It actually controls the devices using the gestures. So, we use PyAutoGUI for controlling devices like mouse and keyboard. Finally, the system is made to do some desired tasks in response to the gestures.

4.5 Summary

To summarize, the whole work is divided into three modules, each for a specific purpose. The first part is to train the model for prediction for which we use keras. The next part deals with getting the user input at real time. We use OpenCV for that. NumPy is also used for computational purposes. And finally, to make the system to perform a specific task, we use PyAutoGUI and by doing so, we are able to control the devices.

PERFORMANCE ANALYSIS

Chapter - 5

5.1 Introduction

Here we are going to analysis how our model was performing in the stream of hand gesture Recognition. Our model is using CNN and using Tensorflow for model training as giving the accuracy of 99% approximately.

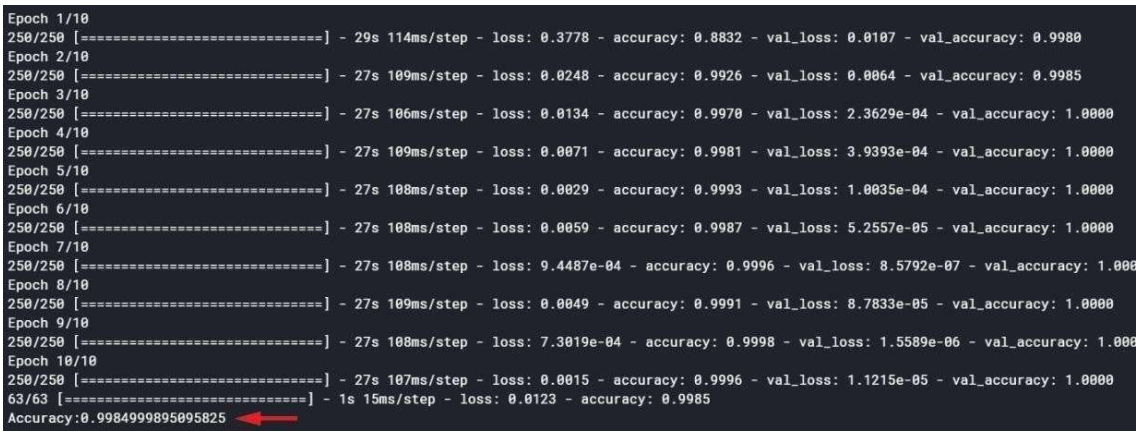

Figure 5 – Training and Testing Accuracy

5.2 Performance Analysis

Here we are providing the chart that how our model was performing while training Process.

Epochs	loss	accuracy	val loss	val accuracy
Epoch 1	0.3778	0.8832	0.0107	0.998
Epoch 2	0.0248	0.9926	0.0064	0.9985
Epoch 3	0.0134	0.997	2.37E-04	1
Epoch 4	0.0071	0.9981	3.94E-04	1
Epoch 5	0.0029	0.9993	1.00E-04	1
Epoch 6	0.0059	0.9987	5.26E-04	1
Epoch 7	9.45E-04	0.9996	8.58E-04	1
Epoch 8	0.0049	0.9991	8.74E-04	1
Epoch 9	7.30E-04	0.9998	1.56E-04	1
Epoch 10	0.0015	0.9996	1.12E-04	1

Table 1 – Value of epochs

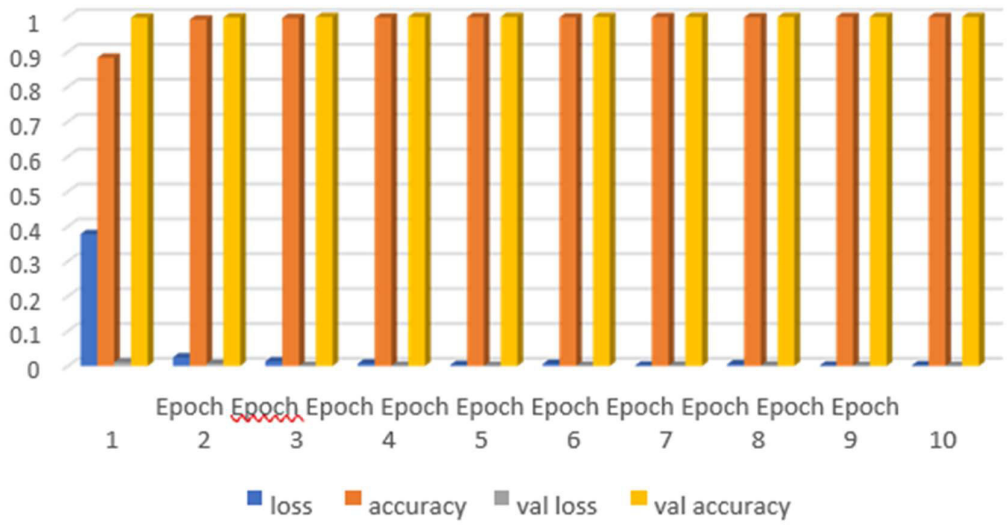

Figure 6 – Performance Analysis

5.3 Summary

This is how our model was performing here we have included the Graph, Table and Charts, our model has reached the accuracy as 0.9984999895095825 that is approximately 99% you can see the results.

FUTURE ENHANCEMENT AND CONCLUSION

Chapter - 6

6.1 Introduction

The following part discuss about the enhancements that can be made to the system in the future. Hand gesture recognition is inevitable in this modern age of communication. Here we propose some of the limitations that exist and a new way of overcoming those limitations

6.2 Limitation/Constraints of the System

It is very difficult to build a system for hand gesture with higher efficiency. Because it involves lot of computations. In many applications, hands only occupy about 10% of the image, whereas the most of it contains background, human face, and human body. So, we need to perform image segmentation and background elimination at real time. This is a constraint of the system.

6.3 Future Enhancements

In order to eliminate the limitation of the system, we can train the model with images thatcontain some noise in the background. In other words, we can train the model with images with the background. The training dataset should be a combination of images with and without background. By this way we can build a robust model that performs less computations.

6.4 Conclusion

To conclude, we can eliminate the existing problem in the system by adding some images to the dataset that contain different background. Thereby a robust model is built with a reasonable accuracy and involving less computations.

REFERENCES

[1] Pavlo Molchanov, Xiaodong Yang, Shalini Gupta, Kihwan Kim, Stephen Tyree, Jan Kautz; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4207-4215

[2] Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition Xiaoyan Wang,1 Ming Xia,1 Huiwen Cai,2 Yong Gao,3 and Carlo Cattani

[3] D. B. Nguyen, S. Enokida, and E. Toshiaki, “Real-time hand tracking and gesture recognition system,” in Proceedings of the International Conference on Graphics, Vision and Image Processing (IGVIP '05), pp. 362–368, CICC, 2005.

[4] Wu J., Peng B., Huang Z., Xie J. (2013) Research on Computer Vision-Based Object Detection and Classification. In: Li D., Chen Y. (eds) Computer and Computing Technologies in Agriculture VI. CCTA 2012. IFIP Advances in Information and Communication Technology, vol 392. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36124-1_23

[5] K. Yamagishi, L. Jing and Z. Cheng, "A system for controlling personal computers by hand gestures using a wireless sensor device," 2014 IEEE International Symposium on Independent Computing (ISIC), Orlando, FL, 2014, pp. 1-7, doi: 10.1109/INDCOMP.2014.7011759.

[6] Vuyyuru, Gopi Manoj. (2020). Hand Gesture Gaming using Ultrasonic Sensors Arduino. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS060668.

[7] Rui Wang, Wei Li, Runnan Qin and JinZhong Wu “Blur Image Classification based on Deep Learning”, IEEE, ISBN 978-1-5386-1621-5 pp. 1-6, 2017

[8] Laila Ma‟rifatul Azizah, Sitti Fadillah Umayah, Slamet Riyadi, Cahya Damarjati, Nafi Ananda Utama “Deep Learning Implementation using Convolutional Neural Network in Mangosteen Surface Defect Detection”, ICCSCE, ISBN 978-1-5386-3898-9, pp. 242-246, 2017

[9] Sebastian Stabinger, Antonio Rodr´ıguez-S´anchez “Evaluation of Deep Learning on an Abstract Image Classification Dataset”, IEEE International Conference on Computer Vision Workshops (ICCVW), ISBN 978-1-5386-1035-0, pp. 2767-2772, 2017

[10] Hasbi Ash Shiddieqy, Farkhad Ihsan Hariadi, Trio Adiono “Implementation of Deep-Learning based Image Classification on Single Board Computer”, International Symposium on Electronics and Smart Devices (ISESD), ISBN 978-1-5386-2779-2, pp. 133-137, 2017

[11] Bansal, Himanshu & Khan, Rizwan. (2018). A Review Paper on Human Computer Interaction. International Journal of Advanced Research in Computer Science and Software Engineering. 8. 53. 10.23956/ijarcsse.v8i4.630.

[12] Yang Shu and Zheng Yang, "Research on human-computer-interaction design of software based on barrier-free principle," 2008 9th International Conference on Computer-Aided Industrial Design and Conceptual Design, Kunming, 2008, pp. 435-440, doi: 10.1109/CAIDCD.2008.4730605.

[13] N. Jmour, S. Zayen and A. Abdelkrim, "Convolutional neural networks for image classification," 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), Hammamet, 2018, pp. 397-402, doi: 10.1109/ASET.2018.8379889.

[14] Sidhant Gupta, Dan Morris, Shwetak N Patel, Desney Tan, SoundWave: Using the Doppler Effect to Sense Gestures, ACM, 2012

[15] Kaustubh Kalgaonkar, Bhiksha Raj, One-handed gesture recognition using ultrasonic sonar, IEEE, 2009

[16] S. P. Tarzia, R. P. Dick, P. A. Dinda, and G. Memik, “Sonar-based measurement of user presence and attention,” in Proceedings of the 11th international conference on Ubiquitous computing. ACM, 2009

[17] H. Watanabe, T. Terada, and M. Tsukamoto, “Ultrasound-based movement sensing, gesture-, and context-recognition,” in Proceedings of the 2013 International Symposium on Wearable Computers. ACM, 2013