



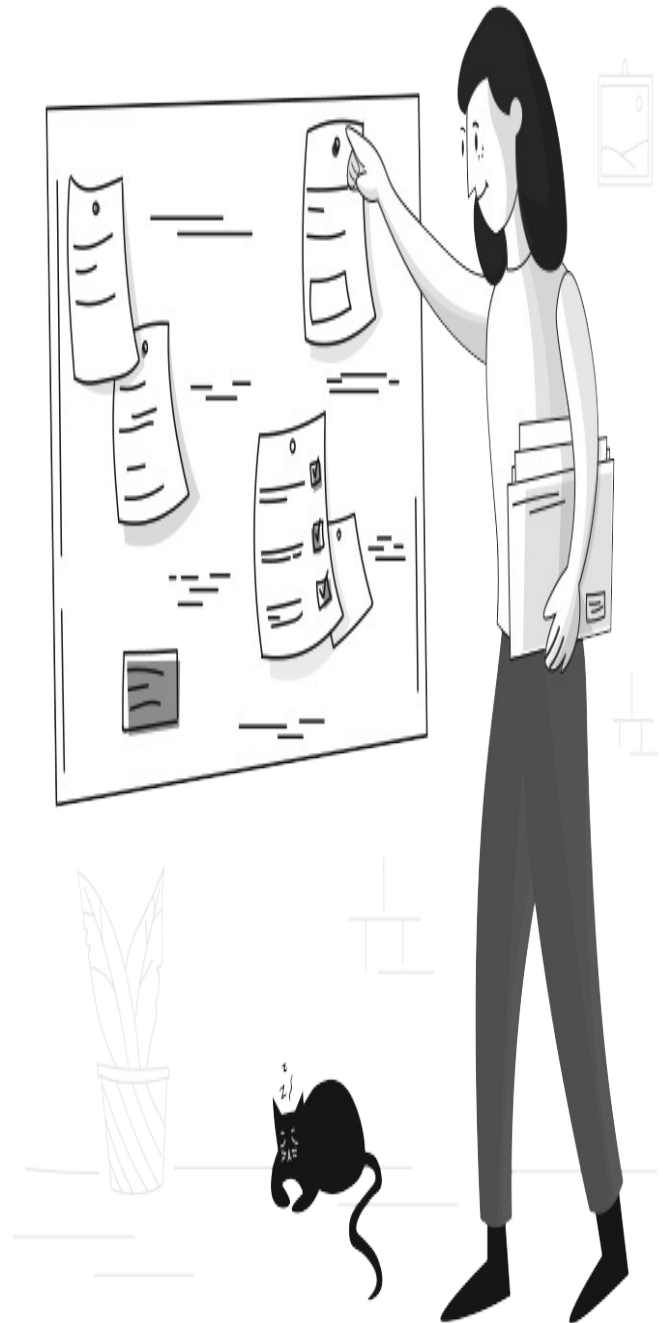
Hand Gesture Recognition

Divyansh Rajput
1901064

IITG

Problem Statement

Creating a Machine Learning model using Python and OpenCV which is able to recognise hand gestures based on the training provided by near infrared images of hand gestures and documenting the results



Database Description

Database used to train and test this model is composed by 10 different hand-gestures (showed above) that were performed by 10 different subjects (5 men and 5 women)

Gestures included in the database are :

- 1) Palm
- 2) Palm moved(side)
- 3) Fist
- 4) Fist moved(side)
- 5) Thumb
- 6) Index finger
- 7) OK gesture
- 8) Letter L
- 9) Letter C
- 10) Hold / Down

Also the database contains 200 near infrared images per gesture per subject acquired by the Leap Motion sensor



Litrature Survey

INTRODUCTION

Gesture recognition has been a very interesting problem in Computer Vision community for a long time. This is particularly due to the fact that segmentation of foreground object from a cluttered background is a challenging problem in real-time. The most obvious reason is because of the semantic gap involved when a human looks at an image and a computer looking at the same image. Humans can easily figure out what's in an image but for a computer, images are just 3-dimensional matrices. It is because of this, computer vision problems remains a challenge.

Methodology and Tools / Algorithms used

OpenCV

OpenCV is the huge open-source library for the 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 handwriting of a human. When it integrated with various libraries, such as NumPy, python is capable of processing the OpenCV array structure for analysis. To Identify image pattern and its various features we use vector space and perform mathematical operations on these features.

Logistic regression

Logistic regression is a statistical analysis method used to predict a data value based on prior observations of a data set. Logistic regression has become an important tool in the discipline of machine learning. The approach allows an algorithm being used in a machine learning application to classify incoming data based on historical data. As more relevant data comes in, the algorithm should get better at predicting classifications within data sets.

Decision Tree

A decision tree is a supervised learning technique that has a pre-defined target variable and is most often used in classification problems. This tree can be applied to either categorical or continuous input & output variables. The training process resembles a flow chart, with each internal (non-leaf) node a test of an attribute, each branch is the outcome of that test, and each leaf (terminal) node contains a class label. The uppermost node in the tree is called the root node.

Random Forest

Random Forest is a robust machine learning algorithm that can be used for a variety of tasks including regression and classification. It is an ensemble method, meaning that a random forest model is made up of a large number of small decision trees, called estimators, which each produce their own predictions. The random forest model combines the predictions of the estimators to produce a more accurate prediction.

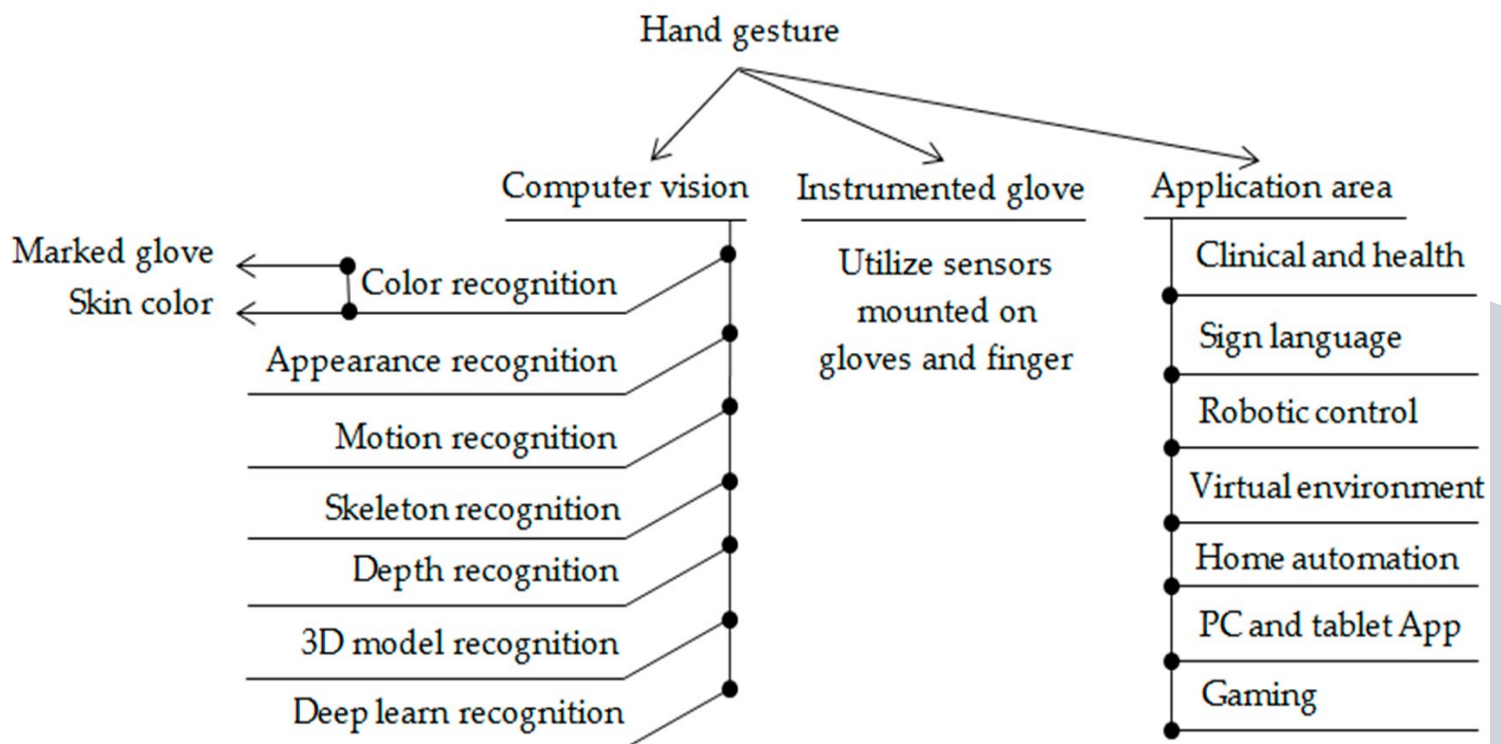
Artificial Neural Networks (ANN)

An artificial neural network is an attempt to simulate the network of neurons that make up a human brain so that the computer will be able to learn things and make decisions in a humanlike manner. ANNs are created by programming regular computers to behave as though they are interconnected brain cells.

Multilayer Perceptron (MLP)

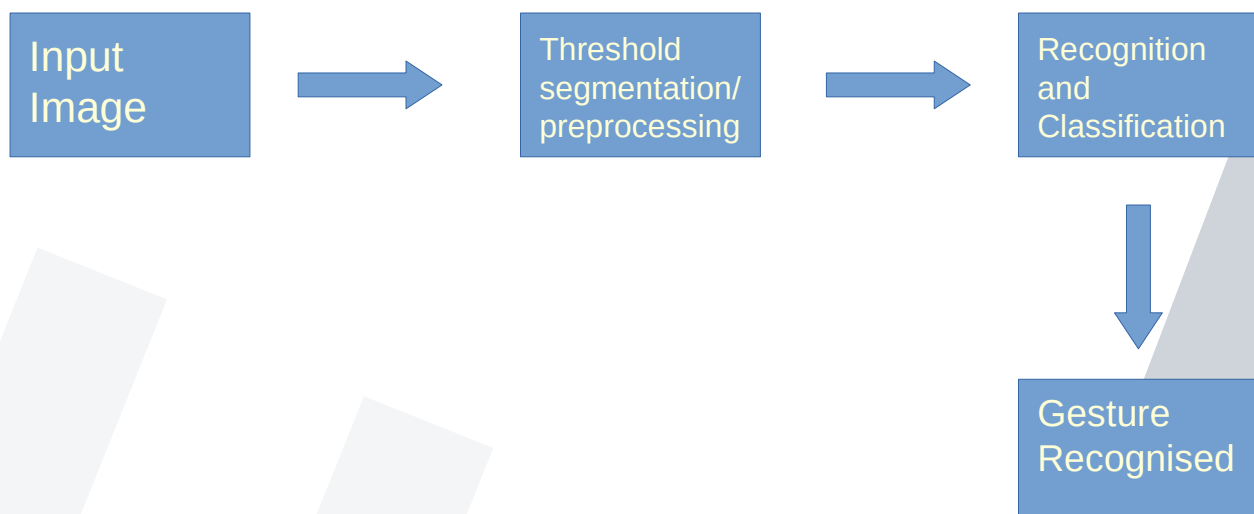
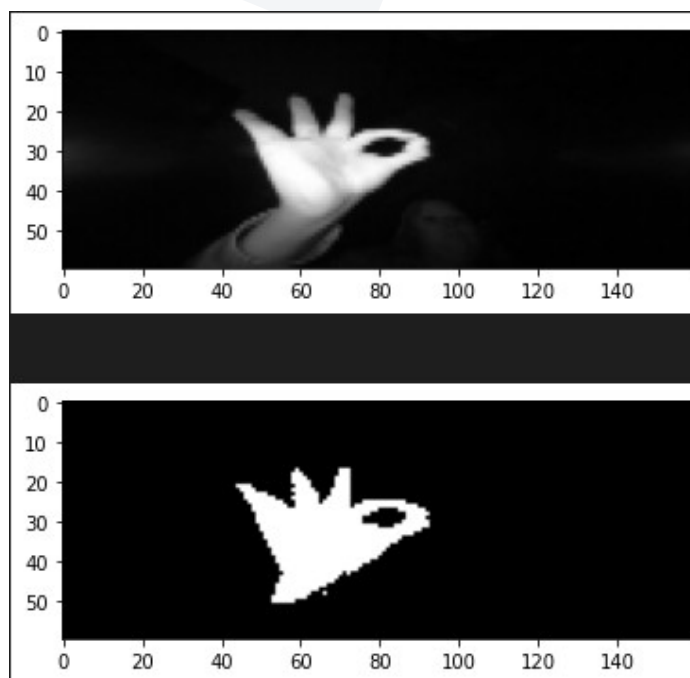
A multilayer perceptron (MLP) is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses backpropagation for training the network. MLP is a deep learning method.

Classifications methods possible for Hand Gesture Recognition/ Application area's



Proposed gesture recognition system

The proposed gesture recognition system composed of three main parts. In the first part, we get image of hand gesture, then after threshold segmentation, the image become a binary image. And after some some preprocessing we extract and process the corresponding matrix using our ML model. Finally, the gesture can be recognized using the Learning module .



- Preprocessing

Due to the nature of image, the hand region may have holes and cracks, which will definitely affect the accuracy of hand gesture. Usually the binary image will be noisy, so image preprocessing is necessary,

which fills the holes. there are two main methods for digital image restoration, texture-based method and non-textured-based method. In this paper, an exemplar image completion based on evolutionary algorithm is proposed. In the non-textured-based method, total variation method is a typical algorithm. An improved total variation algorithm is proposed in this paper. In the improved algorithm, the diffusion coefficients are defined according to the distance and direction between the damaged pixel and its neighborhood pixel. Although in other papers , a novel inpainting methods reach a good result. However, the methods are always too sophisticated and time-costing. Thus we employ some simple morphological operations like erosion and dilation to fulfill our request. In practice, we dilate two times and erode two times.

- Gesture Recognition

In this paper, the pyramidal pooling module and attention mechanism are used to increase the receptive field and classify the details more efficiently. As can be seen from figure 3, the original input image passes through the convolution layer of 3×3 and the maximum pooling layer to get the size of $1/2$ of the original image feature map. Then the four different scale spatial pyramids are pooled to get the size

of $1/4$, $1/8$, $1/16$, $1/32$ feature map respectively. So the different scale features can be captured. Then, global average pooling is used to obtain the weights of global abstract features as channel dimensions at lower levels. Finally, the final probability score of each class is obtained by using fully connected layer and softmax. Comparing with stacked convolution and pooling structure, the proposed structure can not only get the feature maps of different receptive fields quickly, but also pool the high-resolution feature maps globally, and then reduce the channel dimension by 1×1 convolution, which can be used as the weight of adjacent low-resolution feature channels effectively. Guide the formation of abstract features. A large number of experiments on our gesture data show that this structure can accelerate the convergence of the network and improve the accuracy

Conclusion

Gesture plays an indispensable role in Human-Computer-Interaction. This paper comes up with an accurate and efficient method for gesture recognition. First, skin is detected using on rules based on experience and the picture is transformed into binary image. Then expansion and corrosion are adopted. After that, all the contours are extracted and the contour of hand are found. Finally, the gesture can be recognized using the pyramidal pooling module and attention mechanism.

References

ICAACE 2019

IOP Conf. Series: Journal of Physics: Conf. Series 1213 (2019) 022001

IOP Publishing

Doi:10.1088/1742-6596/1213/2/022001

A systematic review on hand gesture recognition techniques, challenges and applications - M Yasen, S Jusoh - PeerJ Computer Science, 2019 - peerj.com

Hand gesture recognition based on computer vision: a review of techniques
M Oudah, A Al-Naji, J Chahl - journal of Imaging, 2020 - mdpi.com

Result (overview)

Logistic Regression Classifier

Accuracy of the model is:

Test 0.43

Train 0.4308333333333335

LR (Stochastic Gradient Descent optimization)

Accuracy of the model is:

Test 0.4008333333333333

Train 0.379375

K Neighbours Classifier

Accuracy of the model is:

Test 0.9941666666666666

Train 0.9920833333333333

Decision Tree

Accuracy of the model is:

Test 0.8991666666666667

Train 0.9039583333333333

Random Forest Classifier

Accuracy of the model is:

Test 0.9975

Train 1.0

MLP classifier

Accuracy of the model is:

Test 0.985

Train 0.9775

From the previous accuracy results data we can say that Linear Regression (both GD and SGD) does not perform well in case of our dataset or rather classifying image data in general where more number of output classes are present.

Where as Decision tree and random forest model performed really well compared to linear regression due to its good multi class support . Random forest regressor and ANN based model performed best among all due to their deep learning capabilities and more feature rich classification possibilities .

some of the model's confusion matrices -

Linear Regression -

	Predicted Thumb Down	Predicted Palm (H)	Predicted L	Predicted Fist (H)	Predicted Fist (V)	Predicted Thumbs up	Predicted Index	Predicted OK	Predicted Palm (V)	Predicted C
Actual Thumb Down	81	1	0	1	19	3	0	2	0	0
Actual Palm (H)	0	77	0	0	3	13	0	0	25	3
Actual L	16	28	11	35	17	0	0	0	4	2
Actual Fist (H)	0	0	0	35	2	1	0	52	21	16
Actual Fist (V)	24	16	0	0	51	14	0	0	0	0
Actual Thumbs up	0	33	21	0	17	0	0	0	45	0
Actual Index	20	0	0	0	14	3	82	0	0	0
Actual OK	0	0	0	21	0	0	0	109	0	0
Actual Palm (V)	0	5	21	32	0	20	0	3	47	1
Actual C	2	22	0	7	14	14	0	0	51	23

Random Forest Regressor-

[illegible]

MLP -

[illegible]

Novelty

The following project can be modified(using CNN and tensorflow) to do much more unique and practically more important stuffs such as -

- Can be integrated with many electronic devices such as TV, mobiles, robots etc to provide functionality control using just hand gestures such as volume controls , movement controls , can help visually impaired persons a lot.
- Virtual reality is another evolving area where these gesture detection models can be useful
- Also helpful in Healthcare sector , lets say we are in a room where voice commands are not a good option as we shouldn't make much noise inside hospital's , in such case gesture recognition can be the way out.