Advance Level Hand Gesture Recognition for Day to Day Life Use

Siddharthsinh Yadav
Dharmsinh Desai University, Nadiad, India

Abstract— In this technical era life is becoming easier day by day. But its not the same for disabled people also. As they can't use all the technologies. Blind people can't use smart phone as efficiently as we can, as they can't see where to touch. So with this hand gesture recognition we can make their life easy as they would be able to do it now with their hand gestures without needing to know where exactly to touch. The system is able to recognize 27 different and easy gestures which anybody can use easily in daily life.

Keywords: Advance Level Hand Gesture Recognition

I. INTRODUCTION

A hand gesture recognition system provide a natural, innovative and modern way of non verbal communication. It has a wide area of application in human computer interaction and sign language. The intention of this paper is to discuss a novel approach of hand gesture recognition based on detection of some shape based features. The setup consist of a single camera to capture the gesture formed by the user and take this hand image as an input to the proposed algorithm. The overall algorithm divided into four main steps, which includes segmentation, orientation detection, feature extraction and classification. The proposed algorithm is independent of user characteristics. It uses 118,562 videos for training purpose and 14,787 for validating and 14,743 for testing accuracy.

II. LITERATURE SURVEY

We are advancing in technology, from a room size computer we have reached to handheld computers. Now we are advancing to touch screen and gesture recognizing computers or laptops. Existing Technologies use some hardware or sensor data for detecting and classifying gestures or uses neural network or convolution neural network for image recognition. These technologies divides the videos into multiple frames and apply convolution neural network on the raw data converted from single frames one after the other. Due to this way of detection it fails in detecting the complete action and may consider a half action as complete. Whereas this system applies a filter of fixed number of layers to number of frames simultaneously.

III. EXISTING SYSTEM

Now a days the Microsoft surface, sticky notes, touch screen, multi touch screen etc. technologies are present but, then cannot fulfil all the requirements of the user means the Microsoft surface do not have portability, as sticky notes also fails when we want to take out some image. In market there are the touch screen devices are present but, they are not as fast as the sixth sense, they requires the more time for the processing. So that, these devices are less user-friendly.

IV. METHODOLOGY

We have used 148092 number of videos as dataset which consist of following 27 gestures : Swiping Down, Swiping Left, Swiping Right ,Swiping Up, Thumb Down, Thumb Up, Turning Hand Clockwise, Turning Hand Counter clockwise, Zooming In With Full Hand, Zooming In With Two Fingers, Zooming Out With Full Hand, Zooming Out With Two Fingers, Doing other things, Drumming Fingers, No gesture, Pulling Hand In, Pulling Two Fingers In, Pushing Hand Away, Pushing Two Fingers Away, Rolling Hand Backward, Rolling Hand Forward, Shaking Hand, Sliding Two Fingers Down, Sliding Two Fingers Left, Sliding Two Fingers Right, Sliding Two Fingers Up, Stop Sign. The model will be given JPG images as input which are obtained from video at 12 frames per seconds. The JPG images have height of 100px and variable width.

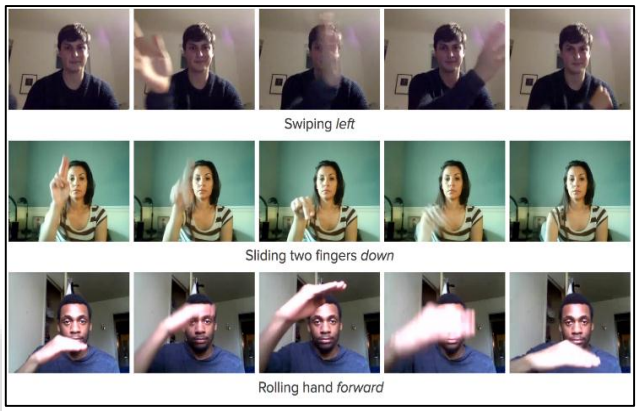

Fig. 1: Sample Movement

The model consists of one input layer, one output layer and 3 hidden layers. First Layer, 3D Convolution Layer, contains 3 filters and it's feeded by multiple frames. Output of this layer is provided as input to the next layer. The output of this layer is further provided to next softmax layer for further processing.

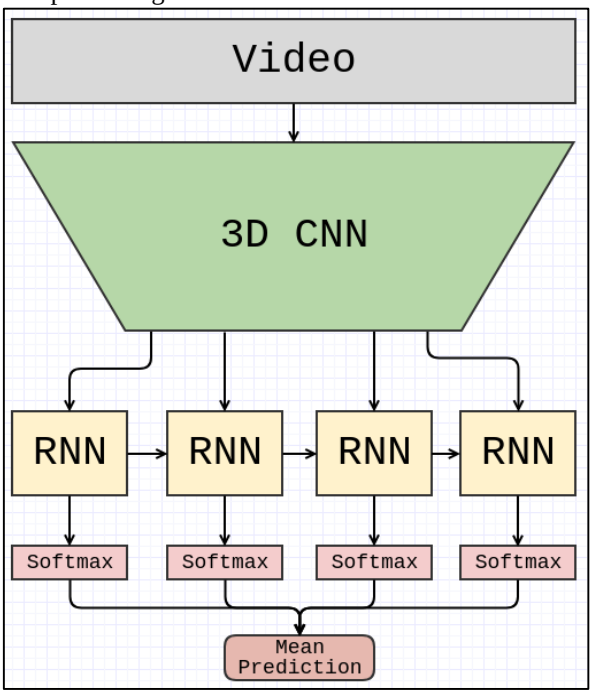

Fig. 2: Overview of model

In other hand gesture recognition techniques, model process on particular frame at a time to make the prediction but in our scenario it works with multiple frames which help to predict more accurate gesture.

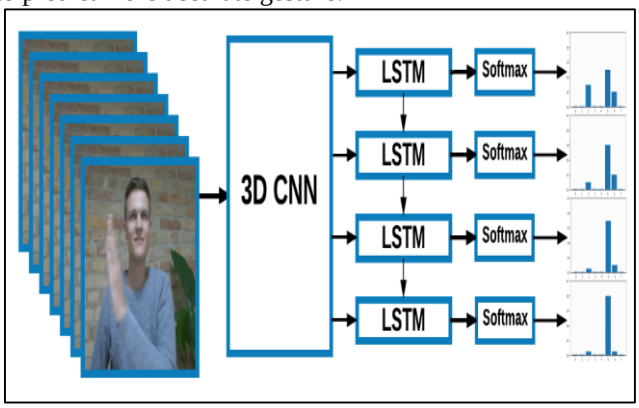

Fig. 3: Fitting video frame into model

This is model diagram and code made in deep learning studio.

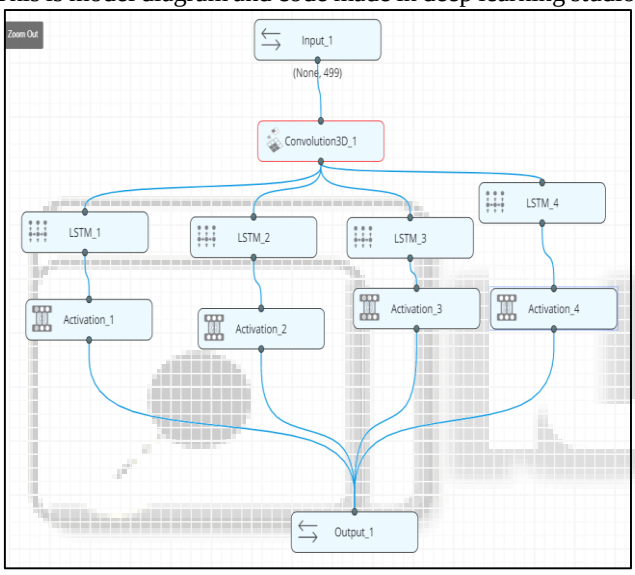

Fig. 4: Diagram of implemented model

Code for the same:

```
aliases = {}
Input_1 = Input(shape=(499,), name='Input_1')
Convolution3D_1 =
Convolution3D(name='Convolution3D_1')(Input_1)
LSTM_2 = LSTM(name='LSTM_2')(Convolution3D_1)
LSTM_3 = LSTM(name='LSTM_3')(Convolution3D_1)
LSTM_4 = LSTM(name='LSTM_4')(Convolution3D_1)
Activation_4 = Activation(name='Activation_4',activation=
'softmax' )(LSTM_4)
Activation_3 = Activation(name='Activation_3',activation=
'softmax' )(LSTM_3)
LSTM_1 = LSTM(name='LSTM_1')(Convolution3D_1)
Activation_1 = Activation(name='Activation_1',activation=
'softmax' )(LSTM_1)
Activation_2 = Activation(name='Activation_2',activation=
'softmax' )(LSTM_2)
model = Model([Input_1],[Activation_2])
return aliases, model
```

In contrast to 2D-CNNs, which are good at processing images, 3D-CNNs use three-dimensional filters which extend the two-dimensional convolutions into the time domain. Videos are processed as 3-dimensional "volumes" of frames. Using such 3D filters in the lower layers of a neural net is helpful, in particular, in tasks in which motion plays a critical role. The output of the network is a sequence of features, each of which can be thought of as a compressed representation of a small input video segment.

The feature sequence is then processed by an LSTM layer, allowing for longer time dependencies. At test time, it exploit the fact that a recurrent network is a dynamical system that can be stepped through time. At training time, each recurrent hidden state is converted into a vector of class probabilities via a softmax layer and the obtained sequence of predictions is averaged across time. The averaged vector is used to compute the loss. One can think of this as a way of asking the network to output the appropriate label as soon as possible, forcing it to stay in sync with what happens in the video. This common approach allows the model to be reactive and to output its best guess about the correct class online and before the full completion of a gesture.

3D-CNN architecture is a sequence of pairs of layers with filter size 1 and 3, in that order. Layers with filter size 1 are used to interpret channel-wise correlations and decrease the number of channels for the next layer. Layers with filter size 3 capture spatial information.

V. USAGE

This system can be used in many other applications to them easier to use. This can also be used with hardware. As an example, it can be used with microcontroller to turn on or off the tube lights, fans. It can be used in audio or video software to give input of pause, resume and fast forward. It can be used in photo viewer to slide the photos and many more other features can be integrated with this system. It is important for many real-world applications such as sign language recognition and human-robot interaction (HRI). Common usage diagram of this system for day to day application.

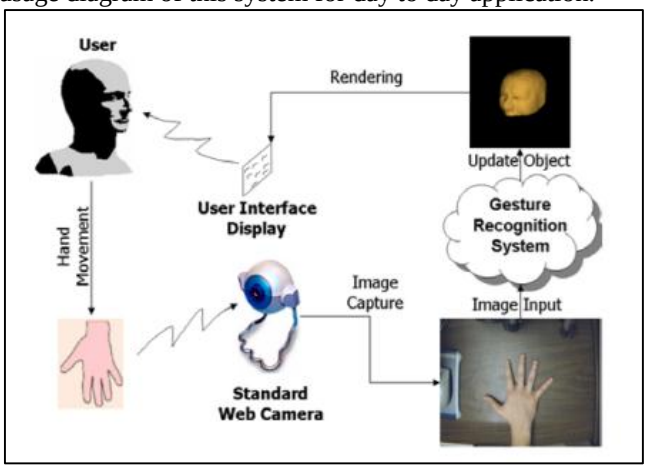

Fig. 5: Usage Diagram

VI. CONCLUSION

It is an effective method for dynamic hand gesture recognition with 3D convolutional neural networks. The proposed classifier uses a fused motion volume of normalized depth and image gradient values, and utilizes spatio-temporal data augmentation to avoid overfitting. The combination of low and high resolution sub-networks improves classification

accuracy considerably. The proposed data augmentation technique plays an important role in achieving superior performance.

VII. FUTURE WORK

This model is now working on 27 different hand gesture but this can be updated to recognize more different gesture and recognize more accurately in low light and in different environment. This system can be added web support for portable use.

After implementation of web server, this system can be used in any hardware with internet and webcam.

REFERENCES

[1] Kundu, Yang He and P.Bahl, "Recognition of Handwritten Words: First and Second Order Hidden Markov Model Based Approach, "Pattern Recognition, vol. 22, no. 3, p. 283, 1989.
[2] J. Triesch and C. von der Malsburg.Robotic gesture recognition. In Gesture Workshop, pages 233–244, 1997
[3] V. Pavlovic, R. Sharma, and T. Huang. Visual interpretation of hand gestures for human computer interaction: A review. PAMI, 19(7):677–695, July 1997.
[4] W. Freeman. Computer vision for television and games. In Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, page 118, 1999.
[5] D.Heckenberg and B. C. Lovell, "MIME: A Gesture-Driven Computer Interface", Proceedings of Visual Communications and Image Processing, SPIE, V 4067, pp 261-268, Perth 20-23 June, 2000.
[6] J. Yang, Y. Xu, and C.S. Chen, "Gesture Interface: Modeling and Learning, "IEEE International Conference on Robotics and Automation, Vol. 2, 1994, pp. 1747-1752.

