

# Image Based Hand Gesture Recognition for Human Computer Interaction

Sree Sankar.J

Assistant Professor, Dept. of ECE, Gojan School of Business and Technology, Tamil Nadu, India

**ABSTRACT:** Hand gesture are non-verbal form of communication. It is commonly used by deaf and dumb people who have problem to communicate with normal people. Various sign language systems have been developed around the world but they are neither cost effective nor accurate. In this paper we propose a model based on deep learning to improve the accuracy of image based hand gesture recognition. A small Visual Geometry Group (VGG) Net Architecture with 16 Convolution layer in Convolution Neural Network (CNN) is used here to process the input images from the web camera. By using this methodology, the system is able to achieve an accuracy of 98% for static hand poses.

**KEYWORDS:** Sign Language, Hand Gesture, CNN, smaller VGG Net

## I. INTRODUCTION

Sign language is a type of language that uses visual model to convey meaning. It is a way of communication through hands and other body parts. The sign language is expressed through various signs. The main difference between the spoken language and the sign language is that in spoken language the sound is coming from the mouth and is heard by the ear whereas the sign language conveys the meaning to others through manual sign stream with non-manual elements. Sign languages are natural language which has its own grammar. For deaf people, sign language is a useful means of communication.

Sign language is not only used by deaf people but also used by dumb people. There exist many sign languages in the world today. Each country has its own sign language and some countries have more than one language. Sign language borrows some elements from spoken language. Sign language is difficult to understand by the normal people. Hence there exist a communication gap between the normal people and deaf and dumb people. To overcome this difficulty various sign language systems has been developed around the world. But some of the systems are proved to be neither cost effective nor accurate.

To improve the accuracy, an image based hand gesture recognition based on deep learning is proposed here. Image processing deals with the changing nature of the image in order to either improve the pictorial information of the image for human interpretation or to render it more suitable for autonomous machine perception. It is very important to realize that these two aspects mentioned here represents two separate, but are equally important, aspects of image processing.

This paper is organized as follows: Section II describes the related works of hand gesture recognition and is followed by Section III which represents the proposed methodology. Section IV deals with the experimental results obtained and eventually Section V presents the conclusion.

## II. RELATED WORK

One of the existing system consists of a software which presents a system prototype which is able to recognize sign language automatically to help the deaf and dumb people to communicate more effectively with each other and also

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

with the normal people. Pattern recognition and gesture recognition are the developing fields of research and are playing key role in our daily life. By considering this in mind the similarities of human hand shape with four finger and one thumb, the software aims to present a real time system for recognition of hand gesture on basis of detection of some shape based features like orientation, centre of mass centroid, fingers status, thumb in positions of raised or folded fingers of hand.

Another existing work is the classical approach. The classical classification method used consists of pre-processing, windowing, feature extraction and classification. Despite of the promising performance that have been shown in recent research, the greatest disadvantage of these classical methods is that some useful information may be discarded while extracting the feature. In this existing system, a novel model based on deep learning is proposed to improve the accuracy of EMG-based hand gesture recognition. A parallel architecture with five convolution layers is adopted. The Surface electromyographic signals which can be noninvasively detected by the surface electrodes have been used. It can be used to generate device control commands for these applications. However, the Surface electromyographic signals can be easily interfered such as the position of electrodes, the individual skin formation, and the presence of electronic equipment. This will lead to inaccuracy in recognition which is the biggest obstacle between application and researches.

### III. PROPOSED METHODOLOGY

The hand gesture images are captured by the web camera. For gesture recognition Convolutional Neural network (CNN) is used here. In CNN, a small VGG Net architecture with 16 convolutional layers are employed. A dataset consisting of multiple hand gestures are used here. Initially the input images are converted into 96x96 frame size. The frame is scaled in order to eliminate the unwanted area of interest. The proposed training and testing block diagrams are shown in the figure 1 and figure 2.

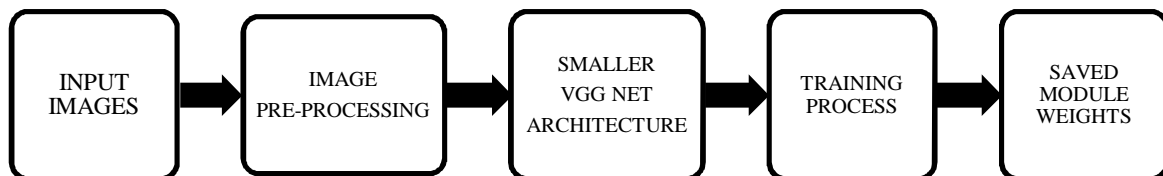


Fig.2.1 Training block diagram



Fig.2.1 Testing block diagram

Here the data pre-processing consists of four steps namely Uniform Aspect Ratio, Image Scaling, Data Augmentation and Cross Validation. It is very important to ensure that all the images are in the same size and aspect ratio before performing any kind of further pre-processing steps. Normally most of the Neural Networks make sure that all the images are in a perfect square shape or not. If the images are not a perfect square, then cropping of images can be done to make it a perfect square and same aspect ratio. After attaining all the images with same aspect ratio Image Scaling is done. Image Scaling is the process of resizing the image. While scaling the image the primitives that makes up the

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

image can be scaled without any loss of image quality. A Convolutional Neural Network (CNN) classify the object even if it is placed in a different orientation. This property of CNN is referred to as invariance.

CNN is invariant to translation, view point, size etc. thus it forms the base for Data Augmentation. Here data augmentation concept is used to build an optimized image classifier with a little amount of training data. The Data Augmentation process artificially creates training images from the given little amount of data through different ways of augmentation techniques like shift, rotation etc. The final step in pre-processing is Cross Validation. Cross Validation process is carried out to increase the accuracy. It involves partitioning a sample of data into complementary subset and performing the training process in one subset and then validating the testing process in another subset.

Here Convolutional Neural Network is used for image based hand gesture recognition. CNN is a specific type of Artificial Neural Network (ANN) that uses perceptrons for supervised learning. CNN is also known as ConvNet. It consists of several layers namely convolutional, nonlinear pooling and fully connected layer. The input image is passed through these layers and the output is finally obtained. CNN has different architecture according to weights used in the process. Here 16 weights are used since a smaller VGG net architecture is used. Python software is used here. Python is a high level general purpose programming language. It supports multiple programming paradigms and it is an open source software. The program is processed by Raspberry Pi 3 model.

## IV. EXPERIMENTAL RESULTS

This section deals with the various experimental results obtained. Figure 3.1 shows the training loss and accuracy graph obtained and Figure 3.2 shows the various hand gestures identified.

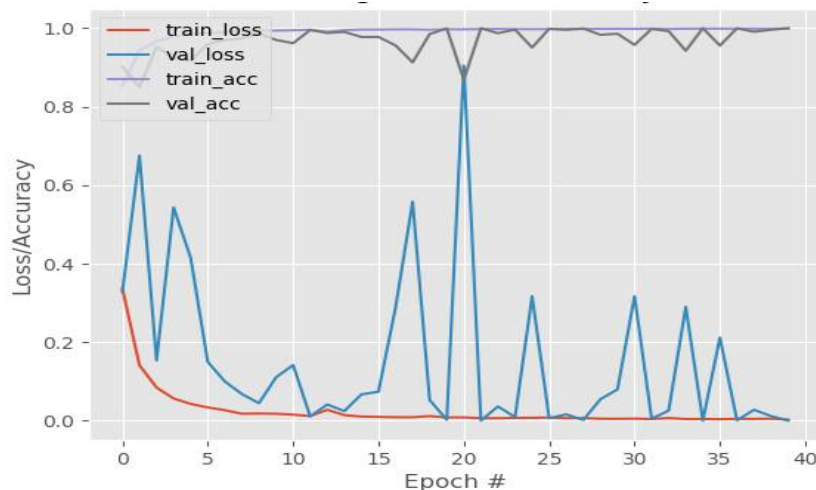


Fig 3.1 Training loss and Accuracy graph obtained

The process of forward pass, loss function, backward pass, and parameter update is one training iteration. The program will repeat this process for a fixed number of iterations for each set of training images (commonly called an epoch). we use 40 epoch for good accuracy.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

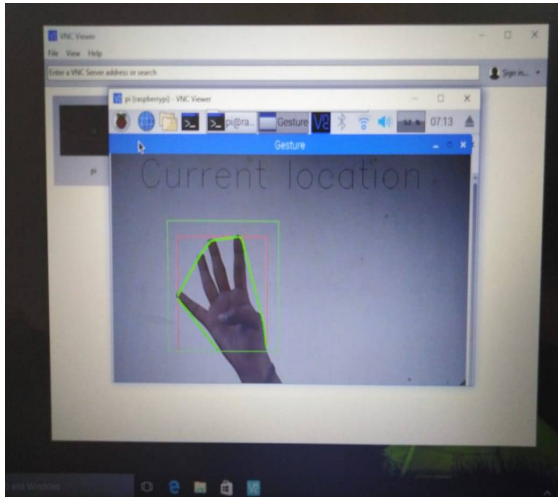


Fig 3.2.1 Output for the input gesture

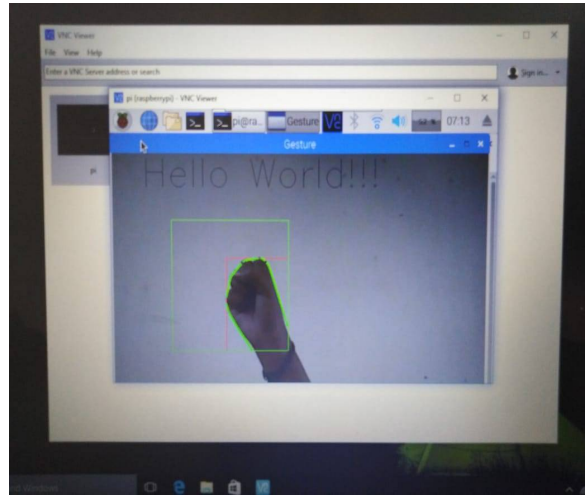


Fig 3.2.2 Output for the input gesture

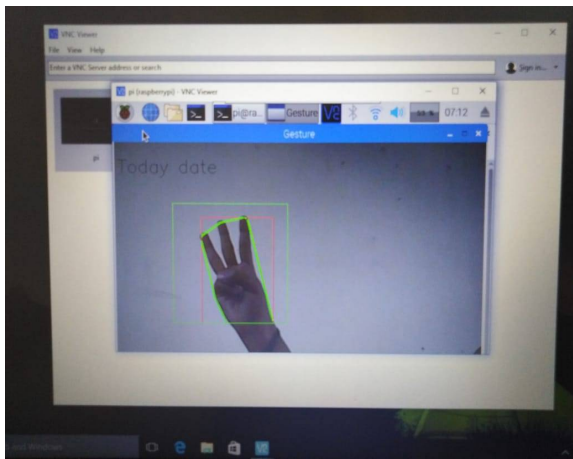


Fig 3.2.3 Output for the input gesture

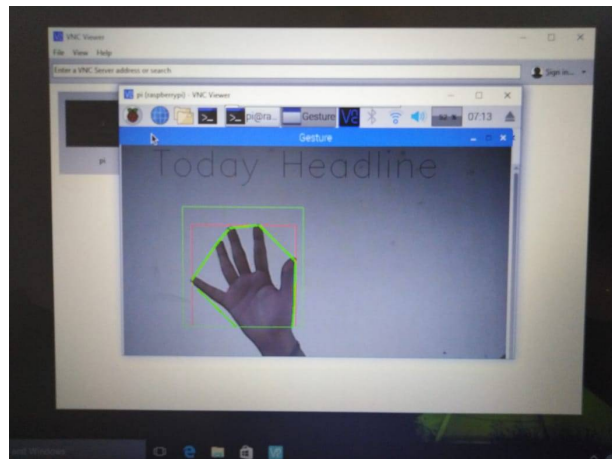


Fig 3.2.4 Output for the input gesture

Fig 3.2 Various Hand gestures identified

The Figure 3.2 shows the various outputs obtained for various input hand gestures. In Figure 3.2.1, this particular input hand gesture represents the current location and is the resultant output. In Figure 3.2.2, this input hand gesture represents the message hello world. Similarly, Figures 3.2.3 and 3.2.4 represents the message today date and today headline respectively.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 5, May 2019

## V. CONCLUSION

Here we have seen the importance of various parameters that causes dramatic changes for deaf and dumb people, it has become a necessary advancement of the image processing. Thus we can avoid deaf and dumb people struggling to communicate with the normal people using sign language. In this paper Convolutional Neural Network is used for hand gesture recognition. Which has been proved to have high accuracy rate. As an extension for this work many other deep learning techniques can also be used. It's our duty and moral obligation to help the people who are struggling to express their perspectives. Thus sign language deduction act as an effective method for deaf and dumb people in future.

## REFERENCES

- [1] Rho, S.; Yeo, S.-S. Bridging the semantic gap in multimedia emotion/mood recognition for ubiquitous computing environment. *J. Supercomput.*, 65, 274–286, doi:10.1007/s11227-010-0447-6, 2013.
- [2] B. Ionescu, D. Coquin, P. Lambert, and V. Buzuloiu, "Dynamic hand gesture recognition using the skeleton of the hand," *EURASIP Journal on Advances in Signal Processing*, vol. no. 13, pp. 2101–2109, 2005.
- [3] H. Francke, J. Ruiz-del Solar, and R. Verschae, "Real-time hand gesture detection and recognition using boosted classifiers and active learning," in *Pacific-Rim Symposium on Image and Video Technology*, Santiago, Chile, pp. 533–547, 2007.
- [4] N. H. A.-Q. Dardas, "Real-time hand gesture detection and recognition for human computer interaction," Ph.D. dissertation, Université d'Ottawa/University of Ottawa, 2012.
- [5] Y. Yin, "Real-time continuous gesture recognition for natural multimodal interaction," Ph.D. dissertation, MIT, 2014.
- [6] Y. Xu, Q. Wang, X. Bai, Y.-L. Chen, and X. Wu, "A novel feature extracting method for dynamic gesture recognition based on Support Vector Machine," in *IEEE International Conference on Image Processing*. Paris, France: IEEE, pp. 437–441, October 2014.
- [7] I. Nigam, M. Vatsa, and R. Singh, "Leap signature recognition using hoof and hot features," in *IEEE International Conference on Image Processing*. Paris, France: IEEE, pp. 5012–5016, October 2014.
- [8] K.-Y. Fok, N. Ganganath, C.-T. Cheng, and C. K. Tse, "A real-time asl recognition system using leap motion sensors," in *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. Xi'an, China: IEEE, pp. 411–414, September 2015.
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [10] Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
- [11] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *International Conference on Machine Learning*. New York, NY, USA: ACM, pp. 1096–1103, 2008.