

Static Hand Gesture Recognition System using Convolutional Neural Networks

V. Umasankara Rao¹ K. Pavani²

¹Assistant Professor ²Student

^{1,2}Department of Computer Science & Engineering

^{1,2}Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

Abstract— Deep learning is driving advances in artificial intelligence. It is a subset of machine learning in artificial intelligence that has networks which are capable of unsupervised learning from data that is unstructured. The neural network is a computer system modeled on a human brain and nervous system where a particular input leads to a specific target output. Glove based gesture detection requires an extra equipment like sensors, specially designed gloves. In this paper, computer vision based gesture detection is used with the convolutional neural networks. Gestures captured from the camera are given to trained network as an input to interpret the gesture. Accuracy will be improved as the convolutional neural network is created with the dataset of hand gestures.

Key words: Convolutional Neural Networks, Hand Gesture, Deep Learning

I. INTRODUCTION

The main important goal of gesture recognition of hand is to create a system which can identify any given hand gesture. People perform various gestures in their daily lives. It is our nature to use gestures in order to improve the communication. Between all kind of gestures that we perform, hand gestures play an important role. Hand gestures can help us to say more information in less time. Nowadays, computers have become an important part of our lives, so why not use hand gesture in order to communicate with them. There are two approaches commonly used i. e, the data glove approach and, vision-based approach.

The DataGlove Approach: Data Glove approach employs mechanical or optical sensor attached to the glove that transforms finger flexions into electrical signals to determine hand posture. The glove is constructed with a series of sensors that detect the hand and finger motion. However, this method requires the glove must be worn and a device with a load of cables connected to the computer. This method is not used for daily communication because of sensors, cables, and glove. It is useful for limited applications.

A. Vision-based Approach

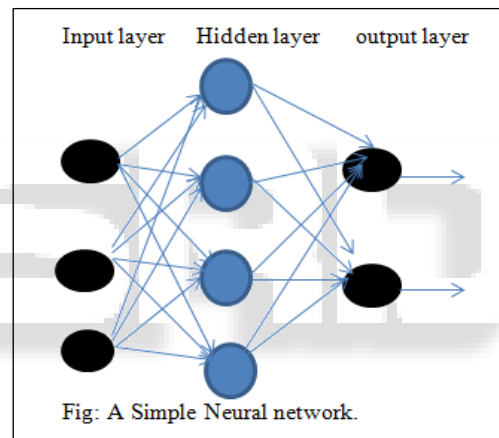
The proposed system provides a natural, innovative and modern way of non-verbal communication. It has a wide area of application in human-computer interaction. We have used the vision-based approach for our gesture recognition. When compared to glove based it doesn't require any sensors, cables and, glove. It is mostly used for daily communication.

In the proposed approach we are using Convolutional Neural Network to process the data efficiently and help in recognizing the gesture. It is simple and fast and it is speed and sufficiently reliable for recognition systems. It can be applied to real-time applications.

B. Deep Learning

Deep learning is driving advances in artificial intelligence. It is a subset of machine learning in artificial intelligence that has networks which are capable of unsupervised learning from data that is unstructured. It is the most interesting and powerful machine learning technique. Top deep learning libraries are available on the Python-like Theano and TensorFlow. The main advantage of this is you can get started with neural networks in an easy way.

Deep learning techniques are so powerful because they learn the best way to represent the problem while learning how to solve the problem. This is called representation learning. A simple neural network is shown below:



Generally, it consists of three parts namely the input layer and hidden layers output layer. We can add and remove the hidden layers based on the user's requirement. By using the activation function we are going to store the information in an input, hidden and output layers.

Activation Functions: There are three types of activation functions:

- Threshold Functions
- Piecewise-linear Functions
- Sigmoid Functions

Compare to sigmoid functions both threshold and piecewise-linear functions have some drawbacks. For network training, we may need the activation function to be differential. The most commonly used activation functions are the sigmoid functions. The formulae for the sigmoid function is

$$\sigma(z) = \frac{1}{1 + \exp(-z)}$$

Neural Networks and Deep Learning: Neural networks and deep learning currently provide the best solutions to many problems. The Neural network is a computer system modeled on the human brain and nervous

system where a particular input leads to a specific target output. The following are the some of the components:

- Unsupervised Pre-trained Networks (UPN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks
- Recursive Neural Networks

C. Convolutional Neural Network

Convolutional Neural Networks are one very interesting sub-field and one of the most influential innovations in the field of computer vision. Convolutional Networks are used in the situation where the data can be expressed as a map. CNN are commonly used to categorize things in an image. CNN takes an image and expressed it as an array of numbers, applies a series of operations to that array, at the end returns the probability that an object in the image belongs to a particular class object. By using Convolutional Neural Networks, we can get almost human results. CNN having the following layers:

- Convolution Layer.
- ReLU (Rectified Linear Units) Layer.
- Pooling Layer.
- Fully Connected Layer.
- Loss Layer.

II. RELATED WORK

The man-machine interface playing very important role in today's life. It is better to implement a model that recognizes user provided patterns and identifies that pattern in the machine. Researchers are made available in face recognition, hand detection, speech recognition, and voice recognition and in more areas.

Hand gesture recognition is a part of Human-Computer Interaction (HCI), the hand is given as an input and machine should recognize the gesture provided by the user and it should give the output based on the gesture. The existing work is saying that the gesture is recognized based on glove attached to the hand. Sensors are fixed to the glove and the information is read by the sensors, it then sent to the microcontroller at the receivers end. The signals from sensors are read and the intelligent system provides the gesture performed by the user.

A computer vision based system able to recognize 14 gestures in real life to handle windows and article within a graphical interface was developed by C.W. Ng et al. in [3]. Abe et al. [4] proposed a system which recognizes hand gestures that detect the bending of the hand's five fingers, based on image-processing as a technology.

Franklin et al. [5], hand gestures are controlled by a robot assistant by using Perseus architecture for gesture recognition. In the work of Cipolla et al. [6], a gesture-based interface for robot instruction is based on un-calibrated stereo vision and active contours.

Ramamoorthy et al. [7] used HMM-based real-time dynamic gesture recognition system which shape characteristics and temporal characteristics of the gesture for recognition. Hand patterns and shape patterns are the novel feature of this work.

Chen et al. [8] proposed gesture recognition of hand by continues monitoring the hand. This proposed system

consists of four modules: real-time hand tracking and extraction, feature extraction form hand, Hidden Markov model (HMM) training for recognition, and recognition of the gesture.

Yin Xiaoming et al. [9] used an RCE neural network based color division algorithm for hand segmentation, extract edge points of fingers as points of interest and match them based on the array of features of the hand, such as the center of the palm.

Xiong et al. [10] explored the utility of movement symmetries of a speaker's hands when they are both possessed in correspondence. For taking care of hand movement symmetries they have utilized an approach in view of the connection calculations by allocating a two-organize calculation of window based relationship and 'gap filling'.

The utilization of profound catching innovation is quickly developing in prominence, and different apparatuses have been installed into the procedure that demonstrated effectively. Improvements, for example, man outlined gloves have been utilized to encourage the acknowledgment procedure and influence the component extraction to step more effective by making certain gestural units less demanding to distinguish and sort. [11].

III. EXISTING SYSTEM

Hand gesture recognition is not new to the computer science these days. But gesture recognition using hand gloves has certain limitations. It is not always possible to wear a glove and detect the gesture because it requires the number of equipment like sensors, specially designed purpose gloves, the microprocessor to read signals from the glove and, wire connection if it is not the wireless sensor. In order to reduce the burden on the users of gesture system it is not efficient to use glove-based hand recognition system.

IV. PROPOSED WORK

The Hand Gestures can be used to solve many challenges. It is not always possible to wear hand glove and read signal from it. A vision based approach; a camera is used to capture gestures from the hand by any end user. All these captured images are given as an input to the model which is already trained.

A. OpenCV

Open source computer vision library is used as an interface between the user and a machine. OpenCV comes with many versions supports many languages like C, Python, C++.

B. Capture Histogram

Histogram solves many problems related to images. It is possible to differentiate between poorly exposed images and perfect images with the histogram. So generating histogram from the captured image is necessary.

C. Hand Contour

Finding hand contour is important to find palm of the hand and it is based on OpenCV method cv2.contourarea(). It finds contour area of hand to further feed to model.

D. Mark Fingers

The necessary step in this approach is to detect fingers. Because the vision-based approach uses a camera to detect the hand and not processing any signals from the hand. The histogram of hand is given as input to this function and it detects the fingers from the hand. After the hand is detected it is now needed to display the hand.

E. Background Subtraction

Not taking much time, it is now needed to subtract the background from the detected hand. This step is needed to have an only hand to be processed. The background can be subtracted by `backgroundsubtractor()` method.

F. Training the model

The much important step in the proposed work is to train the model. Putting sense into a network is the deep neural network. As there are many techniques to train the model using deep learning, here a convolution neural network is used to train the dataset. This dataset contains gesture of hand positioning ok, index pointing, and five fingers.

The Convolutional neural network is created on the dataset with two hidden layers. The Relu is used as an activation function for hidden layers. Softmax is used as an activation function for the output layer. At each epoch, the accuracy is increased. The learning rate of the model will be efficient with convolve layers. The model was trained with convolve2D layers.

As Hand Gesture is captured from openCV and is preprocessed to compare with the model trained already. It is given as an input to the model. The model then compares the dataset which is already trained with the hand gesture generated by OpenCV. It gives better accurate results because it is trained using the deep neural network. As our work produces better results, it is efficient to use vision-based approach than the glove-based system.

1) Algorithm steps:

Step 1: Generate histogram from recognized hand.

Step 2: Find hand contour, palm in order to detect the hand center.

`cv2.contourarea()`

Step 3: Subtract unnecessary background from the hand detected.

`backgroundsubtractor()`

Step 4: Detect and mark fingers.

Step 5: Input the detected fingers to already trained model.

V. RESULTS



From the figure, the hand which is shown is from OpenCV and the number of fingers is displayed as our system detects number of fingers. That image captured by OpenCV is given as an input to the network. The gesture 'victory' is the output from the convolutional neural network, which is trained.



Like-wise six different gestures are given to the network to train. Our approach detects these gestures efficiently by comparing with the vision media. These results are accurate and tested on a computer with 8GB RAM and 1TB hard disk.

VI. CONCLUSION AND FUTURE WORK

As technology inviting us to do research on better things to make them even better, it is our responsibility to provide better solutions to the world. Hand gestures are playing an important role in many applications. We can interpret sign language, virtual reality, military applications and, many more applications can make use hand gestures. Dynamic hand gesture recognition can be used more efficient than this static system. It is the future work of this project. Military applications can make use of this dynamic hand gesture recognition system.

REFERENCES

- [1] Pławiak P, Sośnicki T, Niedźwiecki M, Tabor Z, Rzecki K. Hand body language gesture recognition based on signals from specialized glove and machine learning algorithms. *IEEE Transactions on Industrial Informatics*. 2016 Jun;12(3):1104-13.
- [2] Bobić, Vladislava, Predrag Tadić, and Goran Kvaščev. "Hand gesture recognition using neural network based techniques." *Neural Networks and Applications (NEUREL), 2016 13th Symposium on*. IEEE, 2016.
- [3] C. W. Ng and S. Ranganath, Real-time gesture recognition system and application. *Image Vis. Comput.*, vol. 20, no. 13-14, pp. 993-1007,2002
- [4] K. Abe, H. Saito, and S. Ozawa, Virtual 3-D interface system via hand motion recognition from two cameras. *IEEE Trans. Syst., Man, Cybern. A*, vol. 32, no. 4, pp. 536-540,2002
- [5] D. Franklin, R. E. Kahn, M. J. Swain, and R. J. Firby: Happy patrons make better tippers-Creating a robot waiter using Perseus and the animate agent architecture. In *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Killington, VT, pp. 14-16,1996

- [6] R. Cipolla and N. J. Hollinghurst,: Human–robot interface by pointing with un-calibrated stereo vision. *Image Vis. Computing.*, vol. 14, no. 3, pp. 171–178,1996
- [7] Aditya Ramamoorthy, Namrata Vaswani, Santanu Chaudhury, Subhashis Banerjee,: “Recognition of dynamic hand gestures”. *Pattern Recognition* (36), pp 2069-2081,2003.
- [8] F. S. Chen, C.M. Fu and C.L. Huang,: Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image Vision Computer*, pp. 745-758,2003
- [9] Y. Xiaoming and X. Ming,: Estimation of the fundamental matrix from uncalibrated stereo hand images for 3D hand gesture recognition", *Pattern Recognition*, vol. 36, pp 567-584 ,2003.
- [10] Yingen Xiong, Francis Quek & David McNeill,: Hand Gesture Symmetric Behavior Detection and Analysis in Natural Conversation. In the Proceedings of the Fourth IEEE Int. Conference on Multimodal Interfaces, 2002.
- [11] Cao Dong, Ming C. Leu and Zhaozheng Yin. American Sign Language Alphabet Recognition Using Microsoft Kinect. In IEEE International Conference on Computer Vision and Pattern Recognition Workshops, 2015.

