

## Hand Recognition and Gesture Control using Neural Networks

Senthil Kumar G <sup>1\*</sup>, Sampath T <sup>2</sup>, Mohan Ram M <sup>3</sup>, Muthu Veerapan N <sup>4</sup>

<sup>1,3,4</sup>*Department of Computer Science and Engineering, Chennai Institute of Technology, Chennai, India.*

<sup>2</sup>*Department of Electronics and Communication Engineering, Chennai Institute of Technology, Chennai, India.*

*senthilkumarg@citchennai.net,sampath.t78@gmail.com,mohanramm@cichennai.net,muthuveerapann@citchennai.net*

### Abstract

*In the modern era, lots of things are getting automated and everything is getting digitized. The daily interaction between computers and humans is essential. The development in the interaction between humans and computers is being limited as we need a special type of physical equipment to trigger a command to start a process. To make this procedure increasingly intelligent, the signal acknowledgment has become a vital issue for investigation in the field of PC vision, where PCs can be made to increase a significant level of comprehension from computerized pictures or recordings. A lot of methods have been proposed for recognizing and processing the gesture movement using various sensors and neural networks. In this project, we provide a way for recognizing the hand gesture and processing it using computer vision. The proposed framework utilizes the webcam of the PC framework to take the hand motions as info and the cursor reacts as needs. It can be applied to control electronic appliances and it also uses an optical recognition system that recognizes gestures in different light conditions. The various gestures performed by the user are trained at first to avoid the miscalculation and basically, gestures are being interpreted by mathematical algorithms. Various gestures performed by the user such as rotating clockwise and counter-clockwise, pushing, holding and swiping and with the help of gestures we can also write an alphabet. The gesture processing can be applied to various fields like augmented reality and virtual reality and it's also applicable in the medical as well as engineering field.*

**Keywords:** *Hand recognition, gesture recognition, gesture control*

### 1. Introduction

In this digital era, the computerized gadgets are made to shrewd gadgets that are ready to work without any human intervention. It may very well be controlled remotely or by giving the motions as a contribution to the gadgets. Hand signals are basic developments of an individual's hands and are the nuclear correspondence parts speaking to the contemplations of an individual. Considered as a kind of human development in our everyday life, hand motions have been examined by means of human movement investigation, which give more consideration to the full-outline human stances and exercises. As of late, there are a few overviews which stress on notable hand signal acknowledgment. Developmental anthropologists advise us the utilization of hand motions which has been utilized since the start of mankind's history and are a lot more established than discourse. Motion acknowledgment is a functioning exploration field that attempts to join the gestural direct in Human-Computer Interaction. It has programs in the advanced condition oversee, yet additionally in gesture-based communication interpretation, robot remote control or melodic creation. Signals can range from any physical movement or state from the face or hand. Here the primary concern which interfaces the human and the gadget is PC vision which can be signified as OpenCV. This OpenCV assumes an essential job in propelling keen gadgets. OpenCV (Open source

computer vision) is a library of programming functions mainly aimed at real-time computer vision. OpenCV offers a new framework to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects. Convolution Neural Networks is utilized alongside OpenCV which makes the gadget to perceive the signals. These two has a significant influence on human and PC association (HCI). A particular very such a profound neural system is that the convolutional organize, which is generally named as CNN or ConvNet. It is a profound, feed-forward counterfeit neural system. These frameworks can get to the camera in the computerized gadgets and it initially perceives the Human hands with signals. It comprehends those signals and makes the gadget to react with the move identified with the motion. By and large, the profundity-based hand motion acknowledgment forms fall into three classes: static hand signal acknowledgment, hand direction motion notoriety, and constant hand motion ubiquity. All of 3 sorts of hand signals can use the 3D hand displaying for fine-grained motion (for example finger development) acknowledgment. The fundamental piece of this work is that the motion information can be given as either static or dynamic.

## **2. Literature Review**

### **2.1 Gesture Recognition**

Motion acknowledgment could be a subject in figuring and language innovation with the objective of disentangling human motions by means of numerical calculations. Signals will begin from any substantial movement from the face or hand. Current concentrations inside the field exemplify feeling acknowledgment from face and hand signal acknowledgment. Motion acknowledgment could be a style of tactile movement processing program that grants PCs to catch and decipher human signals as orders. The last meaning of signal acknowledgment is that the capacity of a PC to get a handle on motions and execute orders bolstered those motions. Most customers are at home with the origination through Wii coordinate, X-box and PlayStation games like "Simply Dance" and "Kinect Sports."

### **2.2 CNN (Convolution Neural Networks)**

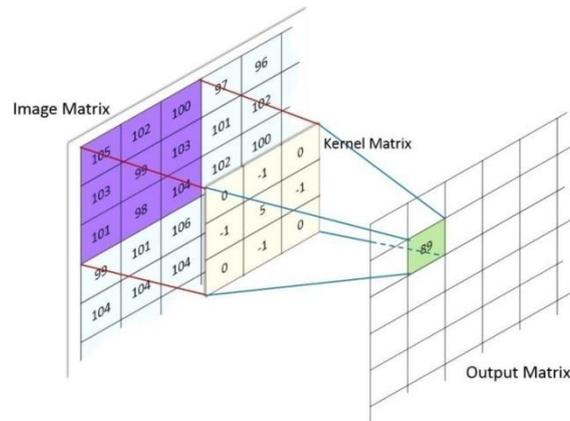
CNN is one of the common types of Neural Networks and it is proved that the Convolution Neural Network is one of the best algorithms in the neural networks and in the scheme of Artificial Intelligence. A neural system having different shrouded layer is by and large alluded to as a Deep Neural Network. Convolution Neural Networks (CNN) is one of the varieties of neural systems utilized vigorously inside the control of Computer Vision. It gets its call from the type of shrouded layers it comprises of. The concealed layers of a CNN by and large comprise of convolution layers, pooling layers, totally related layers, and standardization layers. Here it for all intents and purposes technique that as opposed to utilizing the regular initiation capacities portrayed above, convolution and pooling abilities are utilized as enactment highlights. It has been conclusively demonstrated during that time that neural systems beat different calculations in exactness and speed.

### **2.3 Convolution**

Convolution works on two signs (1D) or photos (2D): you could consider one the "enter" sign (or picture), and the other (called the part) as a "get out" on the enter picture, creating a yield picture (so convolution accepts photos as enter and delivers a third as yield). In layman terms it takes in an info signal and applies a get out over it, basically duplicates the enter signal with the bit to get the changed sign. Scientifically, a convolution of two highlights  $f$  and  $g$  is characterized as which is a

$$(f * g)(i) = \sum_{j=1}^m g(j) \cdot f(i - j + m/2) \quad (1)$$

Speck result of bit and info work



**Figure 1. Convolution matrix**

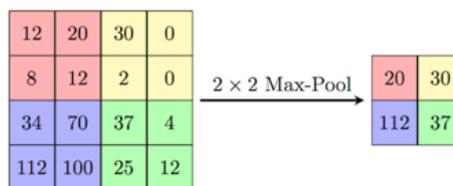
In image processing, the visualization of the kernel is easy as it sliding overran entire image so that the value of each pixel changes.

## 2.4 Pooling

A pooling layer is yet another building block of a CNN after convolution. Its characteristic is to progressively lessen the spatial size of the illustration to lessen the quantity of parameters and computation in the network. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. Pooling layer operates on each feature map independently.

The most common approach used in pooling is *max pooling*.

Max pooling is only a pooling that takes the most extreme qualities from the chose area and the min pooling is one that takes the base qualities from the chose locale.



**Figure 2. Pooling table**

In this way, as it is anything but difficult to see A Convolution Neural Network or CNN is to a great extent a profound neural network which incorporates concealed layers having convolution and pooling abilities further to the initiation trademark for presenting non-linearity.

## 2.5 3D Hand Modeling

3-d hand displaying and checking gauge the explained hand postures and movements. These are key advancements in heaps of HCI bundles, for example, mechanical medical procedure, virtual console, and communication via gestures acknowledgment. As indicated by the scientific categorization proposed by, hand present estimation systems can be classified into two classifications: discriminative methodology and generative

procedures. Additionally, the cross breed three-D displaying which joins discriminative demonstrating and generative demonstrating has been likewise proposed. Next, we can survey different three-D displaying methodology inside the setting of 3-D hand present estimation and track with acknowledge to these three classes.

### **2.5.1 Discriminative 3D Hand Modeling**

Discriminative procedures do never again oblige the three-D hand model with explicit levels of opportunity. Rather, they show the classifiers which conversely map appearance-specific hand pixel capacities to obscure hand parameters, for example, part mark, present parameter, and so on. The classifiers are ordinarily taken in disconnected from a major arrangement of tutoring tests. Also, limit of those methodologies lease the choice tree-based classifiers to support up the estimation in each body autonomously.

### **2.5.2 Generative 3D Hand Modeling**

The generative techniques are celebrated among ongoing spic and span three-D hand demonstrating and following works. They are additionally called model-based thoroughly hand checking procedures, which attempt to fit the unequivocal 3-D DoF form to the discovered hand information. Note that either 2D or 3-d hand photos can be utilized to appraise the 3-D hand variant.

3-D Hand Modeling with a Manipulated Object: All the above 3D hand following procedures can give promising hand present estimation results when the hand is exclusively situated in confinement. Be that as it may, they may never again function admirably while the hand is controlling a thing. To deal with the solid impediment through the controlled thing, Hamer et al. proposed a gadget to recuperate the explained 3-D structure of the hand for the term of thing control. They utilized the human hand model which incorporates 27 bones and rendered each hand area through a chamber with a work approximating the skin. Both hand division and profundity information were utilized to assess the probability likelihood among the hand configuration and the watched RGB-D information.

### **2.5.2 Crossover 3D Hand Modeling**

As of late, the crossover three-D hand displaying systems which join discriminative and generative techniques are additionally proposed. Given a three-advance pipeline wherein the first two stages were basically based at the discriminative profundity elements of the Hough timberland relapse model. The primer hand present estimation was seen through the final verification step which improved a 27 DoF hand model in a generative manner. Simultaneously, it was proposed to blend a discriminative part-basically based posture recovery strategies with a generative posture estimation method dependent on neighborhood improvement. The two works demonstrated that the blend of discriminative and generative thoughts can accomplish present day hand displaying precision as pleasantly as unreasonable efficiency.

## **3. Proposed Solution**

Till date, we are using some hardware devices like data gloves, magic rings for gesture recognition and leap motion sensors for recognizing the Leap motion. Here we are using only the camera which is already available in the device and it is getting implemented this in both gesture recognition and Leap motion. In this idea, the main advantage is that it can be useful in the Medical and the Engineering field. In the medical field of Today's modern world, all the treatments are done using the machines for x-rays and operational purposes. So, these machines and equipment can be controlled by a hand gesture that is easier to use

than operating it by buttons and manually by hand. In the Engineering field, we can use these for construction purposes and for making the designs for 3D models and for developing the blueprint of any architecture. It can also be used for creating virtual environments for chemical bonding with which we can able to understand the reaction of bonding any elements in the periodic table.

By using the packages of python such as OpenCV and NumPy we are recognizing the hand. First, the system distinguishes the palm from the background and then it will recognize our finger. Then it will plot a graph from our thumb to the little finger and then it starts the process. Here we can able to assign the gestures statically or dynamically depends upon the machines and the user. As opposed to the static hand motion acknowledgment which chips away within reach shapes, the hand direction signal acknowledgment thinks about the consecutive information of hand direction and investigates the fleeting character of hand movement. Right now, will overview the 3D hand direction signal acknowledgment draws near. It takes the edge of the fingers and palm to make the developments. By utilizing neural systems, the plotted chart is coordinated with motions and the edge is noted then the gadget or framework reacts for that motion. The hand highlights are typically separated from the bouncing box of the fragmented hand. Both the RGB and profundity data can help in distinguishing the hand region. Here we accept the Region of Interest (ROI) has been resolved in this manner center around the discriminative highlights for the ROI. The hand district's highlights could be generally separated into three classifications: low-level highlights, center level highlights, significant level highlights.

The 8 fundamental motions, for example, right, left, up, down, upper right, lower left, upper left, and lower right, and it likewise has 16 complex signals. The fundamental motion developments can be stretched out by clients as they can assemble their own profiles of groupings. For example, the signal 'square' (motion number is 16) can be part into up, right, down, and left. Thus, a fundamental motion succession of U-R-D-L is produced. Right now, vision camera is utilized so the motion is perceived even in obscurity. What's more, this framework can have diverse vision frameworks, for example, Night vision, Infrared, Black and White, and so on, to perceive the hand developments in each circumstance.

## **4. Implementation of paper**

### **4.1 Modules**

#### **4.1.1 Camera**

It is a device that is embedded in the electronic system. It captures the image and recognizes the gestures by marking plots as a graph in the hand and then it recognizes the angle of the palm and the fingers.

#### **4.1.2 Recognition of an image**

After opening the camera, we have to make sure that it recognizes the image. For that, we are going to make the picture which is statically defined in the code and we need to make sure the camera recognizes it

#### **4.1.3 Hand Detection**

It is the main module in which dynamic recognition is going to occur. In this, the camera will recognize the hand by differentiating the palm from the background and it plots a graph over the fingers.

#### **4.1.4 Gesture Control**

At last, after the completion of the hand detection process, any gesture that is made either static or dynamic is recognized by the system and the resultant output is delivered to get control over the electronic devices

## 5. System Design

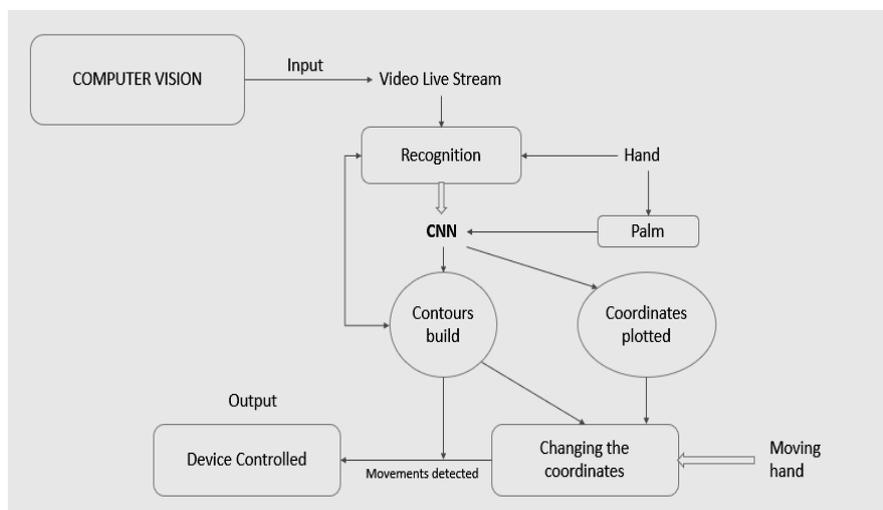


Figure 3. System Architecture

## 6. Experimental Results

We have done several experiments and testing on this application which provides a greater output. The main objective of this project is to reduce the supervision of digital devices by human rather it can be controlled using the hand gestures. After the contour process is finished, and the camera also recognizes that there is an image in front of it. It starts to calculate the defects and according to the number of defects calculated the gesture control is done. There are four types of defects:

**Defect “1”:** It means the camera recognizes the image which is present in front of it. Then it differentiates between the palm and the hand and it also differentiates between the environment and the hand.

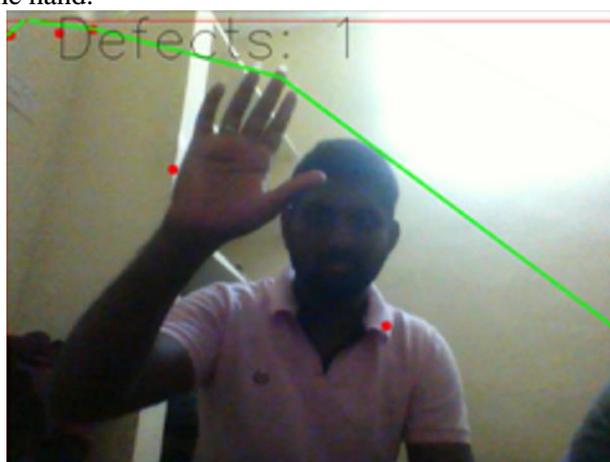
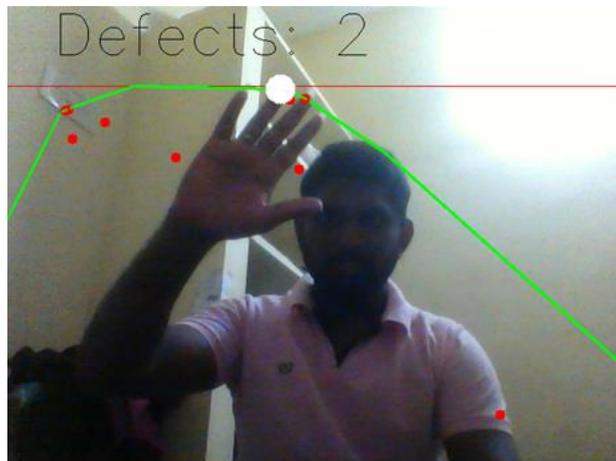


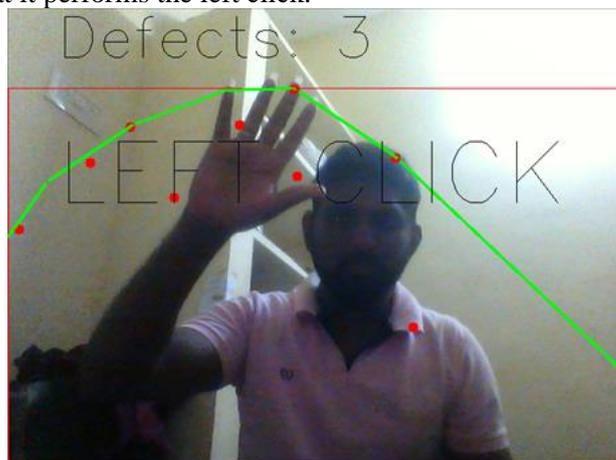
Figure 4. Defect 1- Recognition

**Defect “2”:** It is the process which tells us about tracking the movement of the hand by using a cursor for recognition. It also records the co-ordinates of the movement of the hand. This process is done in the background. The cursor which is present in the screen helps us in knowing the current position of the mouse pointer.



**Figure 5. Defect 2- Movements**

**Defect “3”:** It tells about the left click which is performed using the gesture which is detected by the camera. First it needs to recognize the gesture which is used for the left click and then it matches the gesture data with the data which is already present in the database. After that it performs the left click.



**Figure 6. Defect 3 – Left Click**

**Defect “4”:** It tells us about the right click which is performed using the gesture which is detected by the camera. First it needs to recognize the gesture which is used for the right click and then it matches the gesture data with the data which is already present in the database. After that it performs the Right click.

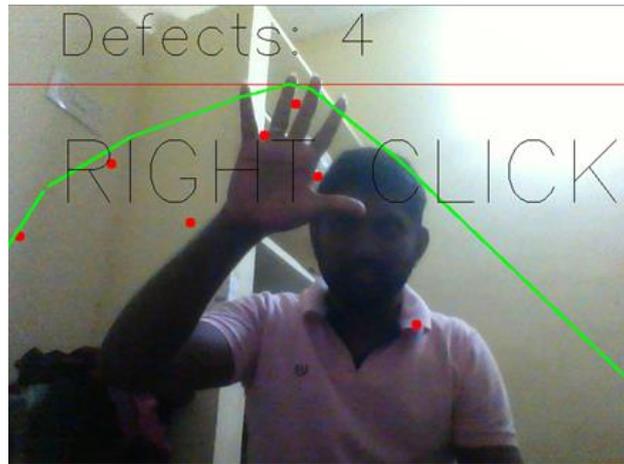


Figure 7. Defect 4 – Right Click

```
Command Prompt - mc.py
X: -17.075 Y: 0.0
X: 5.691666666666666 Y: 1.2
X: 2.045833333333333 Y: 0.0
X: -5.691666666666666 Y: -1.2
X: 0.5375 Y: 1.2
X: -5.691666666666666 Y: 0.0
X: 0.0 Y: 1.2
X: 0.0 Y: 0.0
X: 0.0 Y: 0.0
X: 2.045833333333333 Y: 0.0
X: 0.0 Y: 0.0
X: -2.045833333333333 Y: 0.0
X: -2.045833333333333 Y: 0.0
X: -0.5375 Y: 0.0
X: -2.045833333333333 Y: 0.0
X: 5.691666666666666 Y: 1.2
X: 5.691666666666666 Y: 0.0
X: -2.045833333333333 Y: -1.2
X: 0.0 Y: 0.0
X: 11.383333333333333 Y: 2.4
X: -19.920833333333334 Y: -1.2
X: -5.691666666666666 Y: 0.0
X: 2.045833333333333 Y: 1.2
X: 14.229166666666666 Y: 0.0
X: -11.383333333333333 Y: -1.2
X: -14.229166666666666 Y: -1.2
X: -5.691666666666666 Y: 0.0
X: 11.383333333333333 Y: 1.2
X: -17.075 Y: -1.2
```

Figure 8. Hand gesture co-ordinates



Figure 9. Hand gesture drawing curves

## 7. Conclusion

In this work, we investigated the proposed method for feasibility which classes human hand gestures using open CV with a CNN. After the implementation of CNN the hand's data is being plotted in the graph and the co-ordinates are being marked. Even the

contours are being marked and each and every time the hand moves the live co-ordinates is updated in the background and 4 sets of different cameras are being displayed at the time of execution. This project needs a High resolution camera for the more accuracy and perfect result to be delivered. Future implementations can be done in robotic industry and medical industry where a gesture can be used as a replacement for programming written for the control of the robots and moreover this project is implemented using just OPEN CV and CAMERA a simple implementation for the hand Gesture recognition and control.

## 8. References

- [1] YoungWookkim, Brian Toomajian "Hand Gesture Recognition using Micro-Doppler Signatures With Convolutional Neural Network" IEEE Open Access Journal, November 2016.
- [2] Danilo Avola, Marco Bernardi, Luigi Cinque, Gian Luca Foresti, Cristiano Massaroni "Exploiting Recurrent Neural Networks and Leap Motion Controller for the Recognition of Sign Language and Semaphoric Hand Gestures" IEEE, August 2015.
- [3] Feiyu Chen, HonghaoLv, Zhibo pang, Junhui Zhang, Yonghong Hou, Ying Gu, Huayong Yang, Geng Yang "WristCam: A Wearable Sensor for Hand Trajectory Gesture Recognition and Intelligent Human-Robot Interaction" IEEE, 2018.
- [4] Lun Ge, Hongjun Wang, Jishou Xing "Maintenance robot motion control based on Kinect gesture recognition", IET Journals, October 2018.
- [5] Kang Cheng, Ning Ye, Reza Malekian, Ruchuan Wang "In-Air Gesture Interaction: Real Time Hand Posture Recognition Using Passive RFID Tags", IEEE Open Access Journal, July 2018.
- [6] B.-W. Min, H.-S. Yoon, J. Soh, Y.-M. Yang, and T. Ejima, "Hand gesture recognition using hidden Markov models," in Proc. IEEE Int. Conf. Syst., Man, Cybern. vol. 5. Oct. 1997, pp. 4232–4235.
- [7] F.-S. Chen, C.-M. Fu, and C.-L. Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models," Image Vis. Comput., vol. 21, no. 8, pp. 745–758, Aug. 2003.
- [8] Y. Yao and C.-T. Li, "Hand gesture recognition and spotting in uncontrolled environments based on classifier weighting," in Proc. IEEE Int. Conf. Image Process., Sep. 2015, pp. 3082–3086.