International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

Hand Gesture Recognition System Using Convolutional Neural Networks

Rutuja J.[1], Aishwarya J.[2], Samarth S.[3] , Sulaxmi R.[4] , Mrunalinee P.[5]

[1]*Student, Dept. of Computer Engineering, RMD Sinhgad School Of Engineering, Pune, Maharashtra,India*
[2]*Student, Dept. of Computer Engineering, RMD Sinhgad School Of Engineering, Pune, Maharashtra, India*
[3]*Student, Dept. of Computer Engineering, RMD Sinhgad School Of Engineering, Pune, Maharashtra, India*
[4]*Student, Dept. of Computer Engineering, RMD Sinhgad School Of Engineering, Pune, Maharashtra, India*
[5] *Professor, Dept. of Computer Engineering, RMD Sinhgad School Of Engineering, Pune, Maharashtra, India*

Abstract - *Gesture based communication is perplexing to see yet the total language which includes the hand's development, outward appearances, and body stances. Gesture based communication is centre correspondence media to the general population which can't talk. It's anything but an all-inclusive language implies each nation has its very own gesture based communication. Each nation has its own punctuation for their communication through signing, word requests and articulation. The issues emerge when individuals endeavour to impart utilizing their language with the general population who are unconscious of this language sentence structure. To identify the sign utilizing hand motion and after that convert into printed or verbal structure which will can be comprehended by any individual for example perceives the outcome for each sign.*

***Keywords*: -** *Machine Learning, Character Detection, Speech Processing, CNN.*

1.INTRODUCTION

Gesture based communication is the essential language of the general population who are hard of hearing or deaf and furthermore utilized by them who can hear be that as it may, can't physically talk. It is a complex however total language which includes development of hands, facial articulations and stances of the body. Gesture based communication isn't all inclusive. Each nation has its own local gesture based communication. Each communication through signing has its very own standard of language structure, word orders what's more, articulation. The issue emerges when hard of hearing and unable to speak individuals attempt to convey utilizing this language with the general population who are uninformed of this language sentence structure. So it moves toward becoming important to build up a programmed and intuitive mediator to get them.

Research for communication via gestures acknowledgment was begun in the '90s. Hand signal related research can be isolated into two classifications. One depends on electromagnetic gloves and sensors which decides hand shape, developments and introduction of the hand. Be that as it may, it is expensive and not appropriate for down to earth use. Individuals need something increasingly common. Another depends on PC vision based signal acknowledgment, which includes picture preparing strategies. Thus, this class faces greater multifaceted nature.

A Real time sign recognition system is presented in this paper using Convolutional Neural Network. The model is designed to recognise all non-dynamic gestures of American Sign Language with bare hands of different hand shapes and skin colours which makes it complex for model to correctly recognise the gesture

1.1 AMERICAN SIGN LANGUAGE

American Sign Language (ASL) could be a complete, advanced language that employs signs created by moving the hands combined with facial expressions and postures of the body. It is the first language of the many North Americans United Nations agency are deaf and is one in all many communication choices utilized by those who are deaf or dumb.

The whole framework works in four stages

- Image Acquisition: This is the initial step or procedure of the central strides of advanced picture handling
- Image Enhancement

- Image Restoration
- Color Image Processing
- Wavelets and Multi Resolution Processing
  Compression

Fig 1. Alphabets of American Sign Language

2. BRIEF SURVEY

Numerous specialists are taking a shot close by signal acknowledgment utilizing visual investigation. In [1] an arrangement of equipment pose acknowledgment dependent on a SOM and Hebb crossover vector classifier. As the component vectors utilized in the proposed framework were invariant to area changes in info pictures, the acknowledgment was not hearty to area changes close by signs. In [2]this article multimodal hand motion identification and acknowledgment framework is displayed. Since both infrared and obvious range data is utilized, the proposed framework is more exact than IR-just and less power devouring than camera just frameworks. An epic WTA code based sensor combination calculation is additionally introduced for 1-D PIR sensor flag handling. The calculation melds the information originating from the distinctive PIR sensors in a programmed way to decide left-to-right, ideal to-left, upward, descending, clockwise and counter-clockwise movements. A Jaccard remove based measurement is utilized to group the hash codes of highlight vectors extricated from sensor signals. This [3] incorporates K convex hull for fingertip discovery, pixel division, erraticism, elongatedness of the article. The trial results demonstrate that K raised structure calculation gives increasingly exact fingertip identification. Picture outlines taken by portable camcorder interfaced with the PC are tried by our prepared ANN. In [4], the key Experimentation is done on American Sign Language. American Sign Language, viewed as one of the transcendent gesture based communication for hard of hearing networks in United States. ASL is generally utilized language all through the world. It utilizes just a single hand to show the signals and hence makes it simple for elucidation and comprehension. It includes around 6000 signals and other basic words. The normal words are appeared some particular motion or spelling with the assistance of 26 hand signals demonstrating 26 letter sets of ASL. In this work [5], they have acquainted up-with date the biggest dataset called Ego Gesture for the errand of egocentric signal acknowledgment with adequate size, variety and reality, to effectively prepare profound systems. The dataset is more mind boggling than any current datasets as our information is gathered from the most various scenes. Contrasted with motion order in sectioned information, the execution on motion location is a long way from fulfillment and has substantially more space to improve A robotized vision based American Communication via gestures (ASL) acknowledgment framework was displayed in[13] utilizing the HSV shading model to identify skin shading and edge location to identify hand shape. Another imperative improvement was the HCI framework for perceiving faces and hand motion from a camcorder exhibited in [14]. They consolidated head position and hand signal to control hardware. The situation of the eyes, mouth and face focus were recognized to decide head position. Two new strategies were presented in their paper: programmed motion zone division and introduction standardization of the hand motion. Their acknowledgment rate was 93.6%. The edge identification calculation and skin identification calculation were connected together in [15] utilizing MATLAB for a superior arrangement. The Canny edge identification calculation to detect focuses at which picture splendor changes strongly. They utilized ANN calculation for signal recognizable proof for quick computational capacity. Static hand signal acknowledgment examining three calculations named Convexity deformity, K ebb and flow and Part based

hand motion acknowledgment was created utilizing Microsoft Kinect sensors[16]. Microsoft's Kinect camera takes into account catching pseudo-3D picture called the profundity map which can without much of a stretch section the information picture and track the picture in 3D space. Be that as it may, this camera is very exorbitant. In 17] three strategies were investigated: K bend, Raised Hull, Curvature of Perimeter for fingertip identification.

3. PROPOSED WORK

The hard of hearing and unable to speak individuals endeavor to convey utilizing American Sign Language with the general population who are unconscious of this language sentence structure. So it ends up important to build up a programmed and intelligent mediator to get them.

The proposed system works by following the given steps to recognize the hand gestures and convert them into speech and text and vice-versa.

A. ***Convert ASL into Speech and Text***

1. Capture the image/video and give it as an input to the system.
2. Image Acquisition and Enhancement which is the most fundamental step of digital image processing.
3. All the relevant features are extracted and all the irrelevant and redundant features are ignored
    a. Local Orientation Histogram
    b. Local Brightness
    c. Binary Object Features
4. The appropriate character is detected from the given input.
5. The detected ASL characters are converted into speech and text.

B. Convert Speech or Text into ASL

1. Record Voice/Speech given by the user and give it as an input to the system
2. The received input will be processed and converted into its equivalent text
3. The equivalent text is detected
4. The text is then further converted into its corresponding ASL character

3.1 Algorithmic Strategy

- Convolutional Neural Network (CNN) works by consecutively modelling and small pieces of information and combining them deeper in work
- CNN follows Greedy approach
- In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used.
- CNN image classifications takes an input image, process it and classify it under certain categories (Eg., Dog, Cat, Tiger, Lion). Computers sees an input image as array of pixels and it depends on the image resolution. Based on the image resolution, it will see h x w x d(h = Height, w = Width, d = Dimension). Eg., An image of 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image.
- Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernals), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1. The below figure is a complete flow of CNN to process an input image and classifies the objects based on values.
- CNN works in following four stages
    1. Convolution Layer
    2. ReLU Layer
    3. Pooling Layer
    4. Fully Connected Layer

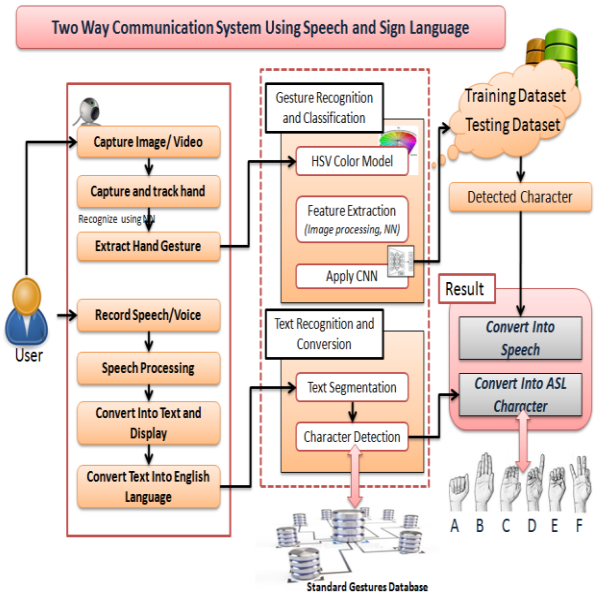

Fig. 1. System Architecture

3.1.1 Experimental Setup

1. Our data set has total 27456 entries and 784 labels for training, out of which 30% are used for validation
2. We are randomly splitting the dataset by subject into training 80% and testing 20%
3. We require processor Intel I3 and above
4. RAM :4GB, Memory: 80 GB
5. It will give better performance on GPU
6. Camera mounted at appropriate position for capturing the input image.
7. Android phone or any recorder for recording voice/speech
8. NetBeans/Eclipse IDE
9. Python for training and testing the dataset using CNN
10. Java used for frontend

3.1.2Dataset Characteristics

Dataset considered for this proposed system has a total of 27456 entries and 748 labels. All the images were stored as per their pixel values after they were converted into a grey scale image. We had an option of using the dataset as a Comma-separated values CSV or to generate the images from the pixel's values and then use them to train our dataset. We went with using the CSV file as it made the process of classification faster. All the missing values in the dataset were first handled and was made sure that there are no missing values in the dataset.

4. RESULT ANALYSIS

The proposed system was implemented using Intel Core i5-42100 CPU @1.70GHz speed and the code was written using Python programming language. Two-way communication system was not implemented in any of the previous systems i.e. a normal person could not communicate with the deaf/dump person.

We have designed a system using which it enables the deaf and dumb community to create recognition and also to give them a standard platform to communicate and express their opinions with every other individual. Using Convolutional Neural Networks (CNN) we have been able to get an accuracy of 95.6% which is higher than all the previously implemented systems.

4.1 REAL TIME SETUP

Real time setup is made possible by using and IDE that supports java and a camera module which captures images of the hand gesture and recognizes the equivalent letter and displays it on the screen. Using an Android device, the person on the other end can record speech or send text message which will then get converted into an equivalent hand gesture.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

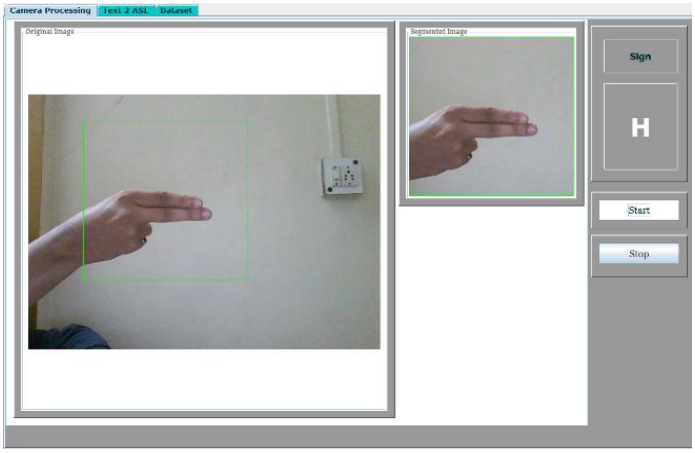

Fig. 2. Real Time Environment of American Sign Language

4.2 COMPARATIVE ANALYSIS OF DIFFERENT ALGORITHMS

A comparison analysis of CNN with KNN and ANN is shown in Table I.

Feature Extraction Algorithm	Testing Rate (%)	Validation Rate (%)	Average Recognition Rate (%)
CNN	99.5	99.9	99.6
KNN	85.9	86.1	85.5
ANN	78.3	77.6	77.9

TABLE I. COMPARISON ANALYSIS OF DIFFERENT ALGORITHMS(IN%)

The graph shows the error recognition rate between CNN and various other algorithms.

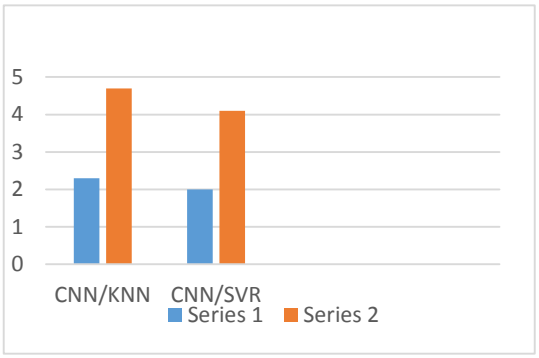

Fig. 3. Error Measure Comparision of CNN with other algorithms

The outcome shows that CNN outperforms any other neural network algorithm used. CNN gives an average Testing Rate of 99.5%, Validation Rate of 99.9%, and an Average Recognition rate of 99.6%. Using CNN in the proposed system gives better accuracy of 99.6%.

4.3 REAL TIME PERFORMANCE ANALYSIS

To do real time hand gesture recognition performance each sign is taken from fiver different people who have different skin tone, features and a white background is maintained to give maximum efficiency. Features are extracted from the images and tested in the previously trained dataset. The number of correct responses out of 10 times of testing each sign is show in Table II.

Class	Numbers of Correct Responses (Out of 10)	Recognition Rate (%)
A	10	100
B	10	100
C	8	80
D	10	100
E	9	90
F	8	80
G	10	100
H	8	80
I	9	90
K	10	100
L	10	100
M	9	90
N	9	90
O	10	100
P	9	90
Q	9	90
R	9	90
S	10	100
T	10	100
U	10	100
V	10	100
W	9	90
X	10	100
Z	10	100

TABLE II. REAL TIME HAND GESTURE RECOGNITION PERFORMANCE ANALYSIS
REAL TIME HAND GESTURE RECOGNITION PERFORMANCE ANALYSIS

Average Recognition rate is calculated as follows:

*Average Recognition rate = (No. of correct Response) / (No. of Total Samples) *100%*

Average Recognition rate of the proposed system is 94.32%.

5.DISCUSSION

- The proposed system is able to detect the hand gestures with more accuracy when compared with the existing system
- It takes into consideration the detailed features including the shape, size, color of bare hand
- Two way communication is possible in the proposed system which is very beneficial and cost effective
- The processing time required for the proposed system is less compared to the existing system
- The proposed model successfully handles two way communication by converting sign language into text, speech and vice versa

This work study and examined how to outwardly perceive every single static motion of American Sign Language (ASL) with uncovered hand. Diverse clients have diverse hand shapes and skin hues, making it progressively troublesome for the framework to perceive a signal. Gesture based communication Recognition is fundamental for the less privileged individuals to speak with other individuals.

6.CONCLUSION

The principal goal of this project is to determine gesture recognition that might enable the deaf to converse with the hearing people. The features extraction is one of the important task such as different gestures should result in different, good discriminable features.

We use CNN algorithm trained dataset to detect the character from the gesture images. With the help of these features and trained dataset we can recognizes ASL alphabets and numbers with accuracy in real time

The proposed system can be made available in multi languages making it more reliable and efficient. It could be made available entirely on the mobile devices which will help in the making the system handier and portable in the near future.

7.REFERENCES

[1] Hiroomi Hikawa, Keishi Kaida, “Novel FPGA Implementation of Hand Sign Recognition System with SOM-Hebb Classifier” 2013 IEEE.

[2] F. Erden and A. E. Çetin, "Hand gesture based remote control system using infrared sensors and a camera," IEEE Trans. Consum. Electron., vol. 60, no. 4, pp. 675-680, 2014.

[3] Md. Mohiminul Islam,Sarah Siddiqua, and Jawata Afnan“Real Time Hand Gesture Recognition Using Different Algorithms Based on American Sign Language” 2017 IEEE.

[4]Dipali Naglot, Milind Kulkarni, “Real time sign language recognition using the leap motion controller.”

[5] Yifan Zhang*, Congqi Cao*, Jian Cheng, and Hanqing Lu “EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition” 2018 IEEE

[6] S. Kim, G. Park, S. Yim, S. Choi and S. Choi, "Gesture-recognizing hand-held interface with vibrotactile feedback for 3D interaction," IEEE Trans. Consum. Electron., vol. 55, no. 3, pp. 1169-1177, 2009.

[7] S. S. Rautaray, and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," Artificial Intelligence Review, vol. 43, no. 1, pp. 1-54, 2015.

[8] D. W. Lee, J. M. Lim, J. Sunwoo, I. Y. Cho and C. H. Lee, "Actual remote control: a universal remote control using hand motions on a virtual menu," IEEE Trans. Consum. Electron., vol. 55, no. 3, pp. 1439-1446, 2009.

[9] D. Lee and Y. Park, "Vision-based remote control system by motion detection and open finger counting," IEEE Trans. Consum. Electron., vol. 55, no. 4, pp. 2308-2313, 2009.

[10] S.H. Lee, M.K. Sohn, D.J. Kim, B. Kim, and H. Kim, "Smart TV interaction system using face and hand gesture recognition," in Proc. ICCE, Las Vegas, NV, 2013, pp. 173-174.

[11] S. Jeong, J. Jin, T. Song, K. Kwon and J. W. Jeon, "Single-camera dedicated television control system using gesture drawing," IEEE Trans. Consum. Electron., vol. 58, no. 4, pp. 1129-1137, 2012.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," IEEE Trans. on Pattern Anal. Mach. Intell., vol. 38, no. 1, pp. 142-158, 2016.

[13] Sharmila Konwar, Sagarika Borah and Dr. T. Tuithung, "An American sign language detection system using HSV color model and edge detection", International Conference on Communication and Signal Processing, IEEE, April 3-5, 2014, India

[14] Yo-Jen Tu, Chung-Chieh Kao, Huei-Yung Lin,"Human Computer Interaction Using Face and Gesture Recognition", Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific, IEEE, Kaohsiung.

[15] Shweta. K. Yewale and Pankaj. K. bharne, "Hand gesture recognition system based on artificial neural network", Emerging Trends in Networks and Computer Communications (ETNCC), IEEE, 22-24 April, 2011.

[16] Marek Vanco, Ivan Minarik and Gregor Rozinaj, "Evaluation of static hand gesture recognition", International Conference on Signal and Image Processing (IWSSIP), IEEE, 12-15 May, 2014.

[17] Javeria Farooq and Muhaddisa Barat Ali, "Real time hand gesture recognition for computer interaction", International Conference on Robotics and Emerging Allied Technologies in Engineering (ICREATE), 22-24 April ,2014.

[18]Peijun Bao, Ana I. Maqueda, Carlos R. del-Blanco, and Narciso García, "Tiny Hand Gesture Recognition without Localization via a Deep Convolutional Network", IEEE Transactions on Consumer Electronics, Vol. 63, No. 3, August 2017.