Convolutional Neural Network-Based Hand Gesture Recognizer

Mrs. K. Kirubananthavalli[1]
Assistant Professor /CSE,
Unnamalai Institute of Technology, Kovilpatti,
Tamil Nadu,India
kirubananthavalli@uitkovilpatti.ac.in

Mrs. C. Ajitha[2],
Assistant Professor /CSE,
Unnamalai Institute of Technology, Kovilpatti,
Tamil Nadu,India
cajitha5@gmail.com

Abstract - **It's difficult to make sense of gesture-based communication since it encompasses so much more than just the shape of the hand. Those who are unable to speak may nevertheless communicate with the rest of society via the use of gestures. It's not a universal language since every culture has its own unique system of hand gestures. When it comes to using sign language, word requests, and spoken language, every culture has its own unique punctuation. Problems arise when people try to communicate using their language with others who are unfamiliar with the sentence structure of that language. The goal is to recognise each sign via its associated hand motion and then to translate that recognition into a written or spoken format that can be understood by everyone.**

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

1. INTRODUCTION

People who are deaf, hard of hearing, or who otherwise lack the physical ability to speak rely on gestures as their primary means of communication. It's an intricate yet comprehensive language, with its own unique articulations of the hands, face, and body. The use of gestures in communicating is limited. Every culture has its own unique system of hand gestures for communicating. Each form of signed communication has its own rules for how words should be ordered and how they should be spoken. The problem arises when deaf or speechless people try to communicate with the general public, who are not familiar with the sentence structure of this language. So it progresses towards becoming crucial to develop a computerised and empathetic intermediary in order to get them.

The study of gesture recognition as a means of communication began in the 1990s. There are two distinct subfields within the study of hand signals. The form, progression, and debut of the hand are all determined by electromagnetic gloves and sensors. However, its high price and lack of practicality make it impractical for everyday life. There is a growing need for a universal necessity among people. One alternative is a technique based on computer vision for recognising signals. This technique makes use of image preprocessing methods. Therefore, this group confronts a more complex set of challenges.

In this study, we provide a Convolutional Neural Network-based system for real-time sign identification. Due of the wide variety of hand shapes and skin tones used in ASL, it is challenging for the model to reliably detect static movements performed with bare hands.

AMERICAN SIGN LANGUAGE

It's possible that American Sign Language (ASL) is a fully developed, sophisticated language in which speakers communicate using a combination of hand movements, facial emotions, and body language. As the native tongue of many deaf North Americans, American Sign Language is only one of several modes of communication available to the hearing impaired.

The whole framework works in four stages

- In sophisticated photo processing, the first, most crucial stage is image acquisition.
- Enhancement of Images
- Image Restoration
- Color Image Processing
- Wavelets and Multi Resolution

- ProcessingCompression

Fig 1. Alphabets of American Sign Language

2. BRIEF SURVEY

Many researchers are trying to perfect visual analysis for signal recognition. An apparatus posture recognition system based on a SOM and Hebb crossover vector classifier is shown in [1]. The suggested framework's insensitivity to spatial variations lack information images meant that the recognition was weak around new indicators. This article shows a multimodal framework for recognising and identifying hand motions. The proposed framework makes use of both infrared and visible range data, making it more precise than IR-only and less energy-hungry than camera-only approaches. For the purpose of managing the flags of 1-D PIR sensors, we also develop a WTA code-based epic sensor combination computation. The formula combines data from the several PIR sensors in a predetermined fashion to establish whether a person is moving from left to right, right to left, up, down, clockwise, or anticlockwise. The hash codes of extracted highlight vectors from sensor data are grouped using a Jaccard removal-based assessment. This [3] uses K convex hull to account for the fingers, the pixels, the randomness, and the length of the article. The results of the trials show that using K raised structure computation improves the accuracy of fingerprint recognition. Picture

Our trained ANN is tested on outlines captured with a handheld camera connected to a personal computer. In [4], American Sign Language serves as the primary experimental subject. When it comes to networks of people who are deaf or hard of hearing in the United States, American Sign Language is often regarded as a universal language. American Sign Language (ASL) is widely used all over the globe. Because it just takes one hand to demonstrate the signals, it's easy to clarify and understand. It has a dictionary of over 6,000 signals and other common terms. With the use of 26 hand gestures illustrating 26 letter sets of ASL, commonly used words seem to have a special motion or spelling. In this study [5], the largest dataset for the task of egocentric signal recognition, dubbed Ego Gesture, has been brought up to date with sufficient size, diversity, and realism to successfully build deep systems. Since we collected this data from so many different settings, it is unlike any other dataset currently available. The execution on motion location is far from perfect and has a lot of room to grow compared to motion order in sectioned data. Using the HSV shading model to determine skin shading and edge location to determine hand form, [13] demonstrated a robot-based vision-based ASL acknowledgement framework. Important progress was also made with the introduction of the HCI framework shown in [14] for recognizing faces and hand activity from a camcorder. They merged the use of head motion and hand gestures to manage the machines. The head position was determined by analyzing the relative positions of the eyes, mouth, and facial focus. In their work, they introduced two novel strategies: motion zone division programming and hand motion standardization. Their rate of recognition was 96. In [15], an improved system was developed by linking the edge identification computation with the skin identification calculation using MATLAB. Using the Canny edge detection formula, we can pinpoint the locations where the brightness of an image abruptly shifts. They used ANN computation for signal recognition evidence because of its high computational speed.

Microsoft Kinect sensors were used to develop a system that could recognize hand motion [16]. The depth map, captured by Microsoft's Kinect sensor, is a pseudo-3D image that can be used to easily segment the data picture and follow it as it moves across 3D space. However, the cost of this camera is prohibitive for most people. Three

approaches were studied in 17]: Fingertip identification thanks to a K bend, raised hull, and curved perimeter.

3. PROPOSED WORK

People who are deaf or hard of hearing often try to communicate with others who are not familiar with American Sign Language by using hand gestures and facial expressions. As a result, it becomes crucial to develop a programmed and intelligent mediator to acquire them.

The suggested system recognises hand gestures and converts them into voice and text, as well as the other way around, by following the stated procedures.

A. Convert ASL into Speech and Text

1. Capture the image/video and give it as an inputto 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 thegiven input.
5. The detected ASL characters are converted intospeech and text.

B. Convert Speech or Text into ASL

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

Algorithmic Strategy

- A Convolutional Neural Network (CNN) achieves its results via the sequential modelling and combination of tiny bits of information.
- CNN's strategy is a greedy one
- One of the most common types of neural networks used for image identification and classification is the convolutional neural network (ConvNets or CNNs). CNNs are often used for object identification, facial recognition, and other similar tasks.
- CNN image classifications work by taking an input picture and determining what it is (for example, a dog, cat, tiger, or lion) and where it belongs. A computer reads a picture as a series of pixels, the size of which is determined by the image resolution. It will perceive h x w x d (h = Height, w = Width, d = Dimension) depending on the picture resolution. As an illustration, consider the following examples: a 6x6x3 RGB picture and a 4x4x1 grayscale image.
- To identify an item with probabilistic values between 0 and 1, deep learning CNN models train and test input images by running them through a sequence of convolution layers with filters (Kernals), Pooling, fully connected layers (FC), and applying a Softmax function. CNN's whole procedure for analysing an input picture and labelling objects based on values is shown in the following diagram.
- CNN's four-step process is as follows:
    1. Convolution Layer
    2. ReLU Layer
    3. Pooling Layer
    4. Fully Connected Layer

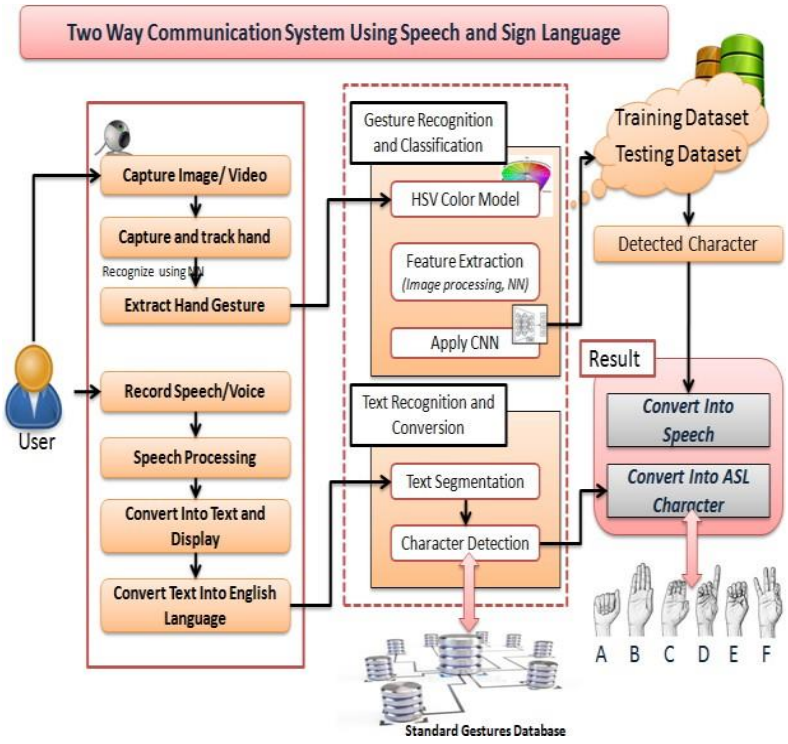

Fig. 2. System Architecture

Experimental Setup

1. First, our dataset consists of 27456 training records and 784 validation labels.
2. Second, we are using a randomised subject-based split to allocate 80% of the dataset to training and 20% to testing.
3. Third, an Intel I3 CPU or above is required.
4. Memory: 4GB; RAM: 4GB; Total: 80GB
5. Five, it'll improve GPU performance.
6. The camera is positioned correctly for taking the input picture.
7. 7. A voice or speech recorder, such as an Android phone
8. NetBeans/Eclipse, the IDEs No. 8
9. 9. Python for utilising Convolutional Neural Networks to train and evaluate a dataset
10. Front-end coded in Java

Dataset Characteristics

In all, there are 27456 observations and 748 labels in the dataset used to generate this algorithm. After being converted to greyscale, each image's pixel values were saved. We could have trained our dataset using the pixel values alone, or we could have used a Comma-separated values (CSV) representation of the data. The categorization procedure ran more quickly using the CSV file, so we opted with it. It was first ensured that there were no missing values in the dataset, and then all the missing values.

4. RESULT ANALYSIS

The suggested system was built using Python code and an Intel Core i5-42100 processor running at 1.70 GHz. No prior system had a two-way communication mechanism; therefore, a hearing person could not converse with a deaf or mute person.

We've built a system that makes it easier for the deaf and dumb to get their names out there and share their thoughts with the rest of society on a level playing field. The 95.6% accuracy we've achieved using Convolutional Neural Networks (CNN) is better than everything that's come before it.

Real Time Setup

Using an IDE that is compatible with Java and a camera module that collects photos of the hand motion, the corresponding letter may be recognized and shown on the screen in real time. The other party may use an Android smartphone to transmit a voice memo or text message that will be translated into a corresponding hand gesture.

When the copyediting is done, the document may be formatted according to its template. Use the Save As option to create a copy of the template file, and give your paper a name that adheres to the guidelines set out by your conference. Select the whole document's contents, then drag and drop your prepared text file into the new document. The formatting window for your manuscript is located in the drop-down menu to the left of the MS Word toolbar.

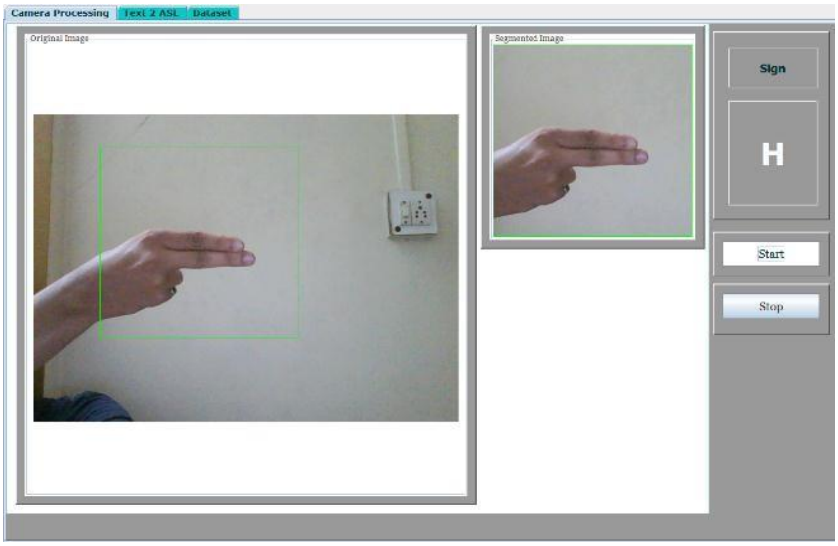

Fig. 2. Real Time Environment of American SignLanguage

COMPARATIVE ANALYSIS OF DIFFERENTALGORITHMS

A comparison analysis of CNN with KNN andANN 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 chart compares CNN to other algorithms' recognition mistake rates.

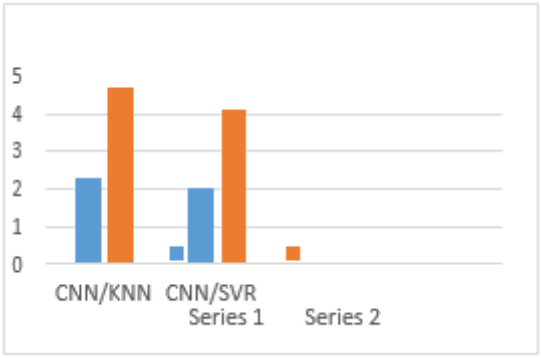

Fig. 3. Error Measure Comparison of CNN with otheralgorithms

CNN beats all neural network algorithms. CNN has a 99.5% Testing Rate, 99.9% Validation Rate, and 99.6% Recognition Rate. CNN provides the suggested technique 99.6% accuracy.

REAL TIME PERFORMANCE ANALYSIS

TReal-time hand gesture detection requires five persons with diverse skin tones and characteristics and a white backdrop. The previously trained dataset is tested using picture features. Number of Table II shows accurate replies out of 10 tests for each indication.

Class	Numbers of Correct Responses (Out of 10)	Recognitin 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.

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 suggested method detects hand motions more accurately than the current system since it considers the form, size, and colour of bare hands.

Two-way communication in the suggested system is useful and cost-effective.

The suggested technology converts sign language into text, voice, and vice versa in less time than the current method.

This research studied how to externally interpret every static motion of American Sign Language (ASL) using exposed hands. Diverse hand shapes and skin colours make it harder for the system to detect a signal. Gestural communication The underprivileged need recognition to communicate.

6. CONCLUSION

This research seeks gesture recognition to help deaf individuals communicate with hearing people. Different movements should produce distinct, distinguishable characteristics, making feature extraction crucial.

CNN algorithm training dataset detects character from gesture photos. These characteristics and training dataset enable real-time ASL alphabet and number recognition.

The suggested multilingual system is more dependable and efficient. It might be totally on mobile devices, making the system more portable in the future.

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]. 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]. 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 ccurate 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