International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 169-173
Published Online June 2023 in IJEAST (http://www.ijeast.com)

AMERICAN SIGN LANGUAGE (ASL) RECOGNITION USING OPENCV

Mehtab Ansari
Dept. of Elect. and Comm. Engg.
HMRITM
New Delhi, India

Ms. Neetu Raj Bharti
Assistant Professor
Dept. of Elect. and Comm. Engg.
HMRITM
New Delhi, India

Abstract: **One of the biggest challenges for thosewho are deaf or dumb is the lack of access to communication. Here, we'll createa OpenCV based system for converting American sign language to text. Computer vision model for the dumb and deaf are a revolutionary system that has the potential to improve the quality of life for those whoare affected by hearing or speech impairments. The computer vision model will detect hand movements and translate them into text in real-time, allowing deaf and dumb individuals to communicate with others more effectively. Traditionalforms of communication such as writing can be difficult and time-consuming, and they also rely on the presence of an interpreter or someone who can read orwrite. Computer vision supply a more efficient and immediate solution to this problem. The CV model are designed to detect hand movements and translate them into text, depending on the user's preferences. This allows deaf and dumb individuals to communicate with others in real-time, without the need for aninterpreter or someone who can read orwrite. The system is configured to display the text or speech on a nearby device, so that others can see and understand the message as well.**

Keywords: **Deaf communication, Sign language interpretation, Cognitive Processing, Computer Vision Model, Deep Learning, Language Processing.**

I. INTRODUCTION

"The capacity to express himself by reacting to the events occurring in his surroundings is one of nature's most priceless gifts to the living species."
India is home to 2.4 million Deaf and Dumb people, or 20% of the world's Deaf and Dumb population. These person lacks most of traits that most people would assert.

In India, individuals who are dumb (unable to speak) and deaf (unable to hear) face many challenges in accessing education, employment, and other opportunities. There are a limited number of schools and organizations that cater to their needs, and societal attitudes towards people with disabilities can be negative. Additionally, many deaf and dumb individuals may not have access to technology or resources that could help them in overcoming these barriers. However, there are also several organizations and individuals working to improve the lives of people who are dumb and deaf in India, such as working to improve access to education and employment opportunities, promoting awareness and acceptance of people with disabilities, and supplying support and resources for individuals and families affected by deafness and speech impairments[1].

Motivation
The misconception that being deaf or hearing challenged is a problem that must be rectified either by surgery or by technological, clinical, or healthcare equipments has continuously afflicted the deaf community, which is an entirely separate society or demographic of people. The term of "culture" states that it is the intellectual development brought about by adequate education or training. We are also able to observe how the deaf community has developed into just this. This inspired us to create an interpreter for sign language that can support their development and establishing objectives.

Available Methods
Both of these two methods is used to identify language through sign gestures.

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 169-173
Published Online June 2023 in IJEAST (http://www.ijeast.com)

- **Gloves method**

It implies that using pair of Smart Glove by the hand of hearing individual during the capture of hand movements empowers them to speak with others.

- **Computer Vision method**

A computer vision model trained different sets of multiple sign language gestures, captures image from webcam and put into the system to process and return corresponding alphbate character, digit or words. classified into static gestures (2d-images) and dynamic (real time live capture of the gestures).
Gloves method have an accuracy of over 90%, but wearing pair of bulky gloves will be uncomfortable, and would not be easy carrying around with microcontroller, electrical circuitary and power supply. And would not be accessible in bed weather.
Whereas Computer Vision model after good appropriate training, is very much scalabe. And later can be transformed mobile devices.

Objective

In this project we will go with a vision based method and attempt to design an interface that will facilitate communication between those who have disabilities and those without disabilities(normal person). We utilized the CNN model for this. We first deployed the model with the interface after training it with photos of gestures made with the hands. Then the model recognizes the disabled person's gesture, transforms it into text, and displays it for both ways of communication.
Our system is able to detecte ASL signs of A-Z alphabets, 0 to 9 digits and below listed 15 words.

Baby	Brother	Dont_like
Friend	Help	House
Like	Love	Make
More	Name	No
Pay	Play	Stop
With	Yes	Nothing

II. LITERATURE REVIEW

- "Deep Convolutional Neural Networks for Sign Language Recognition" G. Anantha Rao, Guntur (DT)

Extraction of complex head and hand movements along with their constantly changing shapes for recognition of sign language is considered a difficult problem in computer vision.

- K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition.

2D CNN are used to extract spatial features of input images while RNN are employed to capture the long term temporal dependencies among input video frames. VGG16 pretrained on ImageNet to extract spatial features and then feed the extracted features to a stacked GRU.

- J. Carreira and A. Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. CVPR, 2017.

3D convolutional networks are used which are able to establish not only the holistic representation of each frame but also the temporal relationship between frames in a hierarchical fashion. Inflate 2D filters of the Inception network trained on ImageNet, thus obtaining well-initialized 3D filters.

- Recognizing American Sign Language Gestures from within Continuous Videos

The fact that the temporal limits of an action are not always obvious in motions in continuous videos is one of the major difficulties. This paper detects their temporal locations from within continuous videos, by collecting an ASL dataset that has been annotated with the time-intervals for each ASL word.

III. THEORETICAL BACKGROUND

i. OpenCV

The vast free and open-source library known as OpenCV is used for machine learning, computer vision, and image processing. It currently plays a significant part in real-time operation, which is crucial in modern systems. Utilizing it, one can analyze pictures and videos to find faces, objects, and even human handwriting. Python is able to process the structure of the OpenCV array for analysis when combined with a variety of modules, such as NumPy. We use vector space and apply mathematical operations to these features to identify visual patterns and their various features.
OpenCV's initial release was 1.0. OpenCV is free for both educational and business purposes because it is distributed under a BSD license. It supports Microsoft Windows, MacOS, Linux, and Android along with implementations in programming languages such as Python, C, C++ and Java. Applications that operate in real time for improved processing efficiency were the primary consideration when OpenCV was developed. Everything is written in C/C++ that has been optimized to take benefit from multi-core processors[2].

ii. Convolutional Neural Network

An advanced form of artificial neural networks known as convolutional neural network, or CNN for short, substitutes the mathematical operation known as convolution for generic multiplication of matrices in at least one of its

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 169-173
Published Online June 2023 in IJEAST (http://www.ijeast.com)

layers. They are employed in image processing and recognition since they were created primarily to process information from pixels.

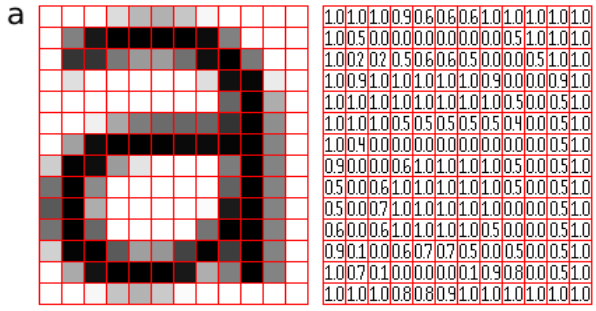

Fig 1; layers of CNN

An input layer, hidden layers, and an output layer make up a convolutional neural network (CNN). Any middle layers in a feed-forward neural network are referred to as hidden layers since the activation function and final convolution hide their inputs and outputs. The hidden layers in a convolutional neural network (CNN) contain convolutional layers. Typically, this involves adding a layer that does a dot product of the input matrix of the layer and the convolution kernel. The activation mechanism for this product, which is often the Frobenius, the inner product, is frequently ReLU. The convolution procedure develops a feature map as the convolution kernel moves across the given input matrices for the layer, adding to the input of the following layer. After this layers like normalization layers, pooling layers, fully connected [3].

iii. Transfer Learning: Inception V3 model

Tranfer Learning is processor of training a computer vision model with the help of lagre per-trained model.

The third version of Google's Deep Learning Convolutional Architectures is called Inception V3. A dataset of one thousand classes from the original ImageNet dataset, which was trained using more than 1 million training photos, was used to train Inception V3[4].

IV. METHODOLOGY

i. Image Acquisition

The acquisition of images is the initial step in our automatic image analysis system. These images are used for model testing and training. This stage is crucial for all other aspects of the system; as a result, if a picture is not taken successfully, the system's remained components may not work as intended or the findings may not be accurate. For improved outcomes at this level, the picture system first needs the image enlarged. For model training, images obtained from kaggle dataset and tested with live feed from webcam[5].

ii. Pre-Processing Technique

Image pre-processing is the initial step to identify the sign. We can use several operators on an image once it enters a computer system and can be represented using the some color space to enhance its representation. Multiple steps are performed to make the image suitable for the feature extraction process. The abnormalities the preprocessed and identified elements in the supplied image for the following purpose:

- Prevent uneven lighting
- Boost contrast inside the image and remove background pixels
- Prevent image noise

In this report work, the techniques used for the preprocessing phase are:

- Image Resizing
- Color transformation (RGB to Gray) and

(a) **Image Resizing**

Remapping might happen when we are adjusting for distortion from lenses or rotating an image, whereas picture scaling is required when it is necessary an increase or decrease the total amount of pixels. When you zoom in on an image, more pixels are used, giving us the ability to see more detail.

(b) Color Transformation

There are numerous ways (Colour Spaces) for a software system to mathematically represent colors once a hardware device has received an image. HSV (Hue, Saturation, Value) and RGB (Red, Green, Blue) are two of the most well-known color spaces. Utilizing an HSV color space has a number of benefits, one of which is the ability to make our system's lighting invariant by using only the HS components.

The photos are taken in RGB (Red, Green, and Blue) format. Grayscale is a spectrum of apparent-colorless hues of gray. Black is the darkest color imaginable. White is the lightest color imaginable. Equal intensity levels of the three major hues (red, green, and blue) for transmission light and reflected light are used to represent intermediate shades of gray.

The number corresponding to the brightness values of the primary colors is precisely proportional to each pixel in a red, green, and blue (RGB) image. White is represented by R = G = B = 255 (decimal) or R = G = B = 11111111, while black is denoted by R = G = B = 00000000. The term "8-bit grayscale" refers to the imaging technique used since there are only 8 bits in the binary code that represents of the gray level.

iii. Noise Removal

To obtain reliable results, high-quality, noise-free photos are required. Our model could produce inaccurate findings as a

International Journal of Engineering Applied Sciences and Technology, 2023
Vol. 8, Issue 02, ISSN No. 2455-2143, Pages 169-173
Published Online June 2023 in IJEAST (http://www.ijeast.com)

result of noisy images. As a result, image de-noise is required. We can do this by getting an image and replacing distorted pixels with an average of its neighbours. A decent picture denoise model must be able to eliminate noise while maintaining edges. In the past, linear models used to be applied. We can apply a median filter to the image to reduce noise. The task of image smoothing is carried out by a median filter.

iv. Feature Extraction

In order to process images accurately, the image must constantly be of extremely high quality. Practically speaking, this is challenging. Obtain images of low or medium quality for a variety of reasons. Consequently, it becomes necessary to raise their caliber. to enhance an image's quality using an algorithm for image enhancement. By emphasizing factors like contrast and brightness modification, this algorithm improves the image.

There are four primary sorts of features once an image has been pre-processed.

- **Global Features**: A single feature vector emerges from the feature extractor after the entire image has been analyzed as one. A histogram of put away pixel values is an easy illustration of a global feature.
- **Grid or Block-Based Features**: Features are taken from every one of the image's various blocks once the image has been divided into blocks. Dense SIFT (Scale Invariant Feature Transform) is one of the primary methods used to extract features from image blocks. Machine learning models are frequently trained utilizing these kind of features.
- **Region-Based Features**: A feature is extracted from all of the regions after the image has been divided into various segments (for example, using thresholding or K-Means Clustering and connecting the segments using Connected Components). Techniques for describing regions and boundaries, such as Moments and Chain Codes, can be used to extract features.
- **Local Features**: The image has several discrete interest spots, and by analyzing the pixels next to the interest points, features are retrieved from the image. Corners and blobs are two of the primary interest spots that may be recovered from a picture; these can be extracted utilizing techniques like the Harris & Stephens Detection and Laplacian of Gaussians. Finally, features can be obtained from the identified interest points using methods like the Scale Invariant Feature Transform. In order to match photos and create a panorama/3D reconstruction or to get images from a database, local features are frequently used.

V. CLASSIFICATION

Figure illustrates the general flow of the suggested approach.

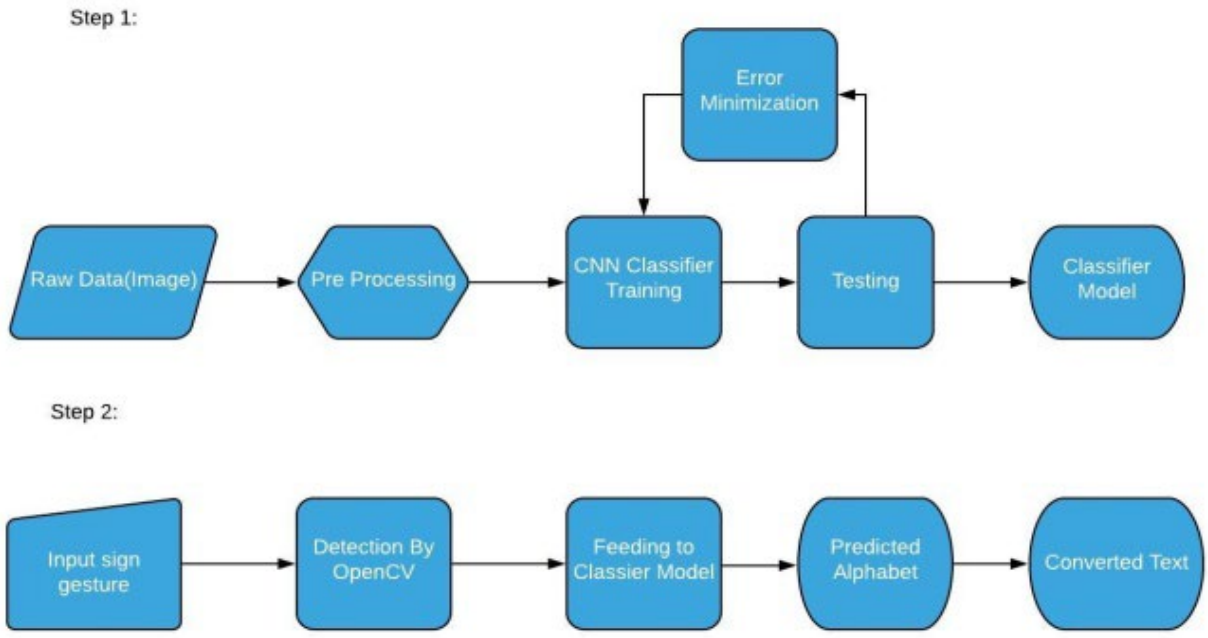

Fig 2; Flow Chart of processing technique

VI. CONCLUSION & FUTURE SCOPE

The computer vision system will offer an interface that allows for simple sign language recognition for communication with hearing-impaired people. As technology continues to advance, it is expected that sign language recognition system will become more smarter, lightweight and portable.

The system is applicable not just in a private household but also in a public setting. The model limited to alphabets, digits and 15 words, after training for more and more words, People who are dumb and deaf will find it very helpful.

Future scopes of improvement in present methodologies are:

- In the upcoming updates, we'll work to improve the system's accuracy and stability.
- Training for more and more words.
- To add other features like convert the text to speech.

VII. REFERENCES

[1]. Pushpendra Jha, Hand motion acknowledgment application for physically impaired individuals," International Journal of Science and Research, vol. 3, no. 8, pp. 765– 769, 2014.
[2]. opencv.org
[3]. wikipedia.org/Convolutional_neural_network
[4]. inception-v3-model, kaggle
[5]. Sign Language MNIST, kaggle
[6]. Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., & Lee, R. G. (2000). The syntax of American Sign Language: Functional categories and hierarchical structure. MIT Press.
[7]. Supalla, T. (2011). Language, culture, and community in deaf America. In P. S. Perneger (Ed.), Deaf around the world: The impact of language (pp. 3-25). Oxford University Press.
[8]. Emmorey, K. (2002). Language, cognition, and the brain: Insights from sign language research. Mahwah, NJ: Lawrence Erlbaum Associates.
[9]. Baker-Shenk, C., & Cokely, D. (1993). American Sign Language: A teacher's resource text on curriculum, methods, and evaluation. Gallaudet University Press.
[10]. Wilcox, P., & Shaffer, B. (2005). The acquisition of sign language by deaf children. In E. Hoff & M. Shatz (Eds.), Blackwell handbook of language development (pp. 580-607). Blackwell Publishing.