IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Sign Language to Text and Speech Conversion Using Image Processing and Machine Learning

[1]N. Harini, [2]B. Geetha Vani, [3]C. Kiranmai, [4]K. Hemavathi, [5]S. Shayastha

[1]Asst. Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Electronics and Communication Engineering,
[1]G. Narayanamma Institute of Technology and Sciences(for Women), Hyderabad, India

Abstract **:** Communication is crucial for understanding one another's feelings and thoughts. However, those who are mute or unable to hear or speak are unable to communicate their ideas or comprehend what others are saying. They can only understand sign language, and since many of us do not know sign language, communication has become quite difficult. With the advancement of technology, every issue may be solved. There is a way for deaf and dumb persons to communicate their thoughts and comprehend those of others through the use of machine learning and image processing. We are developing a human computer interface system in this project that accurately recognizes the language of the deaf and the dumb. As the hand plays a crucial role in communication, the vision-based hand gesture recognition system has been explored in this study. Various strategies for hand tracking, segmentation, feature extraction, and classification are mentioned. Webcams are used to record the images, which are then analyzed using image processing techniques like the OTSU method. The captured gestures are then classified using a linear classification algorithm. The gestures that were recorded are now saved in files with 120 copies of each gesture. Histograms are used to record image gesture.

Keywords: **Sign Language Recognition, Hand Gesture Recognition, Image Processing, OTSU Method, Feature Detection, Feature Extraction, Naïve Bayes Classifier, Machine Learning.**

I. INTRODUCTION

Deaf and dumb persons can communicate via sign language, a nonverbal form of communication. Regional sign languages include ASL (American sign language), ISL (Indian sign language), and BSL.(British Sign Language). However, as English is a widely spoken language in many areas, ASL is seen as a widespread sign language. An image-based, human-computer interface is created utilising machine learning and image processing. Using machine learning, we store all images of sign language in this. When sign language is displayed in front of a camera, an image can be captured, processed using stored images, and output displayed as text before being translated to voice.

There are numerous strategies used in this, including feature extraction, segmentation, and hand tracking. This method is used to process the image. The OTSU method, which performs picture thresholding by dividing pixels into foreground and background, is a key approach in image processing. This allows us to remove any unnecessary context and obtain precise input for picture processing. We can save sign language signals in learned data for machine learning. When classifying data using the Naive Bayes approach, the classifier makes the assumption that the existence of one feature in a class is unrelated to the presence of any other feature.

II. METHODOLOGIES

2.1 Sign Language

An organised language with a phonology, morphology, syntax, and grammar is sign language. A full-fledged natural language, sign language employs numerous modes of expression to facilitate ordinary communication. The communication is transferred from human-human to human-computer interaction through sign language recognition software. Therefore, sensor-based and vision-based approaches are the two basic methods employed in sign language recognition. Vision-based strategy. In this method, a camera captures a gesture, identifies its key features, and records the image. Image processing methods are used to capture the image, and machine learning methods are used to obtain audio and text based on the image. This method's primary benefit is that it provides the highest level of accuracy and resilience.

2.2 Image Processing

Image processing is a technique for applying certain operations to an image to produce an improved image or to extract some relevant information from it. It is a kind of signal processing where the input is an image and the output can either be another image or features or characteristics related to that image.

OTSU Approach: is employed to carry out automatic image thresholding in computer vision and image processing. The algorithm returns a single intensity threshold in its most basic form, dividing pixels into the foreground and background classes. This threshold is established by maximising inter-class variation or, alternatively, minimising intra-class intensity variance. Otsu's approach is comparable

to a globally optimal k-means and is a one-dimensional discrete analogue of Fisher's Discriminant Analysis. It is also connected to Jenks optimisation method.

2.3 Machine Learning

In order to generate predictions or choices without being explicitly taught to do so, machine learning (ML) algorithms construct a mathematical model based on sample data, also referred to as "training data".

Method of Naive Bayes Classification:

A group of classification algorithms called Naive Bayes classifiers using Bayes Theorem-based algorithms. The Naive Bayes classifier makes the assumption that a certain feature's presence in a class has no bearing on the presence of any other feature. The following is how the Naive Bayes Classifier is expressed by utilising a camera to take pictures or videos and then processing those pictures or videos. A vision-based sign language identification system use the OTSU method to extract the image and remove any undesired background noise after analysing the data with static images, recognising the image with the aid of algorithms, and producing words for display. To boost system performance, it applies the Naive Bayes Classification algorithm to convolutional neural networks. This method's primary benefit is that it requires less calculation time and responds quickly to real-time applications.

III. DESCRIPTION OF THE METHOD

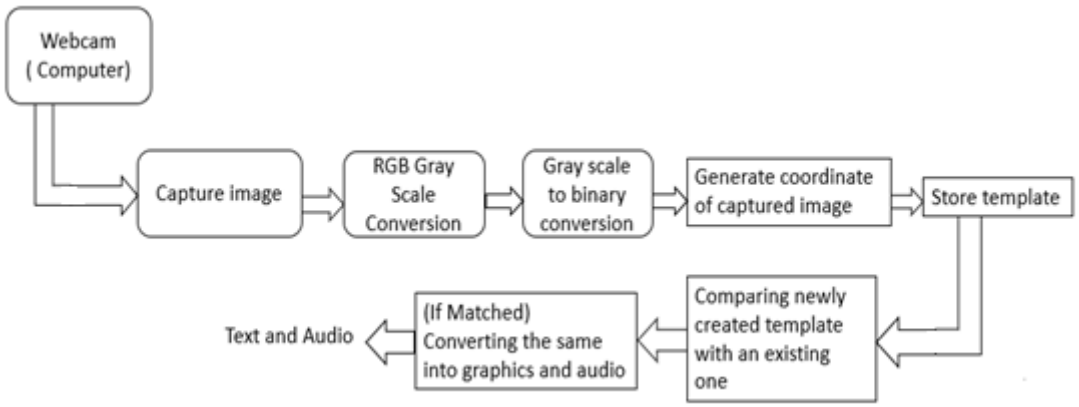

Fig.3.1: Block Diagram

3.1 Initialization and Orientation of Camera

The hardware component that interacts with the camera and produces a standard output for further image processing is called the camera interface block. OpenCV is a free, open-source computer vision library that serves as a user-machine interface. There are numerous versions of OpenCV that support languages including C, Python, and C++.

To make choosing a background simple, it's crucial to carefully select the direction in which the camera points. Pointing the camera at a wall or the ground (or a desktop) are the two practical possibilities. However, since there was only one overhead light, if the camera was aimed downward, light intensity would be greater and shadowing effects would be minimised.We are utilising a 16 MP Intex Night Vision Webcam for the camera. Night Vision is a function included with the Intex 16 MP webcam. This webcam can operate in complete darkness and prov+ides clear video imaging. Additionally, the night vision provides clear visuals in low light. Depending on the situation, the webcam's upper section can be moved. The image that was obtained has a resolution of 640x480, a frame rate of up to 30, and an RGB 24, 1420 colour space.

3.2 Image Acquisition from camera

An image is described as a two-dimensional function, F(x,y), where x and y are spatial coordinates. The intensity of an image at any given position is defined as the amplitude of F at any given pair of coordinates (x,y). We refer to it as a digital image when the (x,y), and amplitude values of F are finite. In other words, a two-dimensional array specifically set up in rows and columns can be used to define an image.A digital image is made up of a finite number of components, each of which has a unique value at a specific position. These components are also known as pixels, image elements, and picture elements. The most frequent usage of a pixel is to indicate a component of a digital image.

Image Matrix: Images are represented in rows and columns. Every element of this matrix is called image element , picture element , or pixel.

The first step in image acquisition is to capture an image using the integrated camera while it is running. These images are then stored in the directory after being captured, and the most recent image is then acquired and compared to images that have been stored in the database for a particular letter. The comparison reveals the gesture that was performed as well as the translated text for the next gesture. The images will be taken using a webcam opened using the basic OPENCV code, and then the image will be taken using frames per second and stored in a different directory from where all the input images are stored. The most recent image taken will then be picked up, and a comparison with earlier images will be made.

The histogram offers several solutions to image-related issues. The histogram can be used to distinguish between well exposed and imperfectly exposed photos. Thus, it is important to create a histogram from the collected image.

3.3 Hand region segmentation and Detection

Hand Contour: This method is used to find the palm of the hand and is based on the OpenCV method cv2.contourarea(). It finds the contour region of the hand to send to the model.

RGB Colour Recognition: Any colour image is essentially a combination of red, green, and blue colours. When implementing a computer vision system, one significant tradeoff is deciding whether to differentiate objects using colour or black and white, and if colour, which colour space to employ (red, green, blue, or hue, saturation, and brightness).Although using intensity alone (black and white) reduces the amount of data to analyse and hence the processing load, it makes distinguishing skin and markers from the backdrop considerably more difficult (since black and white data has less variation than colour data). As a result, it was decided to employ colour

difference. Furthermore, the maximum and lowest HSL pixel colour values of a tiny test patch of skin were calculated manually. These HSL ranges were then employed in a future frame to detect skin pixels (detection was signified by a change in pixel colour to white).
Calibration of Colours
An region of the screen was marked for calibration (2) in order to automatically determine the colour ranges (1). It was then a simple matter of positioning the hand or marker (colour rings) within this area and scanning it to get the ranges' maximum and minimum RGB values(3). The following is a formal description of the initial calibration method: The image is a two-dimensional array of pixels:
Converting a colour image to a binary image: To convert a colour to a grayscale representation of its luminance, first acquire the values of its red, green, and blue (RGB) primaries. A grayscale image, also known as a grayscale digital image, is one in which the value of each pixel is a single sample, carrying just intensity information. Images of this type, commonly known as black and white, are made up entirely of grayscale shades ranging from black at the weakest intensity to white at the strongest. A binary image is a computer image in which each pixel has just two possible values. A binary image is often composed of two colours: black and white, but any two colours can be used. The colour used in the image for the object is the foreground colour, while the remainder of the image is the background colour. Until date, the classification of skin and marker pixels has relied on a basic RGB bounding box.

The most basic approach of image segmentation is thresholding. This method converts an RGB image to a binary image.
Binary images are computer images with only two possible values (0 or 1). Typically, two colours are used for each pixel, black and white, however any two colours can be used. The background pixels are changed to black pixels, and the pixels containing our area of interest are turned to white pixels. It is only the pre processing. We use the OTSU thresholding method for thresholding.

3.4 Extraction of Features and Detection of Orientation

There are various features, or noteworthy points on the item, that can be extracted to create a "feature" description of the thing. SIFT image features provide a set of object features that are unaffected by many of the problems seen in other approaches, such as object scaling and rotation. For image feature generation, the SIFT technique takes a photograph and converts it into a "large collection of local feature vectors." Each of the feature vectors is never changed by the image's scaling, rotation, or translation. It will accept hand movement input in any shape or direction, and the gesture will be detected by the given portion of feature extraction as the SIFT algorithm also includes the orientation
assignment procedure.
As the SIFT algorithm also includes the orientation assignment procedure.
Finally, once the entire procedure has been completed, the application will be translated into its recognised letter or alphabet from the gesture, which may be useful for understanding in layman's language. The following procedure includes handing out a 1-dimensional array of 26 characters corresponding to alphabets, with the image number recorded in the database provided in the array. We are employing the Nave Bayes Classifier, which is based on Linear Classification, for this purpose.
Mathematical Expression: Naïve Bayes Classification
P(h): the probability of hypothesis h being true (regardless of the data). This is known as the prior probability of h.
P(D): the probability of the data (regardless of the hypothesis). This is known as the prior probability.
P(h|D): the probability of hypothesis h given the data D. This is known as posterior probability.
P(D|h): the probability of data d given that the hypothesis h was true. This is known as posterior probability.

3.5 Display as Text and Speech

When a character is chosen based on a recognised sign using voice conversion, the corresponding text is converted to speech. This section falls under Artificial Intelligence, and once the template matching operation is complete, the matched image is converted into text and audio format. Predefined conversion methods are utilised for this purpose.

IV. SOFTWARE REQUIREMENTS

4.1 Python 3.7.4

Python is now the most popular high-level, multipurpose programming language. It supports procedural and object-oriented programming paradigms. Compared to other programming languages like Java, Python programs are typically smaller. Programmers have to type comparatively less, and the language's indentation requirement keeps their work always readable. Python is simple to use for GUI applications, machine learning, image processing, and other applications since it has a large library compared to other languages. Python is more productive than other languages, has extensible libraries, and can be extended to other languages.

Fig.4.1:Python 3.7 Version

For this algorithm python 3.7 version and command window is installed. After installing python 3.7 we store some extended libraries which are required for the algorithm using command window. Then we will write code to store symbols and to process the image and these codes are saved using extension .py with different file names.

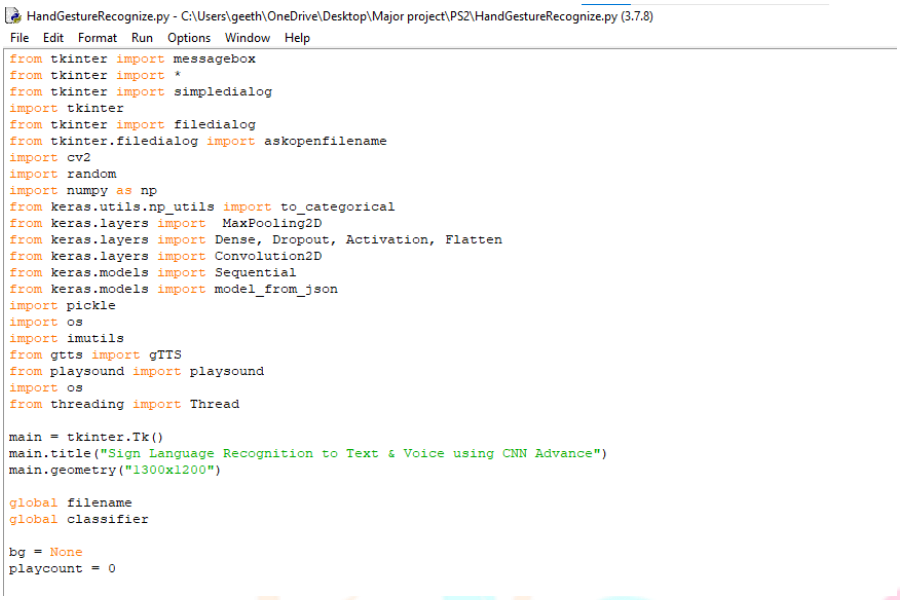

Fig.4.2:Idle Display

Now open the code using idle and click on run option then the system captures image as input and displays output as text in idle python shell as well as CNN display.

V. OUTPUT

The following results are observed:
1)Output shown in convolution neural network
In this when we upload hand gesture the image is captured and checks with the stored images then displays output as text as shown in Fig.5.2.

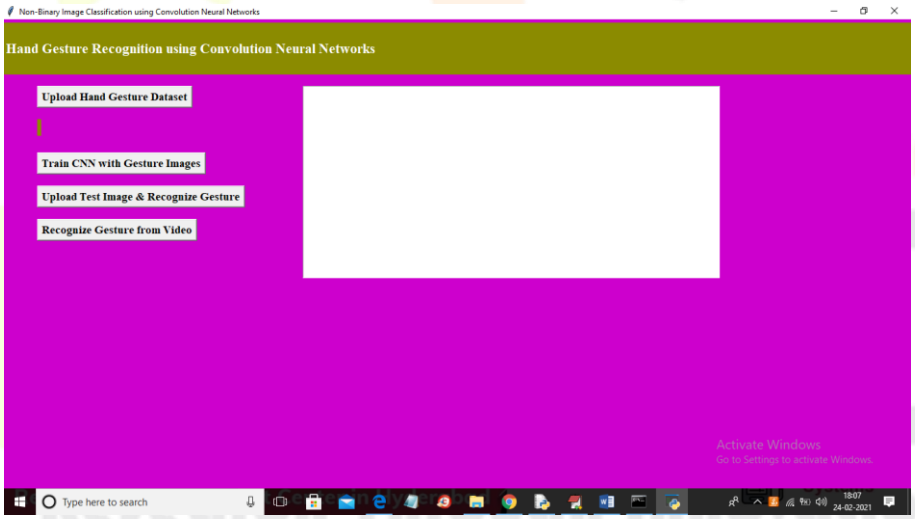

Fig.5.1:Convolution Neural Network

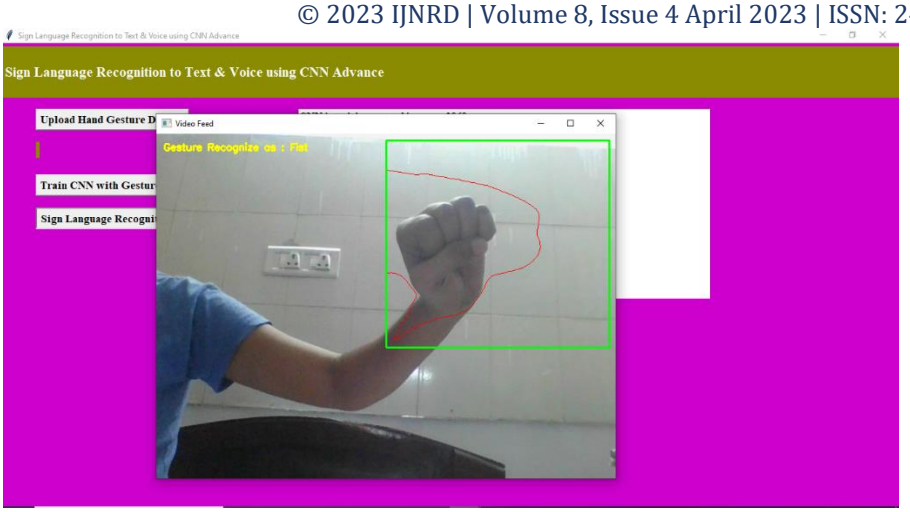

Fig.5.2: Output of Sign Fist in CNN

2) Output displayed in Idle python shell
In this when a sign is shown to camera it displays output for every milli second.

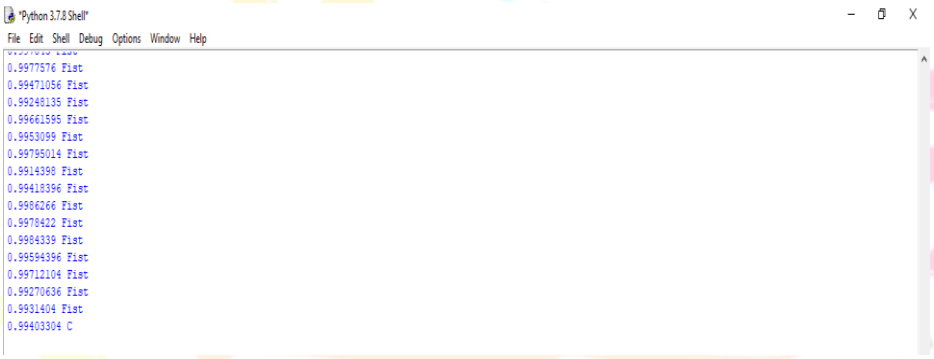

Fig.5.3:Output in Idle Shell

VI. CONCLUSION

To engage in human contact The most crucial aspect of life is communication, but it might be the hardest for the dumb and the deaf. In this project, sign language recognition is used to provide an interface for communication with hearing-impaired people. This is the best method for social use and may be utilised in both public and private settings. Through the use of a low-cost computer, our initiative aims to close the communication gap by allowing sign language to be automatically recorded, recognized, and translated into speech for the benefit of blind people.

Naive Bayes Classification is employed in this project so that we may obtain the highest accuracy with the least amount of noise distortion. It is converted into speech for hearing by those with hearing impairments. Although we are utilising ASL for this experiment, we may use BSL or ISL in the future. By recognising sign language and translating it into text, this project can assist dumb and deaf persons in communicating with others. This can be utilised by blind persons by turning the text into speech.

REFERENCES

[1] Denise Powell, „A Case Study of Two Sign Language Interpreters Working in Post-Secondary Education in New Zealand", International Journal of Teaching and Learning in Higher Education,,2013 Volume 25, Number 3, 297-304.
[2] C.W.Ng and S.Ranganath , „Static Hand Gesture Recognition System using Convolutional Neural Networks", IJSRD - International Journal for Scientific Research & Development,Vol. 6, Issue 01, 2018 ,ISSN (online): 2321-0613.
[3] N.Tanibata, „Automated Extraction of Signs from Continuous Sign Language Sentences using Iterated Conditional Modes", Journal of Engineering and Applied Sciences, August 2014.
[4] H. K. Nishihara, „Recognition of American Sign Language using Image Processing and Machine Learning", International Journal of Computer Science and Mobile Computing, Vol. 8,Issue 3,March 2019,pp:352-357.
[5] Jagdish L. Raheja, „Android based Portable Hand Sign Recognition System", International Journal of Computer Trends and Technology (IJCTT) V40(3):165-171, October 2015. ISSN:1651-1656.
[6] Shweta S.Shinde, Rajesh M. Autee and Vitthal K. Bhosale, „Sign Language Recognition for Deaf & Dumb", International Journal of Advanced Research in Computer Science and Software Engineering 3(9), September - 2013, pp. 103-106.