COMPUTER VISION-ENABLED GESTURE RECOGNITION FOR DIFFERENTLY-ABLED PEOPLE

T. Anusha, Deepti Ravi Kumar, Nandha Kishore, Rahul Raj D, Raswanth E A, Samyuktha Sreekanth

Department of Computer Science and Engineering
(PSG College Of Technology , Coimbatore)
Avinashi Road , Coimbatore , Tamil Nadu - 641004

---***---

Abstract - *This research aims to lower the barrier of day-to-day communication for disabled people by developing a user-friendly, cost-effective system by which we can determine the most appropriate character from the sign that the user is showing to the system. About nine million people in the world are deaf and mute. Communication between differently-abled people and general people has always been a challenging task but sign language helps them to communicate with other people. But not everyone understands sign language and here is where our system will come into the picture. Various machine-learning algorithms have been investigated to facilitate pattern identification and processing. Advanced Python is used to train the model using the image features that were obtained. In response to the sign presented to the model, the trained model accurately predicts the words that are most appropriate using the datasets that are fed into the system. Words that are predicted are generated into a voice output. Sign language provides a way for speech-impaired and hearing-impaired people to communicate with other people. Instead of a voice, sign language uses gestures to communicate. Sign language is a standardized way of communication in which every word and alphabet is assigned to a distinct gesture The solution aims to assist those in need, ensuring social relevance by offering an interface that can help facilitate simple and unrestricted communication between a wide range of people Computer vision and machine learning researchers are now conducting intensive research in the area of image-based hand gesture identification. Intending to make human-computer interaction (HCI) simpler and more natural without the use of additional devices, it is an area where many researchers are researching. Therefore, the main objective of research on gesture recognition is to develop systems that can recognize certain human gestures and use them, for instance, to convey information.*

Key Words: Media pipe, OpenCV, SLR , gTTs , HCI

1. INTRODUCTION

A crucial application of gesture recognition is sign language detection. Current technologies for gesture recognition can be divided into two types: sensor-based and vision-based.In sensor-based methods, data gloves or motion sensors are incorporated from which the data of gestures can be extracted. Even minute details of the gesture can be captured by the data capturing glove which ultimately enhances the performance of the system. However, this method requires wearing a data-capturing hand glove with embedded sensors which makes it a bulky device to carry. This method affects the signer's usual signing ability and it also reduces user amenities. Vision-based methods include image processing. This approach provides a comfortable experience to the user. The image is captured with the help of cameras. No extra devices are needed in the vision-based approach. This method deals with the attributes of the image such as colour and texture that are obligatory for integrating the gesture. Although the vision-based approach is straightforward, it has many challenges such as the complexity and convolution of the background, variations in illumination and tracking other postures along with the hand object, etc. arise. Sign language provides a way for speech-impaired and hearing-impaired people to communicate with other people. Instead of a voice, sign language uses gestures to communicate. Sign language is a standardized way of communication in which every word and alphabet is assigned to a distinct gesture. It would be a win-win situation for both differently-abled people and the general public if such a system is developed where sign language could be converted into text/speech. Technology is advancing day after day but no significant improvements are undertaken for the betterment of specially-abled communities. Nine million people worldwide are both deaf and silent. Differently abled persons and regular people have long found it difficult to communicate, but sign language makes it easier for them to do so. However, not everyone is fluent in sign language, which is where our approach will be useful.

1.1 CHALLENGES IN THE DAY-TO-DAY LIFE OF A BLIND PERSON

1. Large Sections Of The Internet Are Unusable

   Almost all of the world's main operating systems have some sort of software that can assist in turning the text on your screen into audio messages. Unfortunately, the majority of these accessible solutions lack advanced technology. In other words, despite all the amazing things technology can achieve, if you are unable to see a webpage, you might have to listen to its full contents in linear order.

2. Braille Illiteracy Is Growing Due To A Lack Of Infrastructure

   The worrying trend of braille illiteracy is a problem in the blind world. Braille is clunky, challenging to master, and sluggish to process, according to some blind readers. A braille book is much longer than the same book in print because braille takes up much more space on a page.

3. Withdrawing Money From An ATM Can Be Time-Consuming

   Although the ability to use headphones to hear prompts is becoming more prevalent, it is still uncommon. Braille on the buttons doesn't assist if you can't see the matching messages and numbers on the screen, making an ATM inaccessible to the blind if it doesn't communicate.

4. Lowered Expectations Can Be Life-Halting Hurdles

   Due to the widespread misperception that being blind is a disability, blind persons are frequently dismissed on a systemic level before they have the opportunity to show their potential. Blind persons are frequently employed by factories that pay them cents on the dollar for their labour after training them to perform menial activities in "sheltered workshops."

2. LITERATURE REVIEW

1. Conversion of Sign Language into Text -Mahesh Kumar N B

This paper revolves around the establishment of a general framework for creating an SLR model .SLR data processing involves sign representation, normalization and filtering, data formatting and organization, feature extraction and feature selection.

In this research, MATLAB is used to demonstrate the recognition of 26 hand motions in Indian sign language. The suggested system has four modules, including feature extraction, sign recognition, sign-to-text, pre-processing and hand segmentation. The suggested system aids in the reduction of dimensions. It becomes more difficult to visualize the training set and subsequently work on it as the number of features increases. Sometimes, the majority of these traits are redundant since they are connected. Algorithms for dimensionality reduction are useful in this situation. By getting a set of principal variables, dimensionality reduction is the process of reducing the number of random variables being considered. It can be split into two categories: feature extraction and feature selection.

BSL uses a two-handed fingerspelling system in the current systems, as opposed to ASL's use of a one-handed approach (and FSL).

Gesture recognition has several uses, including the understanding of sign language.

Sign language recognition has two different approaches.

- Glove-based approaches

- Vision-based approaches.

Glove-based techniques Signers in this category must put on a sensor glove or a coloured glove. By wearing gloves during the segmentation process, the task will be made simpler. The disadvantage of this strategy is that the signer must wear the sensor hardware and the glove while the device is in operation. LDA is mostly utilised in machine learning, pattern identification, and statistics. It is employed to identify a linear combination of properties that distinguishes between at least two classes of objects or occurrences. LDA describes how to model the distinctions between the various data classes. Continuous measurements are conducted on independent variables for each observation in LDA.

Demerits

- Designing SLR tools that get close to 100% accuracy on a big vocabulary is not yet done.
- Future emphasis should be paid more to the usability factor.
- generate criticism
- Techniques that allow for the quick identification of frequent mistakes while guaranteeing that user input is respected.

2. Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues - Muhammad Al-Qurishi , ,Thariq Khalid , Riad Souissi

The solution's input layer is made up of a hand sign presentation display and an input device based on SLR data collection technique . The second layer is the pre-processing layer, which decodes a sign into the necessary data format and filters gesture data. There may also be extra stages, including sample normalization or integrating data from a video's subsequent frames. Feature extraction is the first operation the system takes after receiving sign data. Visual features, hand movement characteristics, 3D skeleton features, and face features are just a few examples of the many various kinds of features that can be employed as the main source of information. One of the key elements affecting the SLR method's effectiveness is the choice of features to be used in the algorithm training. Before being supplied to the modelling layer, the data are normally processed and converted into a vector format.

Modules involved –

Data collection can be divided into 3 - Hardware-based, vision-based and hybrid

• Hardware-based methods make use of tools or wearable sensors. To recognize sign language, wearable technology frequently attaches sensors to the user or uses a glove-based system. These tools, whether they are gloves, rings, or sensors, can translate sign language into text or speech. Because cumbersome apparatus must be worn, sensor-based techniques are never natural. They instead suggest Real-Sense, a cutting-edge method that can naturally identify and track hand placements.

• Vision-based techniques are less restrictive on users than sensor-based ones. Recent SLR studies sometimes rely on input in the form of visual data. They are limited by the subpar functionality of traditional cameras. Another issue is that simple hand features can lead to ambiguity while complex features take more time to process.

• Hybrid- When compared to other methods, hybrid methods perform as well as or better. This method calls for hardware and vision-based modalities to be calibrated, which can be particularly difficult. This approach is quicker because it doesn't need to be retrained. Since it influences how the models are trained and, consequently, how quickly they may become skilled at discerning between distinct signs or words, feature extraction is a crucial stage for all SLR models. Features are always derived from raw data and correspond to the positions of body components crucial for sign language communication (certain spots on the hands and face). Features are determined using statistical methods with predetermined weights and expressed as vectors.

Essentially, feature selection means reducing the amount of information in the data to a select few relevant statistical characteristics, which are then fed into the machine learning network. The goal is to reduce the number of calculations required to produce an accurate forecast by only including those features that significantly improve the algorithm's ability to distinguish between distinct classes.

3. Signet: A Deep Learning based Indian Sign Language Recognition System -Sruthi C. J and Lijiya A

Convolutional Neural Network architecture is implemented for Indian Sign Language static Alphabet recognition from the binary silhouette of the signer hand region after reviewing several existing methods in sign language recognition. The proposed method was implemented successfully with an accuracy of 98.64%, which is superior to the majority of the currently used techniques. Problems with hand gesture recognition are approached using either a glove-based method or a vision-based one. Since these particular devices capture data directly from signers, gloves offer good accuracy.

Based on features determined using various image or video processing techniques, vision-based systems perform the task of object recognition from images or videos.

Demerits:

The suggested architecture can eventually be expanded with new approaches and modules to create a fully automated sign language recognition system. Soon, facial expression and context analysis will be incorporated into sign language recognition. The resulting recognition accuracy outperforms the majority of the available techniques.

4. Real Time Sign Language Recognition and Speech Generation- Kartik Shenoy, Tejas Dastane,Varun Rao , Devendra Vyavaharkar

Because they are more accurate than gadget methods, image processing and neural network-based sign language recognition systems are chosen. create a neural network-trained sign language recognition system that is both user-friendly and accurate, generating both text and speech from the input gesture.

Merits:

- The proposed solution was put to the test in real-world scenarios, and it was demonstrated that the classification models that were created could identify every taught gesture.
- The system will be improved in the future, and trials using full language datasets will be conducted.

5. Indian Sign Language Recognition System- Yogeshwar Ishwar Rokade , Prashant Jadav

In gesture recognition, sign language recognition forms an important application. It consists of two different approaches.

1) Glove-based approach:

The signer in this case needs to wear a sensor or a coloured glove. Using the glove makes the segmentation phase's duty easier. The drawback of this strategy is that the signer must carry the sensor hardware, including the glove, throughout the entire process.

2) Vision-based approach:

It makes use of the algorithms of image processing for detecting and tracking the hand signs including the signer's facial expressions. This vision-based approach is simple since the signers need not wear additional hardware. In the proposed system vision-based approach is used.

PROPOSED APPROACH:

A. Image Acquisition

The number of signs made use in the system are

A, B, D, E, F, G, H, J, K, O, P, Q, S, T, X, Y, Z.

B. Hand Object Detection:

- Hand Segmentation
- Filter and Noise Removal
- Feature Extraction
- Classification

Table -1: Accuracy chart

Feature Set	AAN Average Accuracy	SVM Average Accuracy
6 features	88.3%	69.92%
8 features	91.94%	89.33%
13 features	94.37%	92.12%

6. Real-Time Indian Sign Language (Isl) Recognition - Kartik Shenoy ,TejasDastane Varun Rao , Devendra Vyavaharkar

The method described in this work uses grid-based characteristics to recognise hand positions and motions from the Indian Sign Language (ISL) in real-time. With the help of this system, the communication gap between the hearing- and speech-impaired and the general public is meant to be closed. The current solutions are either not real-time or offer just a modest level of accuracy. The output from this system is good for both parameters. It can recognise 33 hand positions and a few ISL movements. A smartphone camera records sign language, and the frames are sent to a distant server for processing. It is user-friendly because no other hardware is required, such as gloves or the Microsoft Kinect sensor. Hand detection and tracking employ methods including Face detection, Object stabilisation, and Skin Colour Segmentation. A Grid-based Feature Extraction approach is also used in the image, representing the attitude of the hand as a Feature Vector. The k-Nearest Neighbours method is then used to categorise hand positions. The motion and intermediate hand postures observation sequences, on the other hand, are given to Hidden Markov Model chains corresponding to the 12 pre-selected gestures defined in ISL for gesture categorization. The system can attain an accuracy of 99.7% for static hand postures and an accuracy of 97.23% for gestures using this methodology.

- Dataset Used:
- Pre-Processing:
- Hand Extraction And Tracking
- Feature Extraction Using Grid-Based Fragmentation Technique
- Classification : Recognition of ISL Hand poses using k-NN
- Gesture Classification using HMM
- Temporal Segmentation

33 STATIC HAND POSES	99.7%
12 GESTURES	97.23%

Table-2: Accuracy chart

7. Text to Speech Conversion - S. Venkateswarlu , Duvvuri B K Kamesh Duvvuri ,Sastry Jammalamadaka

The user can now hear the contents of text images rather than reading them thanks to a novel, effective, and cost-beneficial technique proposed in the current research. It combines the principles of Text to Speech Synthesizer (TTS) and Optical Character Recognition (OCR) in the Raspberry Pi. People who are blind or visually challenged can efficiently communicate with computers using this type of device. In computer vision, text extraction from colour images is a difficult task. Using OCR technology, text-to-speech conversion reads English alphabets and numbers that are present in images and converts them into voices. The device's design, implementation, and experimental findings are covered in this publication. The voice processing module and the picture processing module make up this gadget. The device's 900 MHz processing speed was taken from the Raspberry Pi v2 platform.

8. Design and Implementation of Text To Speech Conversion for Visually Impaired People - ItunuoluwaIsewon ,JeliliOyelade Olufunke Oladipupo

There are two basic stages to the text-to-speech (TTS) synthesis process. First is text analysis, which converts the input text into a phonetic or other linguistic representation; second is speech waveform creation, which creates the output using this phonetic and prosodic data. Typically, these two stages are referred to as high-level and low-level synthesis. The input text could be, for instance, scanned text from a newspaper, standard ASCII from email, a mobile text message, or data from a word processor. The character string is then pre-processed and examined to produce a phonetic representation, which is typically a string of phonemes plus some extra information for proper intonation, length, and stress. The information from the high-level synthesiser is ultimately used to make speech sounds using the low-level synthesiser. There have been numerous reported mechanical attempts to produce speech-like sounds since the eighteenth century.

9. P.V.V Kishore, P. Rajesh Kumar, E. Kiran Kumar &S.R.C.Kishore, 2011,Video Audio Interface for Recognizing Gestures of Indian Sign Language, International Journal of Image Processing (IJIP), Volume 5, Issue 4, 2011 pp. 479- 503

The authors of the paper [12] suggested a system that can translate ISL motions from a video feed into English voice and text and recognise them. They achieved this by employing a variety of image processing techniques, including edge detection, wavelet transform, and picture fusion, to separate the shapes in the video stream. Shape features were extracted using Ellipsoidal Fourier descriptors, and the feature set was optimised and reduced using PCA. The system was trained using the fuzzy inference system, and it achieved an accuracy of 91%.

10. Adithya, V., Vinod, P. R., Gopalakrishnan, U., and IEEE (2013). "Artificial Neural Network Based Method for Indian Sign Language Recognition."

The authors of the paper [13] proposed a method for automatically identifying movements in Indian sign language. The suggested method transforms the input image before using it to detect hands using YCbCr colour space and digital image processing techniques. Some of the methods used to extract the features include distance transformation, projection of distance transformation coefficients, Fourier descriptors, and feature vectors. The data were classified using an artificial neural network, and the recognition rate was 91.11 per cent.

3. METHADOLOGY

We have used Media Pipe holistic for real-time image tracking. Separate models for posture, face, and hand components are individually integrated into the Media Pipe Holistic pipeline and are each customized for their respective fields. As a multi-stage pipeline, Media Pipe Holistic is made to treat each region with the proper image resolution. We start by collecting key points from Media Pipe holistic and collect a bunch of data from key points i.e., our hands, our body and on our face and save data in the form of NumPy arrays. We can vary the number of sequences according to our needs, but each sequence will have 30 frames.

• We then build an LSTM model to train with our stored data which helps us to detect action with several frames. The usage of LSTM was advantageous because LSTM networks are an extension of recurrent neural networks (RNNs) which were developed primarily to address failures in RNNs. LSTMs provide us with a large range of parameters such as learning rates, and input and output biases. Thus, there is no need for precise modifications. With LSTMs, updating each weight is much more similar to Back Propagation Through Time (BPTT), reducing complexity to O(1).

• Image processing is the process of analysing a picture and turning the results into sentences. For this, a dataset with a significant number of images and correspondingly detailed captions is needed. To forecast the characteristics of the photos in the dataset, a trained model is employed. This comes under photo data. The dataset is then analysed such that only the most intriguing terms are included in it. Data in text format. We try to fit the model with these two sorts of data. By using input words that were previously predicted by the model and the image, the model's task is to produce a descriptive sentence for the picture, one word at a time.

• We have utilized categorical classification for model training as we include multiple classes of signs However there is no intrinsic ordering of the categories. The number of epochs for the model is determined by us if we increase the number of epochs the accuracy increases but the time taken to run the model also increases and overfitting of the model can happen, for gesture recognition.

• Once training is done, we can use this model for real-time hand gesture detection and simultaneously convert the gesture to speech using OpenCV. OpenCV is a significant open-source toolkit for image processing, machine learning, and computer vision. Python, C++, Java, and many other programming languages are supported by OpenCV. It can analyse pictures and movies to find faces, objects, and even human handwriting. When it is integrated with various libraries, such as Numpy which is a highly optimized library for numerical operations.

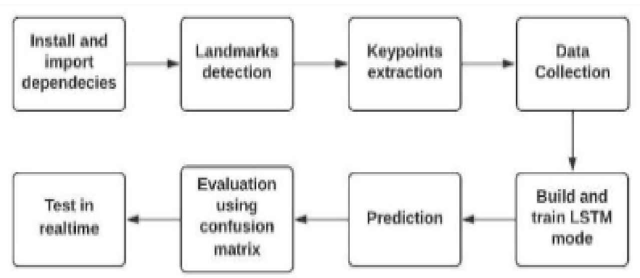

Fig-1: System flow – 1

Table-3: Comparison between existing and proposed model

	EXISTING SOLUTION	PROPOSED SOLUTION
MODEL	CNN Model	LSTM Model
DATASET	MNIST Dataset	Self Generated Dataset Using Different Cameras with Variable Pixel Quality
ACCURACY	93%	98.2%
RECOGNITION BASED ON	Letters (ASL)	Words (ASL)
PREDICTION	Single Frame	Multi-Frames Appended to Sequence

4. SYSTEM ARCHITECTURE

4.1 Input Video Acquisition:

We have used OpenCV for capturing the input video through several devices such as desktop webcam, phone camera and DSLR. Python comes with several libraries for processing images and videos. OpenCV is one among them. A large library called OpenCV helps provide a variety of methods for image and video operations. We can record video from the camera using OpenCV. It lets you create a video capture object which is helpful to capture videos through a webcam and then edit them as you see fit. We can specify the countdown timer instead of choosing one of the specified countdowns. When the appropriate key is pressed, the countdown timer is started, and we display the countdown on our camera using the cv2.putText() function. When the countdown reaches zero, we take an image, display it for a predetermined amount of time, and then write or save it to disc.

4.2 Feature extraction using Media Pipe Holistic:

We have used Media Pipe Holistic for feature extraction. The Media Pipe Holistic pipeline includes separate models for posture, face, and hand components that are each tailored for their particular industries. Media Pipe Holistic, a multi-stage pipeline, is designed to treat each region with the appropriate image resolution. Starting with key points from Media Pipe Holistic, we gather a lot of data from key points such as our hands, our body, and our face and record the information as NumPy arrays. The number of sequences can be changed to suit our needs, but each series will include 30 frames.

4.3 Training the input data with Long Short-Term Memory (LSTM) model:

Here, we employ a unique class of RNN called an LSTM, which enables our network to learn long-term and training of our stored data takes place, which enables us to recognize action over several frames. The second technique involves sending the Inception CNN model's anticipated labels to an LSTM. The input to the LSTM is provided by the final hidden layer of the CNN. We utilized a network with a single layer of several LSTM units after extracting the bottleneck features, then a fully connected layer using SoftMax activation. A regression layer is then applied.

4.4 Prediction of gestures:

We choose the number of epochs for the model; while this increases accuracy, it also lengthens the time it takes to run and increases the risk of overfitting the model, which is problematic for gesture recognition.

4.5 Usage of Text-to-Speech and Displaying output text on screen:

The gestures were translated into vocal output using Google Text-to-Speech. In addition to this, the output text is displayed on the screen. This whole set of the process takes place again for recognizing a sentence.

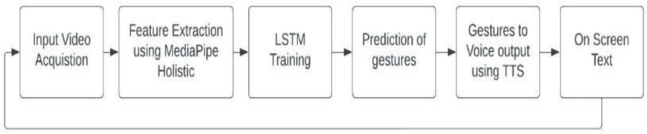

Fig-2 : System flow 2

5. IMPLEMENTATION AND RESULT

5.1 Extract Media Pipe Holistic Key points

The Media Pipe Holistic pipeline integrates separate models for pose, face and hand components, each of which is optimized for their particular domain. However, because of their different specializations, the input to one component is not well-suited for the others. The pose estimation model.

```
mp_holistic = mp.solutions.holistic # Holistic model
mp_drawing = mp.solutions.drawing_utils # Drawing utilities
```

```
def mediapipe_detection(image, model):
    image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB) # COLOR CONVERSION BGR 2 RGB
    image.flags.writeable = False                  # Image is no longer writeable
    results = model.process(image)                 # Make prediction
    image.flags.writeable = True                   # Image is now writeable
    image = cv2.cvtColor(image, cv2.COLOR_RGB2BGR) # COLOR COVERSION RGB 2 BGR
    return image, results
```

Fig-3 : Media Pipe Integration of separate model

5.2 Build a Sign Language model using a Action Detection powered by LSTM layers

The vanilla LSTM network has three layers; an input layer and a single hidden layer followed by a standard feedforward output layer. The proposed system is designed to develop a real-time sign language detector using a. TensorFlow object detection API and train it through transfer learning.

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense
from tensorflow.keras.callbacks import TensorBoard
```

```
model = Sequential()
model.add(LSTM(64, return_sequences=True, activation='relu', input_shape=(30,1662)))
model.add(LSTM(128, return_sequences=True, activation='relu'))
model.add(LSTM(64, return_sequences=False, activation='relu'))
model.add(Dense(64, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(actions.shape[0], activation='softmax'))
```

Fig- 4: Building a model powered by LSTM layers

5.3 Predict sign language in real time using video sequences

The system makes real-time predictions of the signs shown by the user on the camera and appends the labels of the predictions to a sentence belt on the screen. It then voices out the prediction made.

Fig- 5: Real time detection and prediction

5.4 Voice Output for predictions made using gTTs

There are several APIs available to convert text to speech in Python. One such API is the Google Text-to-Speech API commonly known as the gTTS API. gTTS is a very easy-to-use tool which converts the text entered, into audio which can be saved as an mp3 file.

```
from gtts import *
import random
import playsound
default_location="/Users/HP/Desktop/Deepti"
def speak(audio_string):
    tts = gTTS(text=audio_string,lang='en')
    r = random.randint(1,10000000)
    audio_file = 'audio-'+str(r)+'.mp3'
    tts.save(audio_file)
    playsound.playsound(os.path.join(default_location,audio_file))
    print(audio_string)
    os.remove(audio_file)
```

Fig-6 : Voice Output for predictions made using Gtts

5.5 Training and Loss graphs

The model is trained on categorical classification for over 2000 epochs.

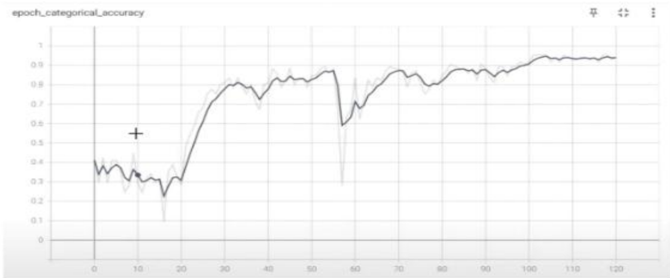

Chart-1 : Categorical accuracy graphs

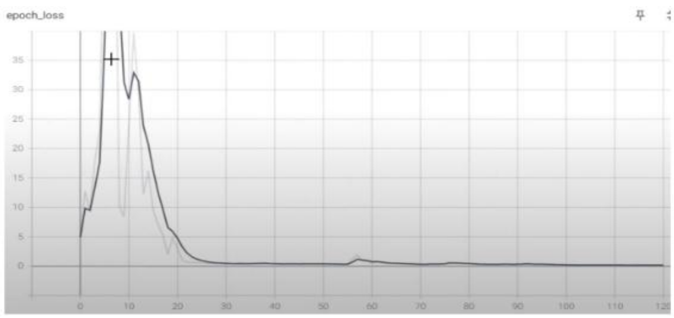

Chart-2 : Training Loss graphs

The trained model is found to be of an accuracy of 98.2%.

The loss over the training epochs is found to be nearly zero

6. FUTURE SCOPE

Motivations to pursue further research to develop an improvised version of the proposed system. Enhancing the recognition capability under various lighting conditions. Implementing and identifying a greater number of gestures and at the same time maintaining higher accuracy. Applying this gesture recognition for accessing internet applications. We intend to broaden the scope of our domain scenarios and integrate our tracking system with a range of hardware, such as digital TV and mobile devices. We want to make this technique accessible to a variety of people

7. CONCLUSION

Our method, called hand gesture recognition, acknowledged that constructing a database from scratch utilizing video sequences and frames has time constraints and that the process is sensitive to variations in gesture. It is a well-known issue for humanity that some people are unable to talk or hear. The technology will provide a user interface that makes communication . Our method, called hand gesture recognition, acknowledged that constructing a database from scratch utilizing video sequences and frames has time constraints and that the process is sensitive to variations in gesture. It is a well-known issue for humanity that some people are unable to talk or hear. The technology will provide a user interface that makes communicating with people with disabilities straightforward. The Long Short-Term Memory Network has performed exceptionally well at recognizing sign language hand gestures in video clips. The technology recognizes sign language more rapidly and accurately than other techniques reported in the literature. Human-computer interaction has several potential uses for hand gestures, which are an effective form of human communication. When compared to conventional technologies, vision-based hand gesture recognition algorithms offer several demonstrated advantages. The current work is only a small step towards achieving the outcomes required in the field of sign language recognition, as hand gesture recognition is a challenging problem. The chosen hand traits worked well with machine learning techniques and could be applied in a variety of real-time sign language recognition systems. The system will continue to be improved, and trials using full language datasets will be conducted in the future. In cases where there is a need for trustworthy solutions to the issue of start and finish gesture identification, it is also planned to evaluate systems that can read dynamic sign language gestures. The proposed approach is a strong foundation for the creation of any vision-based sign language recognition user interface system, even if there is still more work to be done in this fielding with people with disabilities straightforward. The Long Short-Term Memory Network has performed exceptionally well at recognizing sign language hand gestures in video clips. The technology recognizes sign language more rapidly and accurately than other techniques reported in the literature. Human-computer interaction has several potential uses for hand gestures, which are an effective form of human communication. When compared to conventional technologies, vision-based hand gesture recognition algorithms offer several demonstrated advantages. The current work is only a small step towards achieving the outcomes required in the field of sign language recognition, as hand gesture recognition is a challenging problem. The chosen hand traits worked well with machine learning techniques and could be applied in a variety of real-time sign language recognition systems. The system will continue to be improved, and trials using full language datasets will be conducted in the future. In cases where there is a need for trustworthy solutions to the issue of start and finish gesture identification, it is also planned to evaluate systems that can read dynamic sign language gestures. The proposed approach is a strong foundation for the creation of any vision-based sign language recognition user interface system, even if there is still more work to be done in this field.

8. REFERENCES

[1] Mahesh Kumar N B "Conversion of Sign Language into Text ", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018)

[2] Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues - - Muhammad Al-Qurishi , ,Thariq Khalid , Riad Souissi , IEE , vol 9

[3] Muhammad Al-Qurishi , ,Thariq Khalid , Riad Souissi "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues" , IEE , vol 9

[4] Sruthi C. J and Lijiya A "Signet: A Deep Learning based Indian Sign Language Recognition System" , 2019 International Conference

[5] Yogeshwar Ishwar Rokade , Prashant Jadav "INDIAN SIGN LANGUAGE RECOGNITION SYSTEM",July 2017 International Journal of Engineering and Technology 9(3S):189-196

[6] Recognition Kartik Shenoy ,TejasDastane Varun Rao , Devendra Vyavaharkar "Real-Time Indian Sign Language (Isl)" ,9th ICCCNT 2018 July 10-12, 2018, IISC,

[7] S. Venkateswarlu , Duvvuri B K Kamesh Duvvuri ,Sastry Jammalamadaka"Text to Speech Conversion " , October 2016Indian Journal of Science and Technology 9(38)

[8] Isewon ,JeliliOyelade Olufunke Oladipupo "Design and Implementation of Text To Speech Conversion for Visually Impaired People Itunuoluwa ",International Journal of Applied Information Systems (IJAIS) – ISSN Volume 7– No. 2, April 2014

[9] Al-Jarrah, A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," The Journal of Artificial Intelligence 133 (2001) 117–138.

[10] Rajam, Subha & Ganesan, Balakrishnan. (2011) " Real time Indian Sign Language Recognition System to aid deaf-dumb people" , International Conference on Communication Technology Proceedings

[11] Geetha M., Manjusha U. C. "Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation", International Journal of Computer Science and Engineering (IJCSE)

[12] PV.V Kishore, P. Rajesh Kumar, E. Kiran Kumar &S.R.C.Kishore, " Video Audio Interface for Recognizing Gestures of Indian Sign Language", International Journal of Image Processing (IJIP), Volume 5, Issue 4, 2011

[13] Adithya, V., Vinod, P. R., Gopalakrishnan, U., and Ieee (2013). "Artificial Neural Network Based Method for Indian Sign Language Recognition."

[14] A Deep Learning-based Indian Sign Language Recognition System , International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 2, Issue 2, February 2022