WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Research Paper on Real-Time Sign Language Interpreter using Mediapipe Holistic

S.R.Aajmane[1], A.S.Neje[2], B.S.Khedkar[3], S.S.Koulage[4], S.M.Momin[5]

Department of Computer Science and engineering.Dr.J.J.Magdum college of Engineering, Jaysingpur,India
[1]aajmane09@gmail.com[2]nejeabhijeet99@gmail.com[3]baharkhedkar25@gmail.com[4]shreyakoulage@gmail.com[5]sajidmomin29@gmail.com

ABSTRACT:

Deaf and dumb people use sign language to communicate. There are various sign recognition techniques that produce output in the form of words for identified signs. The suggested method focuses on proper sentence interpretation of sign language. In addition to sign recognition, several NLP (Natural Language Processing) techniques are applied. The input is a video of sign language, which is then framed and segmented. The Cam Shift method is utilized for tracking, and the P2DHMM algorithm is used for hand tracking.

One of the fastest-growing areas of research is sign language recognition. In this field, many innovative techniques have lately been created. Sign Language is mostly used by deaf and dumb people to communicate. For hearing-impaired people, sign language is the most natural and expressive method. People who are not deaf never attempt to learn sign language in order to communicate with deaf people. Deaf persons get isolated as a result of this. However, if a computer can be programmed to convert sign language into text and voice, the gap between regular people and the deaf community can be narrowed.

Deaf and dumb persons converse via sign language. Various sign recognition techniques generate output in the form of words or recognized signs. The suggested strategy focuses on correctly interpreting sign language in sentences. Several NLP (Natural Language Processing) approachesare used in addition to signing recognition. The input is a video of sign language, which is then framed and segmented.

As a result, deaf persons become isolated. The divide between normal people and the deaf community could be decreased if an Android app could be developed to convert sign language into textual and audio forms. To identify signs, the Haar Cascade classifier was utilized used. After sign recognition, the continuous words for each sign are supplied as input to the POS (Part of Speech) tagging module. A Word Net POS tagger with its own Word Net Dictionary was used. Finally, the LALR Parser is used to frame the phrase. The suggested sign language interpreter model generates understandable sentences in this approach. The gTTS API is used to turn this sentence into audio once more.

Keywords:HaarCascde,gTTS API,HamNoSysSiGML, HamNoSys, Dimensional Hidden Markov Model Algorithm

1. INTRODUCTION

Communication is the process of passing information from one person to another. The majority of the time, people communicate with signs and speech. Normal people utilise natural language to communicate and engage with one another, whereas deaf and dumb persons employ tactile sign language. People with impairments are finding it increasingly difficult to compete in today's world due to fierce competition in every field. According to a report, India has nearly 2.4 million deaf and dumb people, accounting for roughly 20% of the world's total deaf and dumb population. An interpreter is required for hassle-free interaction between normal people and deaf and dumb people.

Visual Sign Language and Tactile Sign Language are the two types of sign language. Hearing and speech disabled people utilise visual sign language, whereas hearing and sight impaired people use tactile sign language. We're mostly focused on deaf and dumb people's visual sign language. Sign language differs from country to country and is influenced by culture. India utilises ISL (Indian Sign Language), America uses ASL (American Sign Language), and China uses CSL (Chinese Sign Language) (Chinese Sign Language). Sign Language is a deaf and dumb communication system made up of a variety of motions generated by hand shapes, body posture, and face expression. Each motion has a certain significance.

In sign language, alphabets are made up of distinct hand forms, and words are made up of hand shapes with varied orientations. Facial expressions are included in complete visual sign language. Deaf and dumb people can communicate effectively using visual sign language. Though this is true, hearing-impaired people must overcome communication barriers in a culture where the majority of people can hear. Visual Sign Language Interaction will be the focus of this study. Natural language is a skill that allows you to comprehend human language. Linguistics and Artificial

Intelligence are both involved.

NLP is a step for developing a system that can convert the text (words) in human language. POS tagging is the method of NLP and first introduced in 1960. It is an important method for language processing. For many NLP applications it is the simplest and most stable step. Part of Speech tagging is the initial step for machine translation, retrieval of information and etc. Second important method in NLP is parsing. Parsing is the method which is followed by the compiler.

The proposed project focuses on converting sign language into proper sentences as well as developing an application.

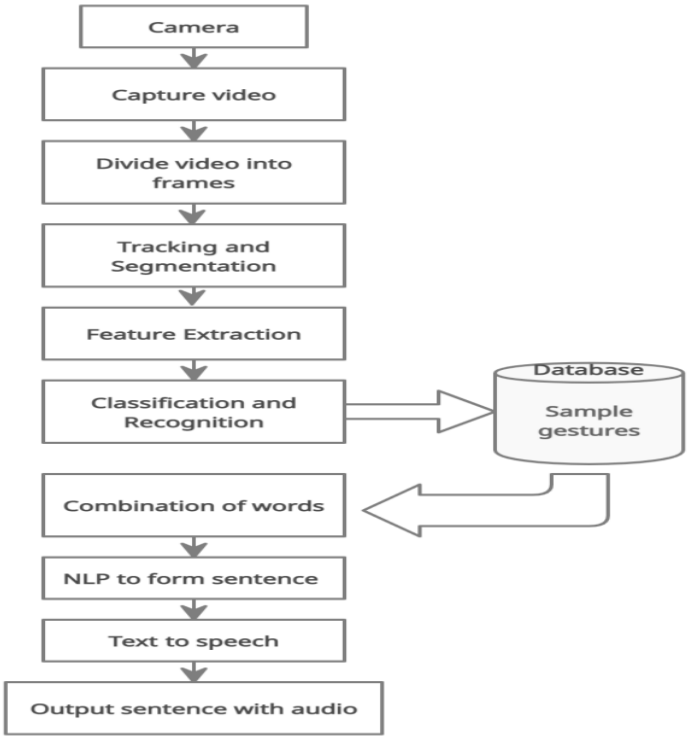


Figure 1. System Architecture of Real Time Sign language Interpreter

The camera on the user's device detects signs made by deaf and dumb people. It is in video format and contains a number of frames; unwanted frames are removed and useful frames are tracked and extracted from thevideo. The posture, hand gestures, and face are now detected and plotted using the MediaPipe Holistic model. These key points are then fed into the A.I model, which predicts a specific sign. Signs are simply words that are used to form meaningful sentences using Natural Language Processing.

2. RECOMMENDED SYSTEM:

Algorithms:

1. LDAAlgorithm:

The Generalization of the Fisher's linear discriminant (FLD) is known as linear discriminant analysis (LDA). LDA mainly used in statistics, pattern recognition and machine learning. It is used to find a linear combination of features that characterizes or separates two or more classes of objects or events. The LDA and FLD are used linear classifier. Its combination also used for dimensionality reduction before later classification. LDA is also closely resembles to principal component analysis (PCA) and factor analysis. Both PCA and factor analysis is linear combinations of variables and they describe the data in a better manner. LDA explains to model the difference between the classes of data. PCA cannot consider the difference in class but factor analysis builds the feature combinations based on differences rather than similarities. There is a difference between Discriminant analysis and factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made. In LDA, the measurements made on independent variables for each observation are continuous quantities. Discriminant correspondence analysis is used to deal with categorical independent variables in LDA

2.SVM is a Support Vector Machine:

According to Wikipedia, SVM is a supervised machine learning model with associated learning algorithm that analyses classification and regression analysis data. In this given a set of training example, we divide data into two classes on the basis of its labelling. If data is labelled it is Putin category of supervised else in the category of unsupervised. When data is labeled then supervised SVM can be used, else SVM is not possible. In case of unsupervised data SVM clustering algorithm is used.

Uses of SVM:

- They are used in text and hypertext classification.
- SVM is used in hand written characters recognition.
- They are used in image classification.

3. Dimensional Hidden Markov Model Algorithm

DCT coefficient of the images is used followed by the selection of Images with the most distinguishable coefficients for P2-DHMMs training Model. So that from the obtained data set, one can extracts the most important information available in subject's images. The algorithm Used for determining the best training images.

3. LITERATURE REVIEW:

1. Indian Sign Language Animation Generation System [1]

   This system was created for Indian Sign Language, as the name implies (ISL). The proposed system takes an English word as input and generates an animation for it. First, HamNoSys based on ISL will be generated in order to generate animation corresponding to words. After that, a HamNoSysSiGML is generated. They used a tool called JA SIGML URL APP to test the accuracy of HamNoSys. The Indian Sign Language Dictionary is used to test the accuracy of animated signs. This system can generate HamNoSys for all basic words encountered in everyday life. They only covered one and two-handed sign symbols.[1]

2. Sign Language Translation [2]

   This system aims at implementing computer vision which can take the sign from the users and convert them into text in real time. The system is divided into four main modules: Image capturing, pre-processing, classification and prediction.

   1. By using image processing the segmentation can be done.
   2. Sign gestures are captured and processed using OpenCV python library.
   3. The captured gesture is resized, converted to grey scale image and the noise is filtered to achieve prediction with high accuracy.
   4. The classification and predication are done using convolution neural network (CNN).

   Aim of this system is to provide communication between normal people and people with hearing disability without need of any specific color background or hand gloves or any sensors. Some systems have used datasets of '.jpg' images. But in the proposed system the pixel values of each image are saved in a csv file which reduces the memory requirement of the system. Also, the accuracy of prediction is high when csv dataset is used.[2]

3. Sign Language Recognition System for Deaf and Dumb People using Image Processing-March-2016.[3]

   This research uses Sign language identification, Hidden Morkov Model, Artificial Neural Network, Data glove, Leap motion controller, Kinetic Sensor. Communication between a deaf-mute and a hearing person has always been difficult . This research examines various approaches to lowering communication barriers by building an assistive gadget for deaf-mute people. The evolution of embedded systems allows for the creation and development of a sign language translation system to assist the deaf. There are a variety of assistant tools available. The major goal is to create a real-time integrated technology that will help physically challenged people communicate more effectively.[3]

4. Hand Gesture Recognition System For the Dumb People [4]

   Authors presented the static hand gesture recognition system using digital image processing. For hand gesture feature vector SIFT algorithm is used. The SIFT features have been computed at the edges which are invariant to scaling, rotation, addition of noise.[4]

5. American Sign Language Recognition System -2018[5]

KshitijBantupalli and Ying Xie developed an American gesture recognition system that uses CNN, LSTM, and RNN to detect video sequences. Inception, a CNN model, was used to extract spatial characteristics from frames, while LSTM was used to extract longer time dependencies and RNN was used to recover temporal features. Various trials were carried out with different sample sizes, and the dataset encompassed 100 different signs done by 5 signers, with a high precision of 93 percent. For longer temporal dependencies, the sequence is passed into an LSTM. To extract temporal features from the softmax layer, the outputs of the softmax layer and the maximum pooling layer are input into an RNN architecture.[5]

4.MODULES

1.Generate and Prepare the Data:

Since we are building this project from the bottom. First thing we need to do is to create the data that we are going to use for training the Neural Network model.For this step we used our computer build-in camera. We captured different gestures for sample signs and they are split in 30folders according to frames. All of the training (prepared) images are stored in *dataset* folder.

2. Generate Features

After the training images are ready we can continue with the next step which is processing all the images and creating the features.

3. Key points Creation:

Following user input, these frames are passed to a function that plots key points and their connections using the mediapipe holistic model and draws these key points on the screen for the user to see using the matplotlib library. Points represent body joints, while lines represent the connection between joints.

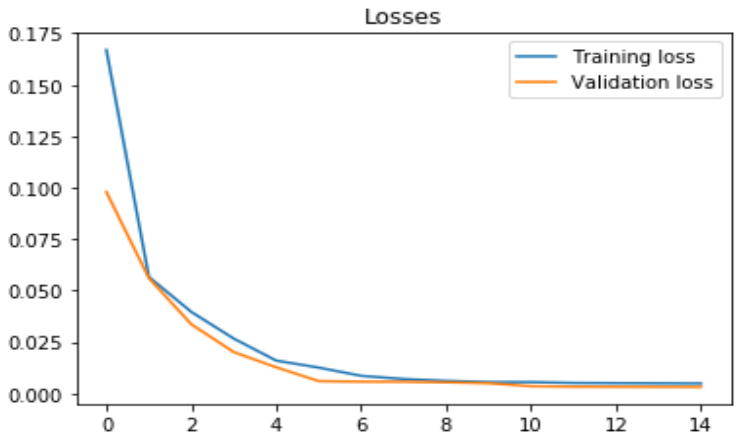
4. Creating Datasets:

We will have a live webcam feed, and every frame that detects a gesture will be created, and key point values will be saved in the form of a numpy array. These numpy array values are kept for various signs. Each sign will contain 30 different numpy arrays. For a single sign there will be 30 different possibilities of signs, each having 30 different numpy arrays. Total 900 numpy arrays will be present for a single sentence.

5. Training a LSTM on the captured dataset:

On the newly generated dataset, we now train an LSTM. To begin, we use keras to load the data. We will create the datasets for this project ourselves because we require datasets in terms of mediapipe holistic keypoint values that are not available on the internet.

Every frame that detects a posture is converted into a numpy array and saved in a directory that includes folders called signs, each of which contains 30 files acquired during dataset creation.The model will then be trained and tested using the flow from directory function, with the names of the number folders serving as the class names for the images loaded.We train LSTM with 21 hidden units. A lower number of units is used so that it is less likely that LSTM would perfectly memorize the sequence. We use the Mean Square Error loss function and Adam optimizer. The learning rate is set to 0.001 and it decays every 5 epochs. We train the model with 100 sequences per batch for 15 epochs. From the plot below, we can observe that training and validation loss converge after the sixth epoch.



Training and validation loss

Fig. Actual accuracy and loss graphs

6. Predicting the Gesture:

Following the successful training of the model with the provided dataset. We now directly convert the camera's live frames into numpy arrays and pass them as input to the model. The model will now predict based on the keypoint values.

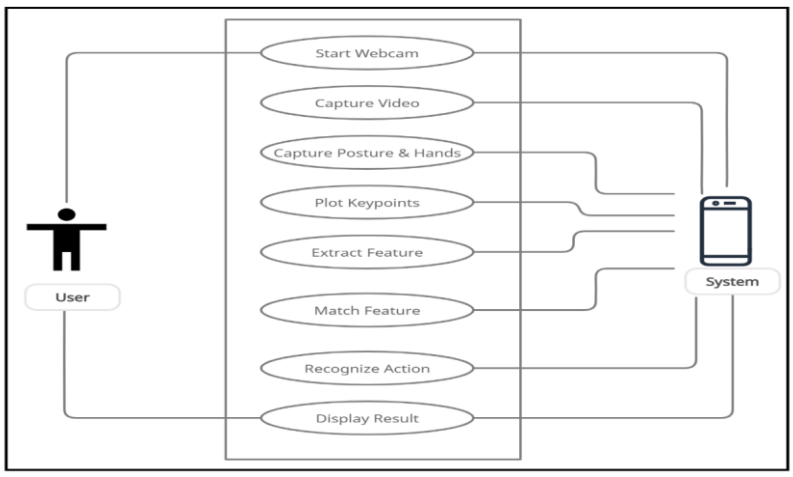


Figure 2.Architecture of Mobile Application

5.RESULT and DISCUSSION:

The training accuracy achieved when training the image dataset without any augmentation was very high (around 90 percent), but the real-time performance was not up to par. Most of the time, it predicted incorrectly because hand-gestures were not precisely centred and vertically aligned in real time. To compensate for this shortcoming, we trained our model by supplementing our dataset. Although the training accuracy was reduced to 89 percent, the real-time predictions were mostly correct. Offline testing of approximately 9000 augmented images revealed 82 percent accuracy.

6. CONCLUSION and FUTURE WORK:

6.1 Conclusion:

We conclude that the communication barrier between deaf and dumb people and normal people can be overcome with the help of AI and ML technology.

In this work, we used various tools to run an automatic sign language gesture recognition system in real-time. Although our proposed work expected to recognize sign language and convert it to text, there is still plenty of room for future work.

6.2 Future Work:

6.2.1 Conversion of text into signs

Now, after using this application the user can understand the meaning of a particular sign language done by the deaf and dumb person. But if he wants to talk to that person, he will not able to talk. This whole communication becomes one way. So, this app can be further upgraded such that when the user enters a message in the application then the application will display a corresponding sign which can be then shown to the deaf and dumb person.

6.2.2 A platform for better communication

A platform can also be created such as a social network platform for various users to communicate with each other. Here instead of showing the application to other person both the users can communicate through their particular mobile application where the other person's message will appear in their specific form. If the user is deaf then he will see various signs and if the user is a normal person then he will see a text message.

ACKNOWLEDGEMENT

We would like to convey our heartfelt thanks and gratitude to one and everyone who has helped us throughout the course of this project. We thanks to our guide, Prof.S.M.Momin for giving us this wonderful opportunity to work on this project. His valuable suggestions and motivations were of immense help.

REFERENCES:

1. 1.Sandeep Kaur, Maninder Singh; Indian Sign Language Animation Generation System; 2015 1st International Conference on Next Generation Computing Technologies (NGCT).
2. Harini; R. Janani, S. Keerthana, S. Madhubala, Venkatasubramanian; Sign Language Translation; 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS).
3. Sign Language Recognition System For Deaf And Dumb People using Image Processing-March-2016
4. Published by Manisha U.Kakde , Mahender G. Nakrani2 , Amit M. Rawate.
5. Sagar P.More, Prof. Abdul Sattar, Hand gesture recognition system for dumb people,
6. K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International
7. Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

2 f