International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 08 | August 2023 www.irjet.net p-ISSN: 2395-0072

Design and Development of Motion Based Continuous Sign Language Detection

Vijayaraghavan R, Hrishi S Kakol, Mithun TP

Student, Dept. of ETE, R V College of Engineering, Bengaluru, Karnataka
Student, Dept. of ETE, R V College of Engineering, Bengaluru, Karnataka
Assistant Professor, Dept. of ETE, R V College of Engineering, Bengaluru, Karnataka

---

Abstract - *Sign language, a structured form of hand gestures incorporating visual motions and signals, is utilised as a communication mechanism to assist the deaf and speech-impaired communities in their daily interactions. Many previous works make use of simpler algorithms rather than efficient ones like MediaPipe holistic to extract features. Also the use of pre-existing data set can limit the accuracy as opposed to developing a custom based one. Real time feed is captured from the webcam and preprocessed by excluding the user's facial features and enhancing only the hands. It is user-friendly because no extra hardware is used. This image is then passed through several layers of a Convolutional Neural Network(CNN) with the usage of an advanced pooling layer, for detection of signs. Gesture recognition is executed with the use of MediaPipe Holistic whose major functionality is detecting feature points of hands and face of the user. The extracted feature points are used to detect the gestures with the help of a Hidden Markov Model (HMM) and Long Short-Term Memory (LSTM) model. Using this methodology, we achieve an epoch categorical accuracy of 91%. With the help of this system, the communication gap between the hearing- and speech-impaired and the general public is meant to be closed.*

Key Words: **ISL, CNN, HMM, MediaPipe Holistic, Image Processing, Machine Learning, Sign Language.**

1. INTRODUCTION

The international federation of the deaf estimates that over 300 sign languages are used by 70 million deaf individuals worldwide. The deaf and speech-impaired community uses sign language as a structured form of hand gestures incorporating visual motions and signals to aid with daily contact. Recognition of sign languages would aid in lowering social obstacles for sign language users. Using this technology, the speech and hearing impaired community can communicate with the rest of the world.

Like spoken language, sign language is not universal and has its own regional variations. Some of the most widely used sign languages worldwide are American Sign Language(ASL), British Sign Language (BSL), Indian Sign Language (ISL), etc. Since most ASL signs aremade with a single hand and are therefore simpler, the majority of studies in this field focus on ASL recognition. The fact that ASL already has a usable standard database is another appealing aspect. ISL is different from sign languages spoken in other countries in terms of syntax, phonology, morphology, and grammar. The Rehabilitation Council of India authorised the teaching materials, ISL grammar, ISL teaching programmes, ISL teacher training courses, and ISL teacher training in 2002. There hasn't been much study done on ISL recognition since the language was just recently established and because tutorials on ISL gestures weren't readily available. Indian Sign Language (ISL) is more dependent on both hands than American Sign Language (ASL), making an ISL recognition system more complicated. The impetus for designing such a useful application sprang from the fact that it would be extremely beneficial for socially assisting individuals as well as raising social awareness. . Systems for sign language development have been created using a variety of methods. These may be broadly divided into sensor-based systems and vision-based systems. Data was collected from various sources and processed in a similar manner in both procedures. The algorithms for extracting indications from photos varied.

Hand shape and hand motions will be retrieved for sign language. As a result, hand characteristics are crucial in hand identification. Fingertips, knuckles, and the palm's centre will be sensed. Various soft computing-based algorithms for gesture detection, such as neural networks, hidden markov models, and long short-term memory (LSTM) models, will be employed using the dataset built particularly for ISL. Fig 1 represents all the signs that are in the ISL.

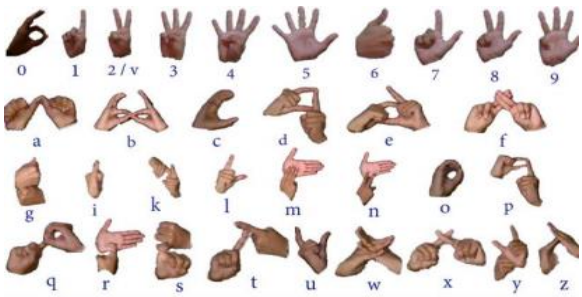

Fig-1: Signs in the Indian Sign Language

2. PROPOSED WORK

To classify dynamic motions, a real-time capture from the webcam at 30 frames per second will be obtained and examined frame by frame. For extracting the hand region from the input video frame, the system will utilise a skin filter. Since HSV colour space is less susceptible to variations in lighting, the image frame will also be transformed into the HSV colour system. From the video frame that was obtained, the ROI will be extracted, and noise will be removed.

To improve the accuracy and detection speed of the system and to aid in subsequent phases of classification, the collected video frame is transformed to a binary picture. This input is fed through a CNN model that is trained with custom dataset, to detect the sign.

ISL includes complex and compound motion based gestures that use both hands and a set of variations within in each. In order to track the motion, two methods were considered. One is the grid based method, where the motion is tracked using relative pixel location of the ROI and subsequently training a HMM model with this data. The other method utilizes feature-points that are extracted using MediaPipe Holistic, which marks the ROI such as fingers, palms and wrists and track the motion of these feature points in a sequence of frames, and train a LSTM model based off of this data, which is used for gesture detection.

Another proposed is translation of English to ISL where a sequence of the signs required to convey the entered message, is displayed to the user.

3. METHODOLOGY

The methodology of the project can be divided into three subsections broadly. They are as follows:

- **Image Processing** - The first subsection deals with feature extraction from the input image or sequence of images. In order to be able to detect hand sings, skin masking is done. A colourful image is transformed into a binary skin mask using this technique. To properly determine if the current pixel is into the skin colour space or not, we employ colour components. As a consequence, a binary image called skin mask is produced. Figure 2 shows the result after skin masking is done.

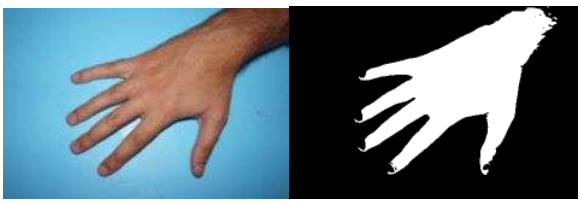

Fig-2: Skin Masking

For gesture recognition, the relative position of hands and fingers is essential. Thus feature points such as fingertips, knuckles, palms and wrist are extracted using MediaPipe Holistic, to further train a LSTM model based on these feature-points.

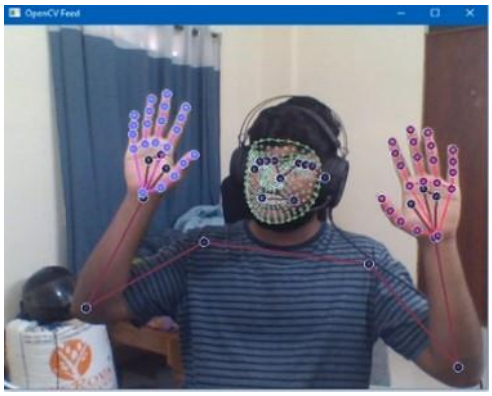

Fig-3: Feature points detected in real-time webcame feed

As seen in Figure 3, there are four basic feature-points, also known as landmarks that we use to orient and classify the image data. They are Pose landmarks, Face landmarks, Right hand and Left hand landmarks.

- **Training of Model** - The present signs are categorized and labeled for future detection, constituting a label map. Further we create TF records for the label map laid out. TF records store the label map in a sequence of binary records The real-time image data is amassed to build a robust data set with real-world scenarios in consideration, in order to make the model more practical and not constrained to formulated testing conditions. It acquires a collection of 100 images for each label – alphabets and numbers, where each image is flipped to make detection more universal. These images are then segmented and labeled to create corresponding ".npy" files which essentially map out the ROI from the entire image into a more computable array format. Tensorflow library is used to create a CNN model with the configuration that is set and using the dataset that was just created, with labels. The dataset is split in a ratio of 70:30 where 70% of the dataset images are used for training the model and 30% of the images are used for testing the model. In order to track the motion, two methods were considered. One is the grid based method, where the motion is tracked using relative pixel location of the ROI and subsequently training a HMM model with this data. The other method utilizes feature-points that are extracted using MediaPipe Holistic, which marks the ROI such as fingers, palms and wrists and track the motion of these feature points in a sequence of frames, and train a LSTM model based off of this data. Based on trial runs, LSTM method proved to be more accurate. Here, the dataset is created by considering a number of

gestures from ISL, and for each gesture – a set of 50 frames is collected in sequence with the feature-points marked and this is iterated 50 times for each gesture. The extracted feature-points (key-points) are formulated in the form of an array with the corresponding co-ordinates of each feature-point. Similar to hand sign recognition, a label map with the gestures chosen for training is created. The extracted data, in the form of an array, that contain the co-ordinates of the feature-points are labelled by associating each frame-sequence data with corresponding entity in the label map.

- **User Interface and Translation of English to ISL -** In order to make this work complete, it only made sense to include translation of English to ISL as well. Here, the program takes a text input (string input) and breaks it down into phrases and words. Further, it checks if there is a gesture available in the repository of the ISL gestures that is stored locally, for each phrase and then breaks down words into individual characters. Once this is done, a window pops up to display the gestures and signs required to represent the input text, in accordance with ISL. This is done using the OpenCV module, by using a locally saved repository of all the signs and gestures available in ISL. Using GUIZero, a minimalistic user interface was developed. Each of the above mentioned components can be easily accessed through this UI, without facing any trouble.

4. IMPLEMENTATION AND RESULTS

The results obtained from hand sign detection for individual alphabets and numbers is as shown below. Figure 4 shows some of the alphabets recognized with a practical background.

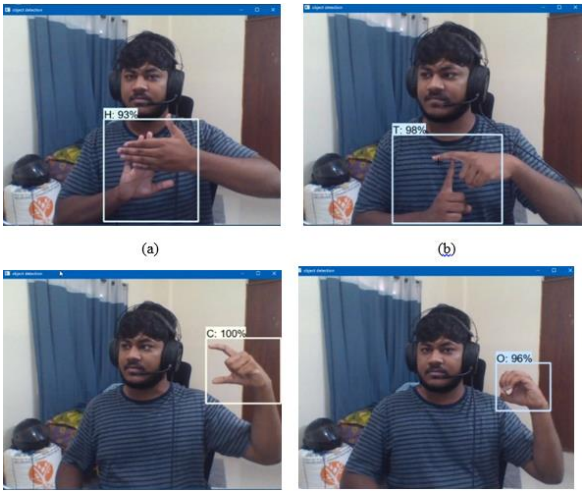

Fig-4: Alphabets as detected by the program

In conjunction of ideal surroundings, background and proper presentation of hand gestures, accuracy levels of above 90% is achievable. Some things to be noted to achieve ideal accuracy is to maintain distinctive colors, emphasis on displaying the hand gestures more than the whole body and to eradicate any disturbances that may arise. Figure 5 shows the confusion matrix for all the alphabets optimized for true positive values which are continually passed through the model for the betterment of results.

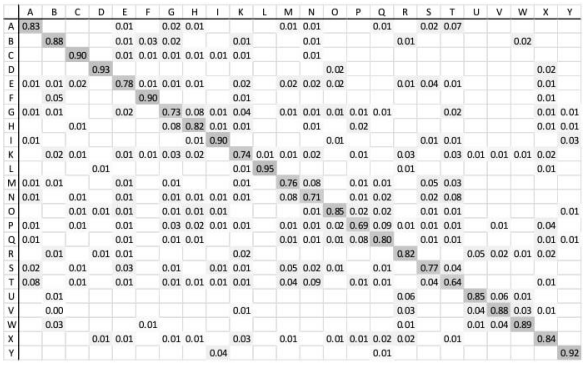

Fig-5: Confusion matrix obtained during testing of Alphabet recognition

The results obtained from compounded hand gestures detection in a motion based scenario is as shown below. Some of the gestures recognized are shown below in Figure 6.

Fig. 6. Motion gestures as detected by the program

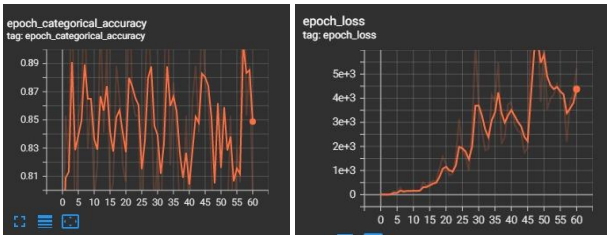

Fig. 7. Epoch Accuracy and Epoch Loss of the trained model

We observe that many of the gestures in ISL, take cues from experiences of reality to simplify some of the commonly used phrases. These may include phrases like “Good afternoon”, “Good morning”, “Sorry”, “Stress”, “Happy birthday”, etc. Thus we can observe that the usage of pre-trained alphabets cannot be implemented here and a new model was developed. This new model has a maximum epoch setting of 5 passes per 50 orientations of feature point arrays. The epoch accuracy and epoch loss values obtained using Tensorboard, are shown below in Figure 7.

5. RELATED WORK

There has been significant amount of work in this sector in order to tackle the difficulties faced by the speech and hearing impaired. Existing solutions have the limitations of requiring the user to wear data gloves or colourful gloves and requiring a clear background with only the user's hand in the frame. Most systems rely solely on hand gestures, while other sign languages additionally include facial expressions to represent some phrases. This remains a significant difficulty in the field of sign language recognition. More focus should be placed on identifying elements that can completely distinguish each sign regardless of hand size, distance from the source, colour features, or lighting circumstances[1][2].Another present work employs inception, a CNN (convolutional neural network) for detecting spatial information, and an RNN (recurrent neural network) for training on temporal variables. It solely utilises the American Sign Language dataset. When shown pictures of users with various skin tones or wearing a wide range of clothing colours, the model suggested here performs less accurately.[3] The alphabet is recognised by segmenting the hand in the collected frames and analysing the position of the fingers. The characteristics of each finger, including the angle between them, the number of them that are totally open, entirely closed, or partially closed, and their individual identity, are used to identify each finger[4][5]. A support vector machine was used to classify the gestures after extracting the hu-moments and motion trajectories from the picture frames. Both a webcam and an MS Kinect were used to test the system[6]. Additionally, evaluation of several models is performed and well discussed. The models that employ convex hull for feature extraction and KNN (K-nearest neighbours) for classification have the greatest accuracy of all the ones that were evaluated, coming in at roughly 65%. Additionally, by employing a larger dataset and a more effective classifier, the accuracy may be improved.[7]. Another system that detects in real-time using grid-based characteristics was proposed. The then-current methods either offered only moderate accuracy or did not operate in real-time, whereas this system focused primarily on increasing accuracy and utilising real-time images. It can only accurately recognise poses and movements made with a single hand and employs a grid-based feature extraction approach.

6. CONCLUSION AND FURTHER PROSPECTS

As previously mentioned, there exists a shortage models and research to modernise the usage of ISL within India as well as in a global platform. Thus the effort towards it in creating a robust system to detect ISL was an impeding necessity. Therefore the original objective to develop such a system to automatically detect and classify ISL was developed without the need of any physical aids. This was achievable to a satisfactory accuracy with the usage of algorithms like CNN, LSTM, HMM and modern feature extraction techniques like the media pipe holistic. In addition to this a custom repository of data towards the ISL was developed to further aid the learning process. Much improvement can be made in developing the system to handle various kind of unsupportive scenarios, grow and train the datasets to a much more accurate percentage and to make the system more accesible and condusive to handle it.

The system presented is limited to Indian Sign Language. This could be expanded to various other systems of sign language around the world. It can also be made to extend support for translation of Sign languages into vernacular languages. Another defining factor for the speech and hearing impaired individuals is to be able to recognize and comprehend speech through lip movements. Including this feature in the system, would make it more robust and convenient. The translation of English into sign language can be improved by customizing a repository of motion gestures incorporated in the Indian Sign Language to be used for its translation.

REFERENCES

[1] Gilorkar, Neelam K. and Manisha M. Ingle. “A Review on Feature Extraction for Indian and American Sign Language.” International Journal of Computer Science and Information Technologies, 2014, pp. 314-318

[2] V.Nair, Anuja & V, Bindu. “A Review on Indian Sign Language Recognition.” International Journal of Computer Applications, 2013, pp. 0975 – 8887

[3] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 4896-4899K. Elissa,

[4] R. K. Shangeetha., V. Valliammai. and S. Padmavathi., "Computer vision based approach for Indian Sign Language character recognition," 2012 International Conference on Machine Vision and Image Processing (MVIP), 2012, pp. 181-184

[5] Tavari, Neha V. and Prof. A. V. Deorankar. “Indian Sign Language Recognition based on Histograms of Oriented Gradient.” International Journal of Computer Science and Information Technologies, 2014, pp. 3657-3660

[6] Raheja, J.L., Mishra A. & Chaudhary A. “Indian sign language recognition using SVM.” Pattern Recognition and Image Analysis, 2016, pp. 434–441

[7] K. Amrutha and P. Prabu, "ML Based Sign Language Recognition System," 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), 2021, pp. 1-6

[8] K. Shenoy, T. Dastane, V. Rao and D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1-9