Real Time Sign Language Detection and Interpretation

Shreyas Shankar - 16IT138
Department of Information technology
National Institute of Technology
Karnataka, India
shreyas.shankar920@gmail.com

Dhvanil Parikh - 16IT217
Department of Information technology
National Institute of Technology
Karnataka, India
dhvanilhparikh@gmail.com

Madhusudhan NH - 16IT2225
Department of Information technology
National Institute of Technology
Karnataka, India
madusudannh010689@gmail.com

Abstract - A huge number of individuals around the globe experience the ill effects of hearing incapacity. This expansive number exhibits the significance of building up a sign language recognition system framework changing over communication through signing to content for sign language communication to progress toward becoming more clear to comprehend without an interpreter. Sign Language is the most common and expressive route for the meeting hindered. This execution shows a technique which perceives the Indian Sign Language (ISL) and converts into an ordinary text content. It is utilizing inception model. To enable communication between speech impaired and visually impaired people, a simple text to speech function is used to speak out the sentences interpreted. Exploratory outcomes give agreeable division of hints under assorted foundations and moderately high precision in gesture based communication acknowledgment. The inception model was trained on self produced datasets for various signs. The exactness of the signs distinguished was over 90% and the time taken was negligible.

Introduction

Sign Language is the methods for correspondence among the deaf and dumb network. Communication via sign Language develops and advances normally inside hearing impeded network. Sign Language correspondence includes manual and non-manual signs where manual signs include fingers, hands, arms and non-manual signs include face, head, eyes and body. Communication via sign Language is a well-organized language with a phonology, morphology, linguistic structure and punctuation/grammar. Sign Language is a finished common language that utilizes distinctive methods for articulation for correspondence in regular day to day existence. Communication through signing contrasts from different dialects as it has no expressed word. The structure of the verbally expressed language utilizes words consecutively though a gesture based communication utilizes various body developments in parallel. Sign Language recognition framework exchanges the correspondence from human-human to human-computer interaction. Sign Language mediators are utilized by deaf and dumb individuals to speak with the conference world. The point of the sign language based communication acknowledgment framework is to show an effective and precise component to translate content, in this way the "dialog communication" between the deaf and hearing individual will be smooth. There is no standardized communication via sign language for all deaf individuals over the world. In any case, communications via sign language are not widespread, similarly as with spoken dialects, these vary from region to region.

There are two primary methodologies utilized in the sign language based communication acknowledgment that is Glove/Device based and Vision based. In the glove based technique the client needs to wear a gadget which conveys a heap of links in order to interface the gadget to a PC. Such gadgets are costly and diminish the instinctive nature of the gesture based communication correspondence. Conversely, the Vision based technique requires just a camera and legitimately manages picture motions. It is a two stage process: sign catching and sign examination. Vision based techniques give a regular habitat to the client and diminishes the intricacies as in the glove based strategy. Here we have utilized vision based for detection.

Each nation has its own sign language based communication with an abnormal state of syntactic varieties. The sign language which exists in India is normally known as Indian Sign Language (ISL). It has been contended that maybe a similar communication via gestures is utilized in Nepal, Sri Lanka, Bangladesh, and fringe districts of Pakistan.

Communication is one of the essential necessity for survival in the public arena. Deaf and dumb individuals convey among themselves utilizing communication through signing however ordinary individuals think that it's hard to comprehend their language. Broad work has been done on American communication through sign language recognition yet Indian gesture based communication contrasts fundamentally from American gesture based communication. ISL utilizes two hands for communicating(20 out of 26) while ASL utilizes single hand for imparting. Utilizing two hands frequently prompts lack of definition of highlights because of covering of hands. Moreover, absence of datasets alongside change in communication through signing with area has brought about controlled endeavors in ISL motion discovery. Our task goes for making the fundamental stride in crossing over the correspondence hole between ordinary individuals and hard of hearing and unable to speak individuals utilizing Indian gesture based communication. Successful expansion of this task to words and typical statements may not just influence the deaf and dumb to impart quicker and simpler with external world, yet in addition give a lift in creating self-ruling frameworks for comprehension and supporting them.

The major contributions of this implementation are as follows :-

- A novel system to aid in communicating with those having speech and vocal disabilities
- A real-time approach to bare hand detection using skin color segmentation with minimal noise and false positives.
- An improvised and fast method to hand posture detection.

Our proposed model was trained on self generated datasets for different signs.

The execution worked very well when tried for correspondence by framing sentences distinguished by the model. The precision of the signs identified was over 90% and the time taken was negligible.

Literature Survey

1. Chuan-Kai Yang, Quoc-Viet Tran and Vi N.T. Truong have proposed a framework perceiving static hand indications of letter sets in American Sign Language from live recordings and converting into content and discourse. AdaBoost and Haarlike classifiers have been utilized for the classification during training process. After the training procedure, the classifier can perceive diverse hand stances. Procedure of testing the framework comprises of three phases: preprocessing stage, classification stage and text to speech arrange. In "Preprocessing Stage", frames from the video stream are separated and techniques for image processing are utilized to acquire the features from the picture. In "Classification Stage", the preprocessed pictures in the preprocessing stage are utilized as info and classification is finished by utilizing Haar Cascade Algorithm. In "Text To Speech Stage", content perceived by the classifier is changed over to speech by utilizing SAPI 5.3. Performance measures of the framework are: 98.7% precision, 99.9% specificity, 98.7% recall, 98.7% F-score and 98.7% sensitivity.

2. Anis Diyana Rosli, Adi Izhar Che Ani, Mohd Hussaini Abbas, Rohaiza Baharudin, and Mohd Firdaus Abdullah have proposed a spelling glove work perceiving the letters of American Sign Language letters in order. The framework has been planned focusing on deaf and dumb individuals to speak with ordinary individuals. Right off the bat, the letter set of the sign based communication is framed by the designed glove. At the point when the communication through signing is framed, the bending sensor recognizes the situation of each finger and yields different resistance value. At that point Microcontroller that is associated with the spelling glove arranges the situation of bowing of finger dependent on yield voltage created. After Microcontroller finds the mix position of each finger in library, LCD shows right letter set. Recognition rate of the framework is 70%.

3. Md. Mohiminul Islam, Sarah Siddiqua and Jawata Afnan have proposed another Hand Gesture Recognition consider dependent on American Sign Language. The framework works in four stages for signal acknowledgment including image acquisition, preprocessing, feature extraction and feature

recognition. In "Image Acquisition" step, a database of 1850 pictures of 37 signs is made by gathering picture tests of each indication of the sign based communication from various individuals. "Preprocessing" step readies the picture got from camera for feature extraction venture by expelling noise and trimming picture to get portion of wrist and fingers of a hand for sign location. "Feature extraction" step applies diverse calculations for feature extraction of sign language recognition framework including K convex hull for fingertip identification, eccentricity, elongatedness, pixel segmentation and rotation. Artificial Neural Network, Backpropagation Algorithm is utilized for training. Gesture acknowledgment rate of the framework is 94.32%.

4.Jun-Wei Hsieh, Teng-Hui Tseng, Wan-Yi Yeh and Chun-Ming Tsai have proposed a gesture based communication acknowledgment framework so as to identify English letters and numbers. The color information, skeleton information and depth information which are obtained from the input Kinect are utilized for recognizing palm region of hands. At that point Otsu thresholding strategy is utilized for extracting palm and morphology shutting operation is utilized for shutting the openings in the palms. At that point SURF descriptors and features are separated.

5. Chana Chansri and Jakkree Srinonchat have introduced an investigation of perceiving Thai gesture based communication. The proposed framework gets the color and depth data from the Kinect sensor for hand identification. At that point Histograms of Oriented Gradients method is utilized for extracting features of pictures. At long last the separated highlights are prepared by utilizing Neural Network. The accuracy rates are gotten from various separations from Kinect sensor, for example, 08.m, 1.0m and 1.2 m are arranged by 83.33%, 81.25%, 72.92%.

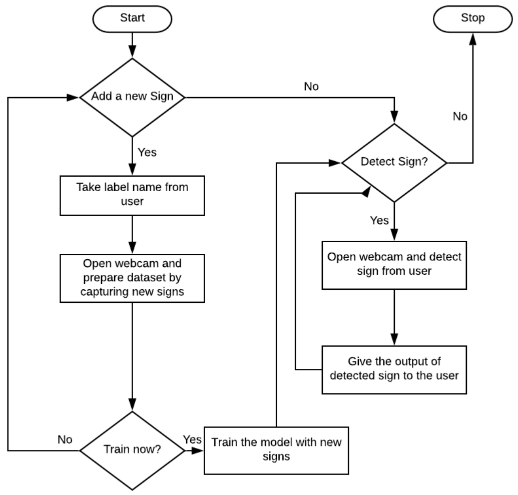
Author	Work Done	Takeaways
Chuan-Kai Yang, Quoc-Viet Tran and Vi N.T. Truong	Used AdaBoost and Haarlike classifiers and Haar Cascade Algorithm for classification.	Frames from the video stream are extracted and methods of image processing are used to obtain the features from the image and then fed to the classifier.
Anis Diyana Rosli, Adi Izhar Che Ani, Mohd Hussaini Abbas, Rohaiza Baharudin, and Mohd Firdaus Abdullah	Proposed a spelling glove work recognizing the letters of American Sign Language alphabet	No significant take away as we have used vision based sign detection instead of glove based.
Md. Mohiminul Islam, Sarah Siddiqua and Jawata Afnan	Their system works in four steps for gesture recognition including image acquisition, preprocessing, feature extraction and feature recognition	Image acquisition i.e. a database of 4350 images of 29 signs is created by collecting image samples of each sign of the sign language from different people.
Jun-Wei Hsieh, Teng-Hui Tseng, Wan-Yi Yeh and Chun-Ming Tsai	The color data, skeleton data and depth data which are obtained from the input Kinect are used for detecting palm area of hands	Bare hand detection using skin color segmentation with minimal noise and false positives.
Chana Chansri and Jakkree Srinonchat	Histograms of Oriented Gradients technique is used for feature extraction of images. Finally the extracted features are trained.	Inspired us to use a CNN based Inception model for training and classification of the signs.

Problem Statement and Objectives

Problem of creating an interface for Indian sign language for people with hearing and speech impairment based on Real Time Personalised Sign Language Detection using a Google inception model.

- Survey the currently existing methods for sign language translation.
- Implement a resource efficient model to classify thesigned language on a real-time scale.
- The objective of the proposed work is to efficiently recognize the signs of Indian Sign Language (ISL).
- Translate the accurate meaning of the recognizable signs.
- Indian Sign Language is a combination of invariant moments and shape descriptor features.We have to consider only manual signs which comprises of hand gesture of isolated signs which is a difficult task.

Methodology


Start
Stop
Add a new Sign
No
No
Yes
Detect Sign?
Take label name from user
Yes
Open webcam and detect sign from user
Open webcam and prepare dataset by capturing new signs
Give the output of detected sign to the user
No
Train now?
Yes
Train the model with new signs

The system consists of Inception model. The pre-trained Inception-v3 model achieves state-of-the-art accuracy for recognizing general objects with 1000 classes, like "Zebra", "Dalmatian", and "Dishwasher". The model extracts general features from input images in the first part and classifies them based on those features in the second part.This model is responsible for predicting the word in the vocab, which is being signed in the frames captured by the webcam.

There are a great deal of exceptionally vigorous and precise communication via gestures interpretation frameworks accessible, as talked about in the writing survey. Be that as it may, we chose utilizing Inception model over those for the principle reason of speed and on gadget preparing and deduction. Different models are incredibly extensive and prepared on a specific variation of communication through signing, and in this way can't be customized by the client. Utilizing Inception model enables us to give the clients a chance to prepare without anyone else vocabulary with couple of precedents and for their specific use case. Also the prepared vocabulary is spoken out by the interface aimed to cater visually impaired people. This application is likewise prepared to recognize Indian Sign Language and furthermore client can prepare his/her own behavior.

Work Done

For the implementation of the application we use a webcam device to capture the signs. The implementation details are as follows:

The application was implemented as a Graphical User Interface to make it easily accessible for any user, without download and installation. The application is a single page python application.Tkinter is used for GUI. Python sound api is used for text to speech.

The visual feed for the system is captured using the Browser Webcam API. The frames at a set interval are then processed by the model. The Inception model is implemented in Tensorflow. During the training phase the images are captured and stored in the memory temporarily. These images with their corresponding label are then fed together to the model. During inference a frame is sent to the model at a fixed intervals. All the data is collected and used within the directory of the particular user itself so the problem of user data privacy is alleviated.

The parameters used were as follows:

Dataset - 150 images of each sign generated using image acquisition discussed in the literature survey.

Bottlenecks - A bottleneck generated for each image by extracting the hand occupied area from the image frame.

Training steps - 500

Model used - Inception

Type	Patch size/ Stride	Input size
conv	3 x 3 / 2	299 x 299 x 3
conv	3 x 3 / 1	149 x 149 x 32
conv padded	3 x 3 / 1	147 x 147 x 32
pool	3 x 3 / 2	147 x 147 x 64
conv	3 x 3 / 2	71 x 71 x 80
conv	3 x 3 / 1	35 x 35 x 192
pool	8 x 8	8 x 8 x 2048
linear	logits	1 x 1 x 2048
softmax	classifier	1 x 1 x 1000

Results and Analysis

The application was built successfully. The GUI implemented was simple and unobtrusive to ensure that it is easy for anyone to get started. The application can function on low end hardware and requires only a webcam and a microphone to function. The above figures contain snapshots of the application interface. The interface is clean and easy to follow with informative instructions for the user.

For the analysis of the application, we conducted tests on a group of 7 individuals who had never used the application before. In the test the individuals were not allowed to speak directly to the smart home device and were required to use the application as an interface to the devices. The individuals were not given any background information regarding how to use the application.

Each individual was asked to perform 5 different queries to the system, each with a different difficulty level. The 5 queries were designed to be representative of the real world usage scenario for the application. The queries consisted of simple interactions to common queries to complicated queries.

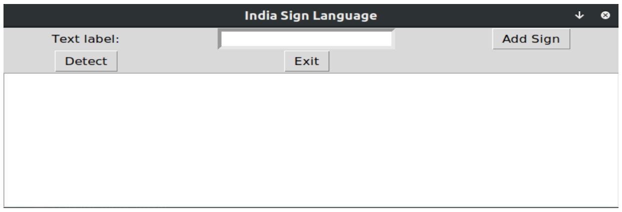
The 5 queries are as follows:-

1. Is it simple to use?
2. Can we add more signs in a simple manner?
3. How much time does it take to add a new sign?
4. Is the application predicting the sign language properly?
5. Is the resulting sentence spoken out properly properly?

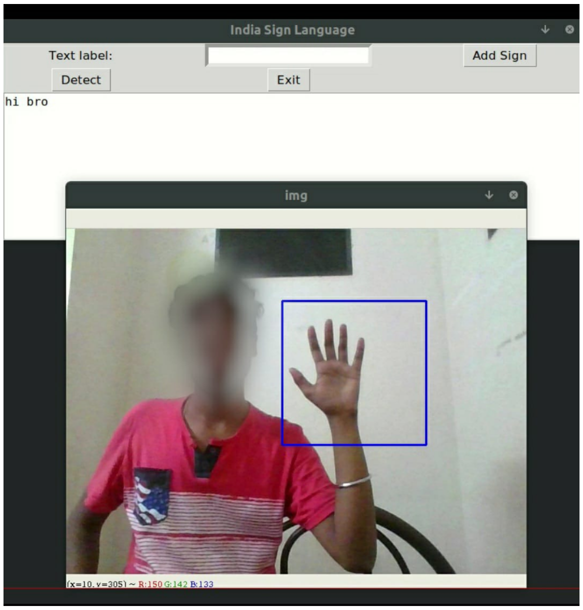
The clients were required to begin starting with no outside help and add their sign to the application, and afterward play out the given inquiry. The questions were given to them one by one and were in expanding request of trouble. The time taken by the client was noted from the season of beginning the question to the time the whole inquiry was finished. A period limit was additionally authorized, past which the errand was esteemed unsuccessful.At the finish of the test, the people were given a structure to get criticism on the application. The form consisted questions with 5-scale response options. The questions ranged from ”Satisfaction with the user interface” to ”Satisfaction with the predicted results” to ”Overall satisfaction with the result”. The clients were additionally furnished with a content field to write in a particular criticism they had on their execution. Figure demonstrates the consequences of the adequacy investigation. The diagram shows the level of clients that had the capacity to finish the given queries. The queries expanded in trouble from Query 1 to Query 5. We can see that the level of clients who finished the given undertaking diminished with the undeniably unpredictable queries. This is normal conduct for new clients. In any case, another fascinating perception is that for the last question the level of individuals who finished the assignment expanded. One conceivable clarification is that the clients turned out to be increasingly more acquainted with the framework as they continued utilizing it and towards the end had the capacity to make even complex questions effectively. This demonstrates is simpler for the clients to adjust to the application.The client timing subtleties are shown in Fig. 6. We can see that couple of clients (2, 4, 5) had the capacity to finish the given undertaking in a sensible measure of time. They had the capacity to comprehend the framework and figure out how to utilize it rapidly. In any case, we can likewise see that in numerous assignments clients were not ready to finish the errand by any means. The clients achieved the set time point of confinement and subsequently the endeavor was considered as a fail.

The time efficiency of the system was calculated to be 0.151 goals/sec and the overall relative efficiency was around 51%. The failed tasks indicate that the system was not easy enough.Time delay is around 2 seconds.
To learn quickly and get started using it. We also observe that users who have some background knowledge in working with computers were able to work with the system much better
and were able to finish most of their tasks within the given time limit. This is expected behaviour since the users who have prior experience with computers can easily learn the new interface.

We now discuss the responses of the Five-Scale Likert Analysis form filled by the individuals. There form consisted of 5 questions which asking for the satisfaction of the user for each aspect of the application from the Ease of getting started to the User Interface to the ease of use. We observe that the users who were familiar with aspects of computers and who were able to complete the task provided positive feedback. On the other hand users who were not able to complete the tasks provided negative feedback. This is expected behaviour since the system must adapt to all types of users and if the users are not satisfied the feedback will be negative. So from the analysis of the tests conducted and user feedback from the data obtained, we can observe that the application is effective and efficient for users who are familiar with the computers in general but there is room for improvement for users who are not familiar with the technology.



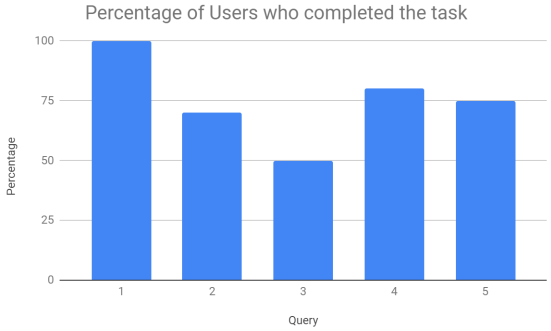
GUI Interface



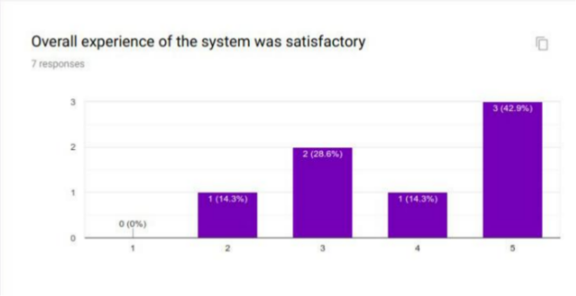
```
hi bro
```

```
(x=10, y=305) ~ R:150 G:142 B:133
```

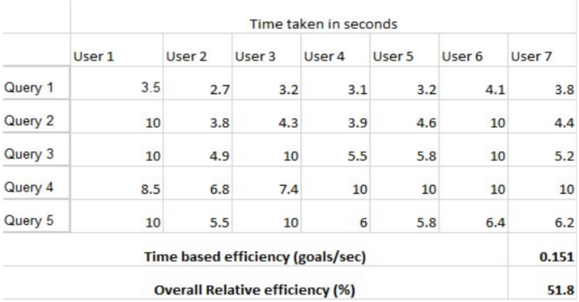
Real Time Sign Language Detection



Analysis of User Queries



Overall Experience



	Time taken in seconds						
	User 1	User 2	User 3	User 4	User 5	User 6	User 7
Query 1	3.5	2.7	3.2	3.1	3.2	4.1	3.8
Query 2	10	3.8	4.3	3.9	4.6	10	4.4
Query 3	10	4.9	10	5.5	5.8	10	5.2
Query 4	8.5	6.8	7.4	10	10	10	10
Query 5	10	5.5	10	6	5.8	6.4	6.2
Time based efficiency (goals/sec)							0.151
Overall Relative efficiency (%)							51.8

Time Efficiency

Limitations

The model uses very high computation power to train and detect the signs. So we had to restrict the number of images per sign to 150. More number of images would have resulted in more accuracy at the cost of high computation power and more time. Also, as the training set included just 150 images per sign, it was difficult for the model to classify extremely similar signs like ‘d’ and ‘p’ accurately, but was able to classify them significantly.

Conclusion

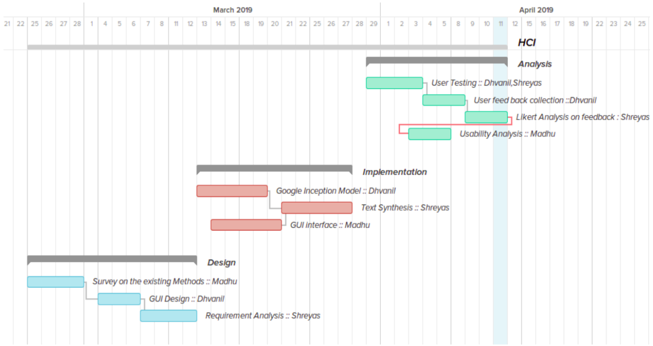
There has been a surge in the number of available and accessible sign language detection applications. These applications are generally used by physically challenged persons and can be accessible to people with disabilities(dumb). In this project we proposed an interface that would detect sign language and show/tell us what they want to convey . The system consists of a Google Inception model which recognizes the symbols from webcam footage and also trains dynamically. We implemented the proposed system in a python GUI. The system prints the interpreted statements in the python GUI and spoken out using text to speech. This makes the system capable to aid communication between speech impaired people and normal or visually impaired people. We conducted tests on the application on a control group and performed various analysis on the received feedback from the users.

The work done was able to create a novel system to aid in communicating with those having speech and vocal disabilities.The achievement of a real-time approach to bare hand detection using skin color segmentation with minimal noise and false positives was significant. The significance of the proposed approach was an improvised and fast method to hand posture detection.

Acknowledgement

The authors would like to thank the individuals who participated in the testing phase and provided valuable feedback on the application.

Individual Contribution



Gantt Chart

The contributions of each member can be seen in the the chart.Dhvanil worked on the survey of existing methods, implementation of the User Interface and the Sign Language Recognition Model, the user testing on control group and Google Inception Model. Shreyas worked on the Likert Analysis, user testing on control group and Text Synthesis. Madhu worked on the survey of existing methods, design of the user interface, useability analysis and user testing on control group.

References

1. “Deafness and hearing loss” World Health Organization, [online] Available: http://www.who.int/mediacentre/factsheets/fs3001/en/.

2. V. N. T. Truong, C. K. Yang, Q. V. Tran, "A translator for American sign language to text and speech", 2016 IEEE 5th Global Conference on Consumer Electronics, pp. 1-2, 2016.

3. Yi Li, "Hand gesture recognition using Kinect", 2012 IEEE International Conference on Computer Science and Automation Engineering, pp. 196-199, 2012.

4. P. Subha Rajama, G. Balakrishnan, "Recognition of tamil sign language alphabet using image processing to

aid deaf-dumb people", Science Direct Procedia Engineering, vol. 30, no. 4, pp. 861-868, 2012.

5. R. K Shangeetha, V Valliammai, S Padmavathi, "Computer vision based approach for Indian sign language character recognition", IEEE International conference communication technology and system design, vol. 30, pp. 861-868, 2012.

6. Xiaohui Shen, Gang Hua, Lance Williams, Ying Wu, "Dynamic hand gesture recognition: An exemplar-based approach from motion divergence fields", Science Direct Image and Vision Computing, vol. 30, no. 3, pp. 227-235, 2012.