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 - Millions of people around the world suffer from hearing disability. This large number demonstrates the importance of developing a sign language recognition system converting sign language to text for sign language to become clearer to understand without a translator. Sign Language is the most natural and expressive way for the hearing impaired. This implementation presents a methodology which recognizes the Indian Sign Language (ISL) and translates into a normal text. It is implemented using inception model. Experimental results show satisfactory segmentation of signs under diverse backgrounds and relatively high accuracy in sign language recognition. The inception model was trained on self generated datasets for different signs. The accuracy of the signs detected was above ninety percent and the time taken was minimal.

Introduction

Sign Language is the means of communication among the deaf and mute community. Sign Language emerges and evolves naturally within hearing impaired community. Sign Language communication involves manual and non-manual signals where manual signs involve fingers, hands, arms and non-manual signs involve face, head, eyes and body. Sign Language is a well-structured language with a phonology, morphology, syntax and grammar. Sign language is a complete natural language that uses different ways of expression for communication in everyday life. Sign Language differs from other languages as it has no spoken word. The structure of the spoken language makes use of words sequentially whereas a sign language makes use of numerous body movements in parallel. Sign Language recognition system transfers the communication from human-human to human-computer interaction. Sign language interpreters are used by deaf and dumb people to communicate with the hearing world. The aim of the sign language recognition system is to present an efficient and accurate mechanism to transcribe text, thus the “dialog communication” between the deaf and hearing person will be smooth. There is no standardized sign language for all deaf people across the world. However, sign languages are not universal, as with spoken languages, these differ from region to region.

Sign language consists of making shapes or movements with your hands with respect to the head or other body parts along with certain facial cues. A recognition system would thus have to identify specifically the head and hand orientation or movements, facial expression and even body pose.

There are two main approaches used in the sign language recognition that is Glove/Device based and Vision based. In the glove based method the user has to wear a device which carries a load of cables so as to connect the device to a computer. Such devices are expensive and reduce the naturalness of the sign language communication. In contrast, the Vision based method requires only a camera and directly deals with image gestures. It is a two step process: sign capturing and sign analysis. Vision based methods provide a natural environment to the user and reduces the complications as in the glove based method. Here we have used vision based for detection.

Every country has its own sign language with a high level of grammatical variations. The sign language exists in India is commonly known as Indian Sign Language (ISL). It has been argued that perhaps the same sign language is used in Nepal, Sri Lanka, Bangladesh, and border regions of Pakistan.

Communication is one of the basic requirement for survival in society. Deaf and dumb people communicate among themselves using sign language

but normal people find it difficult to understand their language. Extensive work has been done on American sign language recognition but Indian sign language differs significantly from American sign language. ISL uses two hands for communicating(20 out of 26) whereas ASL uses single hand for communicating. Using both hands often leads to obscurity of features due to overlapping of hands. In addition to this, lack of datasets along with variance in sign language with locality has resulted in restrained efforts in ISL gesture detection. Our project aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using Indian sign language. Effective extension of this project to words and common expressions may not only make the deaf and dumb people communicate faster and easier with outer world, but also provide a boost in developing autonomous systems for understanding and aiding 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 implementation worked quite well when tested for communication by forming sentences detected by the model. The accuracy of the signs detected was above ninety percent and the time taken was minimal.

Literature Survey

1. Chuan-Kai Yang, Quoc-Viet Tran and Vi N.T. Truong have proposed a system recognizing static hand signs of alphabets in American Sign Language from live videos and translating into text and speech. AdaBoost and Haarlike classifiers have been used for the classification during training process. After the training process, the classifier can recognize different hand postures. Process of testing the system consists of three stages: preprocessing stage, classification stage and text to speech stage. In “Preprocessing Stage”, frames from the video stream are extracted and methods of image processing are used to obtain the features from the image. In “Classification Stage”, the processed images in the preprocessing stage are used as input and classification is done by using Haar Cascade Algorithm. In “Text To Speech Stage”, text recognized by the classifier is converted to speech by using SAPI 5.3. Performance measures of the system are: 98.7% precision, 98.7% recall, 98.7% sensitivity, 99.9% specificity and 98.7% F-score.

2. Anis Diyana Rosli, Adi Izhar Che Ani, Mohd Hussaini Abbas, Rohaiza Baharudin, and Mohd Firdaus Abdullah have proposed a spelling glove work recognizing the letters of American Sign Language alphabet. The system has been designed targeting deaf-mute people to communicate with normal people. Firstly, the alphabet of the sign language is formed by the designed glove. When the sign language is formed, the bending sensor detects the position of each finger and yields various resistance value. Then Microcontroller that is connected to the spelling glove categorizes the position of bending of finger based on output voltage produced. After Microcontroller finds the combination position of each finger in library, LCD displays correct alphabet. Recognition rate of the system is 70%.

3. Md. Mohiminul Islam, Sarah Siddiqua and Jawata Afnan have proposed another Hand Gesture Recognition study based on American Sign Language. The system works in four steps for gesture recognition including image acquisition, preprocessing, feature extraction and feature recognition. In “Image acquisition” step, a database of 1850 images of 37 signs is created by collecting image samples of each sign of the sign language from different people. “Preprocessing” step prepares the image received from camera for feature extraction step by removing noise and cropping image to obtain portion from wrist to fingers of a hand for sign detection. “Feature extraction” step applies different algorithms for feature extraction of hand gesture recognition system including K convex hull for fingertip detection, eccentricity, elongatedness, pixel segmentation and rotation. Artificial Neural Network, Backpropagation Algorithm is used for training. Gesture recognition rate of the system is 94.32%.

4. Jun-Wei Hsieh, Teng-Hui Tseng, Wan-Yi Yeh and Chun-Ming Tsai have proposed a sign language recognition system in order to detect English letters and numbers. The color data, skeleton data and depth data which are obtained from the input Kinect are used for detecting palm area of hands. Then Otsu thresholding method is used for extracting palm and morphology closing operation is used for closing the holes in the palms. Then SURF descriptors and features are extracted.

5. Chana Chansri and Jakkree Srinonchat have presented a study of recognizing Thai sign language. The proposed system receives the color and depth information from the Kinect sensor for hand detection. Then Histograms of Oriented Gradients technique is used for feature extraction of images. Finally the extracted features are trained by using Neural Network. The accuracy rates are obtained from different distances from Kinect sensor such as 08.m, 1.0m and 1.2 m are in order of 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 by using Neural Network	Inspired us to use a CNN based Inception model for training and classification of the signs.

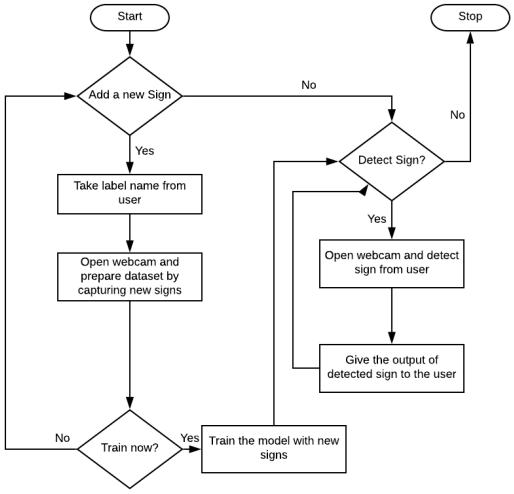
Problem Statement and Objectives

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



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 lot of very robust and accurate sign language translation systems available, as discussed in the literature review. However, we decided on using Inception model over those for the main reason reason of speed and on device training and inference. The other models are extremely large and trained on a particular variant of sign language, and thus cannot be personalized by the user. Using Inception model allows us to let the users train on their own vocabulary with few examples and for their particular use case. This application is also trained to detect Indian Sign Language and also user can train his/her own actions.

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. 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. Are the signs detected quickly?

The users were required to start from scratch and add their sign to the application, and then perform the given query. The queries were given to them one by one and were in increasing order of difficulty. The time taken by the user was noted from the time of starting the query to the time the entire query was completed. A time limit was also enforced, beyond which the task was deemed unsuccessful.At the end of the test, the individuals were provided with a form to get feedback 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 users were also provided with a text field to write in any specific feedback they had on their performance. Figure shows the results of the effectiveness analysis. The graph displays the percentage of users that were able to complete the given queries. The queries increased in difficulty from Query 1 to Query 5. We can see that the percentage of users who completed the given task reduced with the increasingly complex queries. This is expected behaviour for new users. However, another interesting observation is that for the last query the percentage of people who completed the task increased. One possible explanation is that the users became more and more familiar with the system as they kept on using it and towards the end were able to make even complex queries easily. This indicates that it is easier for the users to adapt to the application.The user timing details are displayed in Fig. 6. We can observe that few users (2, 4, 5) were able to complete the given task in a reasonable amount of time. They were able to understand the system and learn how to use it quickly. However, we can also observe that in many tasks users were not able to complete the task at all. The users reached the set time limit and thus the attempt 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.

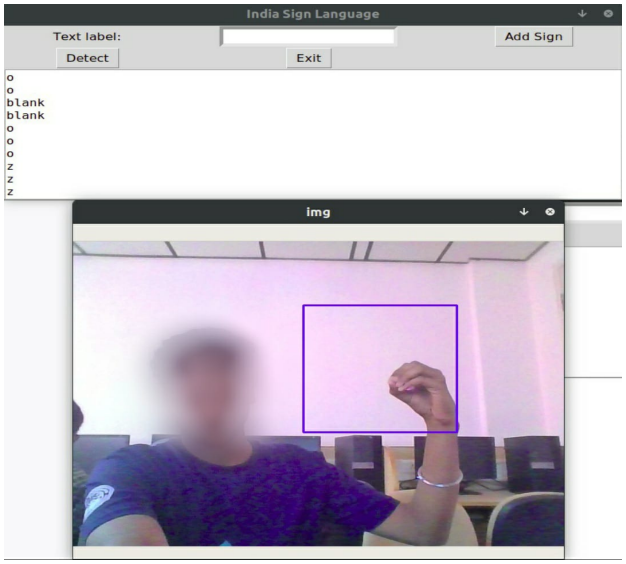
To learn quickly and get started using it. We also observe that users who have some background knowledge in working with computers were able to work with the system much better

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.

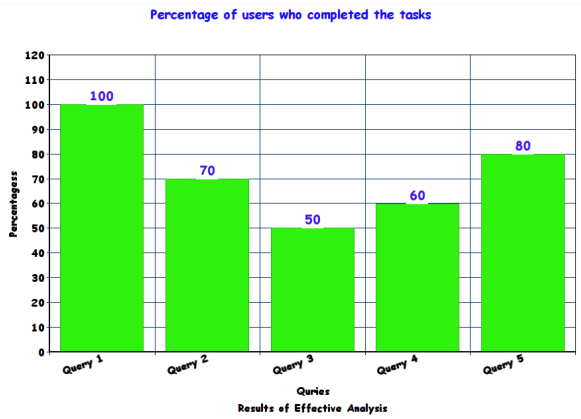
We now discuss the responses of the Five-Scale Likert Analysis form filled by the individuals. There form consisted of 5 questions which asking for the satisfaction of the user for each aspect of the application from the Ease of getting started to the User Interface to the ease of use. We observe that 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.

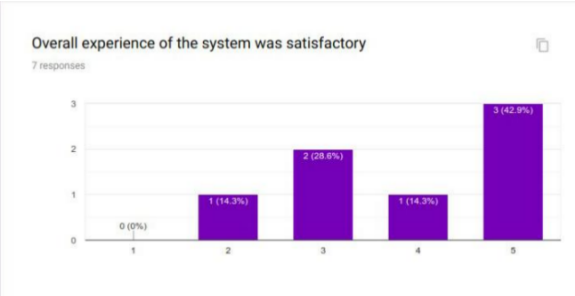
GUI Interface



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

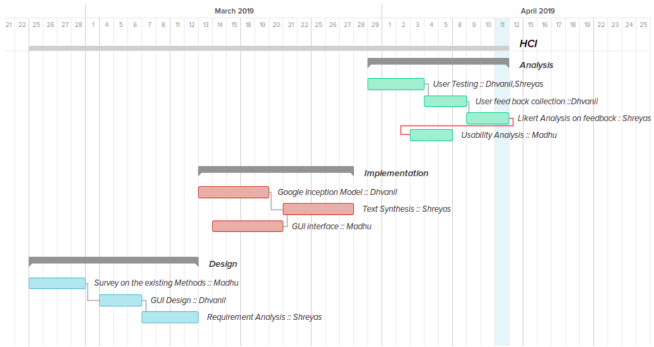
There has been a surge in the number of available and accessible sign language detection applications. These applications are generally used by physically challenged persons and can be accessible to people with disabilities(dumb). In this project we proposed an interface that would detect sign language and 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. 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.