Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

DETECTION OF SIGN LANGUAGE ON VIDEO CONFERENCING PLATFORMS

Miss Radhika Sharma, B.Tech Final year Student, Dept. Of Computer Science Engineering, Medi-Caps University, Indore, M.P.
Miss Varsha Kothari, Assistant Professor, Dept. Of Computer Science Engineering, Medi-Caps University, Indore, M.P.

Abstract
Three years since the outbreak of Novel Coronavirus, the world is still amid the pandemic. Everything shifted from traditional collocated work to remote work. One of the reasons behind this successful shift is the increased use of video conferencing platforms as a means of communicating. Sadly, with this fast-changing technological life, people with speech and hearing impairment are usually forgotten and left out. They have to struggle to bring out their ideas and voice their opinion. Sign Language although a medium of communication between deaf and mute people still has no meaning when conveyed to a non-sign language user. Hence, broadening the communication gap. To bridge this gap, we demonstrate a demo application of sign language detection over a web browser to showcase its usage possibility over video conferencing platforms. The recognition system is built using supervised learning and OpenCV, to predict the meaning of sign gestures in real-time.

Keywords: Deaf and mute, Video-conferencing, Supervised learning, OpenCV, Sign language Detection, Speech and hearing impairment.

I. Introduction

In this era, where inclusivity and sense of equality are considered a basic necessity for all. But this gets challenged when an individual who is fluent in sign and uses it to communicate with someone who isn't. We often take our ability to interact with others for granted. More than 70 million around the world uses sign language. But significant barriers to communicating in sign language are depriving these specially-abled people to enjoy even basic interaction with others. Although deaf and mute people can settle everything in writing, but there might be such situations where the corporation of manual sign language interpreters becomes a necessity. Hiring a translator can be exorbitant. The objective of this paper is to tear down this communication barrier a bit without having to learn sign explicitly or need of an interpreter.

With the recent rise of videoconferencing platforms, we identify the problem of signers not "getting the floor" when communicating, which either leads to them being ignored or to a cognitive load on other participants, always checking to see if someone starts signing.[1]

A Sign Language Detection System can be thought as a must-have feature for every video conferencing platform which comes into contact with people, who are hard-of-hearing or are not able to speak, on a regular basis. The proposed system revolves around the idea of a camera-based sign language recognition system that will provide the users with an interactive tool, capable of detecting hand gestures in real-time along with providing the percentage accuracy and further translating into text. This text is rendered as a header over the detected frame. The Model is trained over the input dataset using a neural network, provided for a particular sign, and further uses this training to detect the gestures in real-time.

II. Literature

The The detection of hand gestures with more precision is the key area for many researchers. Different techniques and algorithms, such as glove-based method, sensor-based approach, computing

approaches such as neural network, fuzzy logic have been used to provide an easy-to-use environment for the detection of sign language.

Authors of our second reference [2] implemented the sign language recognition system based on the image processing technique. This system recognizes the ASL sign alphabets using an edge detection algorithm. It also includes the removal of noise by smoothing algorithm. As a result, it will display the text meaning of character alphabet, but this system is limited only to the detection of one hand.

Author of our third reference in [3], used a MYO armband for data collecting and Neural Network for processing and detecting Chinese Sign Language (CSL). The latency was good, but one had to wear armband all the time.

Authors of our fourth reference proposed Sign Language Detection System based on Deep Multi-layered Convolution Neural Network. The system detects static and dynamic hand gestures with an accuracy of 99.89% but the system does not apply to real-time detection of sign gestures [4].

Author of our fifth reference [5] developed the system using an Eigenvector-based image processing algorithm to recognize several Indian sign language signs for live video sequences with 96.25% accuracy. This system eliminated the difficulty of glove-based approach, but it is limited only to recognizing ISL alphabets.

In glove-based method (reference six [6]), the sensors in the gloves can detect the movement of hands and pass the information to the computer. This approach has high accuracy in gesture recognition but is quite expensive and inconvenient to the user.

Same approach was used by our seventh reference [7]for recognizing ISL alphabets and numerals. For easy hand segmentation, the signer must wear a red and blue colored glove while collecting data. And it's all done with Principal Component Analysis (PCA). The recognition rate is 94 percent. However, the system only recognizes static motions and ignores signs in which both hands overlap.

Flex sensors also came into the picture, which were developed for mute audiences researched in our eighth reference [8], in which the user's hand is attached with the flex sensors. The Flex sensor responds individually to the bend of each finger. Taking this value, the controller starts to react with speech, each flex sensor holds unique voice stored in the APR Kit and for each sign it will play a unique voice. The work is done only for certain alphabets and not for the words or phrases, and the obtained accuracy is very low.

author in reference nine [9], used Kinect to capture the 3D video stream and the joints of interest in the human skeleton. It had 97% accuracy of words detection but it does not recognize differences at finger level and also gestures performed in forward or backward sequence.

Authors of our tenth reference [10] implemented a tool to recognize hand gestures for Indian Sign Language using the Convolution Neural Network (CNN) Deep Learning algorithm for mute people. This system gives the user a frontend that can be used for daily purposes. Therefore, this system can further improve the success rate by implementing sentences and phrases for better user experience.

Reference eleven proposed a system capable of detecting Bangla sign language for deaf and mute communities using YOLOv4 as an object detection model and generate both textual sentences and speech in real-time. It also proposed three new signs for the task of sentencing generation. In addition, this system will be more useful if it is implemented in several other sign languages [11].

But as far as video conferencing and video meetings are concerned there is not much that has been published so far. Recently, a former journalist said that Google is developing sign language detection for videoconferencing. an intern at Google Research, recently published a blog post describing some of the work that the Google has been doing to make videoconferencing technology more accessible to people communicating through sign language.

According to the literature survey, most of the above focused on improving accuracy and correcting gestures, but we are trying to add sign language recognition feature on video conferencing platforms, also Google is working on it. Video conferencing applications do not have a mechanism for detecting

sign language and we are therefore trying to do so. It can help these specially-abled people to communicate easily and effectively.

III. Proposed Methodology

Sign language recognition has been widely studied across different domains and sign languages. As sign language corpora are usually small[12],previous works take one of two approaches to reduce the network's parameters: (1) using pose estimation on the original videos[1], [9]; or (2) using pre-trained CNNs to get a feature vector per frame[3], [4], [10], [11].
We used the second approach, i.e., a pre-trained neural convolution network (CNN) to implement our demonstration application due to its high precision for image classification and feature extraction.
Steps involved:

a) Creating Dataset:

Instead of working over all the signs of ASL, we selected few signs for implementing purpose. Then collected 16 pictures of each selected sign through python OpenCV, which is huge open-source library for computer vision and machine learning as shown Fig.1 Collected images were taken considering different positions of and direction to increase the detection accuracy.

Figure 1: Dataset

b) Labeling Dataset

Image annotation is a necessary step in supervised learning to make our dataset a useful component for training our machine learning model. Labeling maps the dataset to its specific output that we expect as our detection result. We labeled our
images with LabelImg, an open-source annotation tool. It creates an xml file that contains the orientation information for each labeled part of the image, as shown in fig.2.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

Figure 2: Labelled image and its Xml file

c) Training Dataset

We trained our labeled dataset using Convolution Neural Network (CNN) for detecting hand gestures in real-time. Instead of building and training CNN from scratch we used a pre trained TensorFlow object detection Model. Through transfer learning, we trained the pre-trained TensorFlow object detection model Zoo to our smaller and specific purpose dataset for 5000 steps.

We used ssd_mobilenet_v2_fpnlite_320x320_coco17_tpu-8 architecture as our training model. SSD is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network. This makes SSD easy to train and straightforward to integrate into systems that require a detection component. It has a small convolutional filter to predict object categories and offsets in bounding box locations. It uses separate predictors (filters) for different aspect ratio detections, and applying these filters to multiple feature maps from the later stages of a network in order to perform detection at multiple scales[13].

d) Detection in real time

Real time video can be thought as sequence of multiple frames. For real time detection, the system will take hand gesture as input data. It would detect and track hands and fingers placed within its field of view and presents motion tracking data as a series of snapshots called frames. Each frame is resized to dimension of 800x600. Each frame of tracking data contains the measured positions, orientations and other information about each entity detected in that snapshot. Using this information, it will try to match the detected frame to it correct sign and display the label as output as shown in Fig.3.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

Figure 3: Workflow of Sign language Detection System

IV. Experimental Results

Each sign for which the model was trained is detected with high accuracy for both hands as shown in Fig 4 and summarized result can be seen in Table1.

Table 1: Experimental Results

Sr.No	Gestures	Accuracy Achieved
1.	Hello	98%
2.	Yes	99%
3.	No	99%
4.	Thank You	94%
5.	See You	84%
Average accuracy		94.8%

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, May : 2023

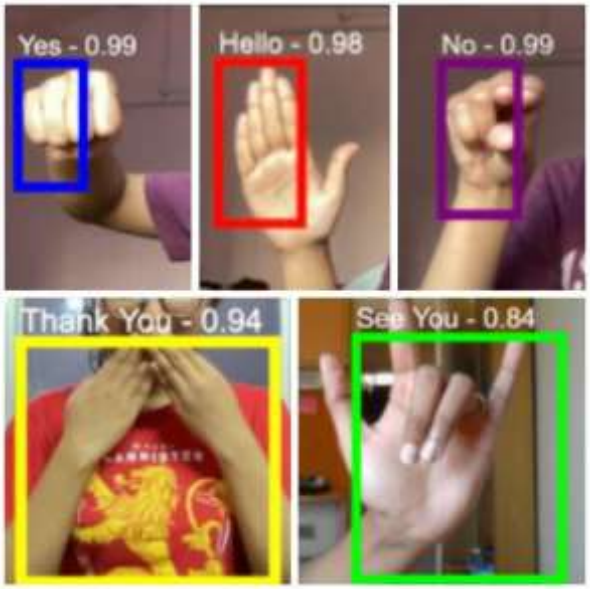

Figure 4: Detection Results

V. Conclusion

Technology cannot be labelled as "accessible" until and unless each and every individual regardless of their disability, are able to operate it easily. There are about a billion people in the world who cannot access applications in the same way as we do. Making an application accessible to all is not just the law, but it is extremely easy if we put our minds to it. Inclusion is all about understanding, respecting and ensuring that everyone's voice and opinions are heard and carefully considered. It is crucial to create an environment in which everyone feels respected. The main purpose of this work is to bridge the barriers of communication between the hearing impaired, mute and the general public. More audience can take the leverage of video conferencing platforms if sign language detection can be embedded as a feature. We trained our model with images of different direction and orientation; thus, our system is capable of detecting precise and exact gesture in every possible angle with average accuracy of 94.8 % in real time. Our system does not require any tool like gloves, sensors and any kind of equipment, the only tool required is web camera. The system is simple, swift, flexible and fast for every user with self-explanatory interface. With society's lack of fluency in sign language, sign language detection feature over video conferencing platforms will allow these specially-abled people to blend easily in the society.

VI. Future Scope

- This software can be scaled over application like Zoom, Microsoft teams etc, for the ease of every individual.
- In future AI assistants like Alexa, google duo can be deployed to respond to sign language using AI.
- Just like WhatsApp and Google where we can use feature of Voice Search (Speech-to-text translation), this project may be further developed to convert the sign Language to text using Natural Language Processing.

References

[1] A. Moryossef, I. Tsochantaridis, R. Aharoni, S. Ebling, and S. Narayanan, "Real-Time Sign Language Detection using Human Pose Estimation," Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.04637.

[2] S. Shrenika and M. Madhu Bala, "Sign Language Recognition Using Template Matching Technique," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Mar. 2020, pp. 1–5. doi: 10.1109/ICCSEA49143.2020.9132899.

[3] Z. Zhang, Z. Su, and G. Yang, "Real-Time Chinese Sign Language Recognition Based on Artificial Neural Networks *," in *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2019, pp. 1413–1417. doi: 10.1109/ROBIO49542.2019.8961641.

[4] R. Bhadra and S. Kar, "Sign Language Detection from Hand Gesture Images using Deep Multi-layered Convolution Neural Network," in *2021 IEEE Second International Conference on Control, Measurement and Instrumentation (CMI)*, Jan. 2021, pp. 196–200. doi: 10.1109/CMI50323.2021.9362897.

[5] J. Singha and K. Das, "Recognition of Indian Sign Language in Live Video," *International Journal of Computer Applications*, vol. 70, no. 19, pp. 17–22, May 2013, doi: 10.5120/12174-7306.

[6] R. Y. Wang and J. Popović, "Real-time hand-tracking with a color glove," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 1–8, Jul. 2009, doi: 10.1145/1531326.1531369.

[7] D. Deora and N. Bajaj, "Indian sign language recognition," in *2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking*, Dec. 2012, pp. 1–5. doi: 10.1109/ET2ECN.2012.6470093.

[8] Y. 'Raushan, A. 'Shirpurkar, V. 'Mudholkar, S. 'Walke, T. 'Makde, and P. 'Wahane, "SIGN LANGUAGE DETECTION FOR DEAF AND DUMB USING FLEX SENSORS," *International Research Journal of Engineering and Technology (IRJET)*, vol. 04, pp. 1–3, Mar. 2017.

[9] F. Nasir, U. Farooq, Z. Jamil, M. Sana, and K. Zafar, "Automated Sign Language to Speech Interpreter," in *2014 12th International Conference on Frontiers of Information Technology*, Dec. 2014, pp. 307–312. doi: 10.1109/FIT.2014.64.

[10] T. Aadit, J. Deepak, P. Janhavi, and D. Jyoti, "Video Chat Application for Mutes," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Mar. 2021, pp. 181–185. doi: 10.1109/ESCI50559.2021.9397044.

[11] D. Talukder and F. Jahara, "Real-Time Bangla Sign Language Detection with Sentence and Speech Generation," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, Dec. 2020, pp. 1–6. doi: 10.1109/ICCIT51783.2020.9392693.

[12] D. Bragg *et al.*, "Sign Language Recognition, Generation, and Translation," in *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, Oct. 2019, pp. 16–31. doi: 10.1145/3308561.3353774.

[13] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.