iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** IV **Month of publication:** April 2020

DOI: http://doi.org/10.22214/ijraset.2020.4048

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IV Apr 2020- Available at www.ijraset.com

Virtual Assistant for Deaf and Dumb

Niraj Singh Tomar[1], Sukanya Deshpande[2], Abhishek Jagnade[3], Dr. Srinivas Ambala[4], Avadhoot Bhosle[5]
[4]Associate. Prof, [1, 2, 3, 5]Computer Engineering, G.H. Raisoni College of Engineering & Management, Pune, Maharashtra

Abstract: Voice is the future of computing interfaces, such as Robotics or the Internet of things. It is very difficult for those who have hearing or speaking disabilities. Almost all products or applications which are being developed today have voice-controlled features in them. It creates an indifference among the normal people and those with hearing or speaking disabilities.
These applications should be able to be used by all sections of the potential users equally, so we propose a system that will bridge the gap between all the potential users whether normal or disabled (hearing or speaking impaired). Our systems take in input1 in the form of sign language video, it first breaks the gesture video into frames and then apply Convolutional Neural Network (CNN) on these gesture-frames so as to extract meaningful text, which is then spoken aloud by the device (on which the system is running) this is fed to Google Assistant/Amazon Alexa as input2, the response is converted to text and displayed on the screen and also spoken aloud by the device. This way people with hearing or speaking disabilities can communicate with a virtual assistant.
Keywords: Sign language, gesture-frame, CNN, Google Assistant/Amazon Echo

I. INTRODUCTION

In the world of the digital era, communication is everything whether in the form of text or voice or gesture. We come across many people who have hearing disabilities or who cannot speak. It is very difficult for them to convey their message or feeling to normal people (who don't know sign language). In addition to this, they also feel left out from using the latest technologies. So we came through this problem and got motivated to make this project.

Currently, we all are surrounded by high-end technology which includes Visual recognition, object classification, video classification, however, we are dealing with processing of the gestures which are further broken into frames, having various texture, colors making it difficult to process the gesture frame image. As of now if we provide a machine with a gesture-frame photograph it puts out the meaning of gesture item however it should process the entire gesture and extract the real meaning of the gesture and produce it to the user. Hence it challenges the current technology to expand its knowledge far beyond its vision and should inculcate the high-quality structured information for the meaning extraction process.

Traditionally we are dealing with the formulated structure where we extract the meaning of the gesture-frame retrieved from the dataset. This system might crash down even if there are two similar gesture items but being shown in a different context. Therefore, it requires a dataset in huge quantity and diversity which increases its performance to its end.

II. RELATED WORK

A. Challenges in sign recognition

For every word used for conversation, there is a sign or gesture for it. As there are plenty of sentences, there are plenty of signs. American Sign language comprises of 26 signs known as American manual alphabets, these can be used to make words in the form of gestures also there are separate signs for various other words say - how, where, when, what etc. So, there can be plenty of such combinations of gestures.

Sign recognition involves complex methods such as motion modeling, analyzing the motion, recognition of pattern in the motion and machine learning.

The nature of environment such as the background light, speed of gesture affects the prediction. The difference in the viewpoints makes the sign to appear different in the 2-D space.

B. Existing System

There are various devices and systems available for translating the sign language into text, these devices include

1) *Hand Wearables:* The hand wearable holder is made of filaments which have elasticity and flexibility. These flexible filaments enable the hinges and joints to fit in any hand of any size. There are sensors in these wearables which record the motion of sign and the model predicts the meaning. These wearables are expensive and complex in manufacturing.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IV Apr 2020- Available at www.ijraset.com

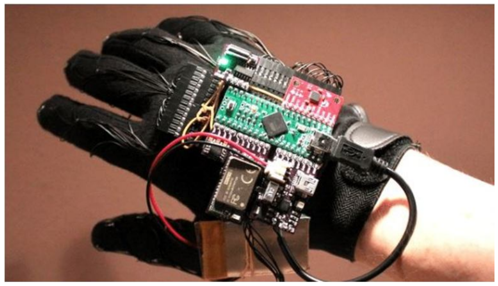


Figure 1 Hand-glove wearable for gesture prediction

2) *DFRobot 3D Gesture Sensor:* This integrates gesture analyzation and motion tracking, it can be used to detect both clockwise and anti-clockwise signs. Then the microprocessor chip analyses the meaning.

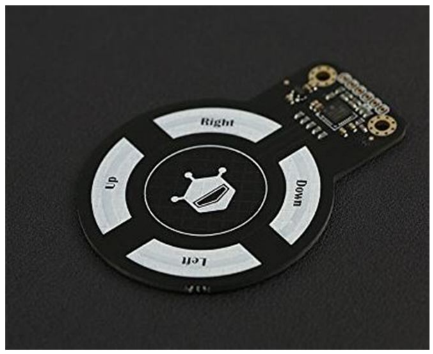


Figure 2 DFRobot 3D gesture sensor

C. Training the Model

The model needs to be trained for each gesture separately.
This is done by enacting each sign in front of the camera and recording a video of the sign.

Figure 3 Recording training video

This gesture-video is then converted into series of gesture-frames and given a label or the meaning of the sign.

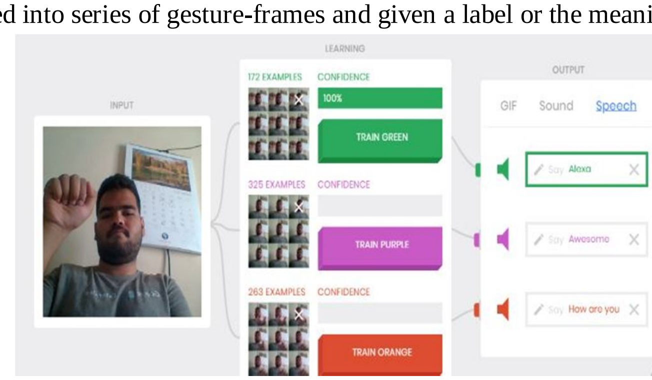


Figure 4 Training the model

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IV Apr 2020- Available at www.ijraset.com

D. Challenges in Training

There are various challenges in training the model such as:

1) The background of the trainer should be bright. So that, the model can identify the gestures properly both while training and testing.
2) The video of each sign should involve multiple people of various complexion and body type so as to make the prediction accurate/.

III. VIRTUAL ASSISTANT

Generating a textual meaning of the sign from a series of gesture-images which requires understanding of the sub-signs composing the sign is a peculiar job.

The meaning is extracted from the series of gesture-frames using convolutional neural network (CNN).

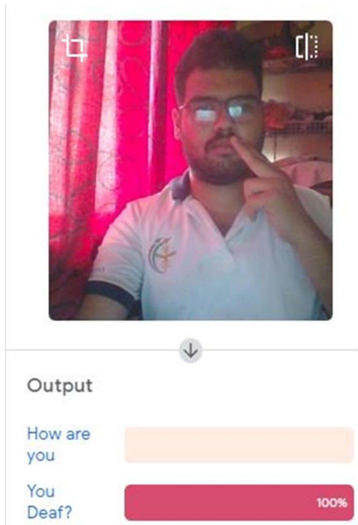


Figure 5 Meaning out of gesture

This meaning is then fed into a voice assistant such as Google Home or Amazon Echo.

Figure 6 Response by Amazon Echo

The response produced by the voice assistant is the final output.

IV. CONCLUSION

We introduced a sign-language and voice assistant(Google Assistant/ Amazon Alexa) based communication system, which takes a gesture video as input and produces a meaning out of it, which is then fed into the Google Home/ Amazon Echo. The response from the Echo is produced to the user on the browser in the form of text and sound. We first predicted sets of gesture-frame labels from gesture-frame images, showing that modeling dependencies matters. Then, we produced meaning out of it. Which is then fed into Echo for the response.

Finally, the user demonstrates the superiority of our system against state-of-the-art sign-language and voice assistant (Google Assistant/ Amazon Alexa) base communication approaches.

REFERENCES

[1] Dimitris Kouremenos, Stavroula-Evita Fotinea, Eleni Efthimiou,Klimis Ntalianis - A prototype Greek text to Greek Sign Language conversion system (Behaviour and Information Technology Volume 29, 2010 - Issue 5)

[2] Lavanya S, Kalyan R V N S, Vadivu G - A GPU based deep Convolutional Neural Network System to detect and convert Indian Sign Language for mobile devices (International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 4923-4930).

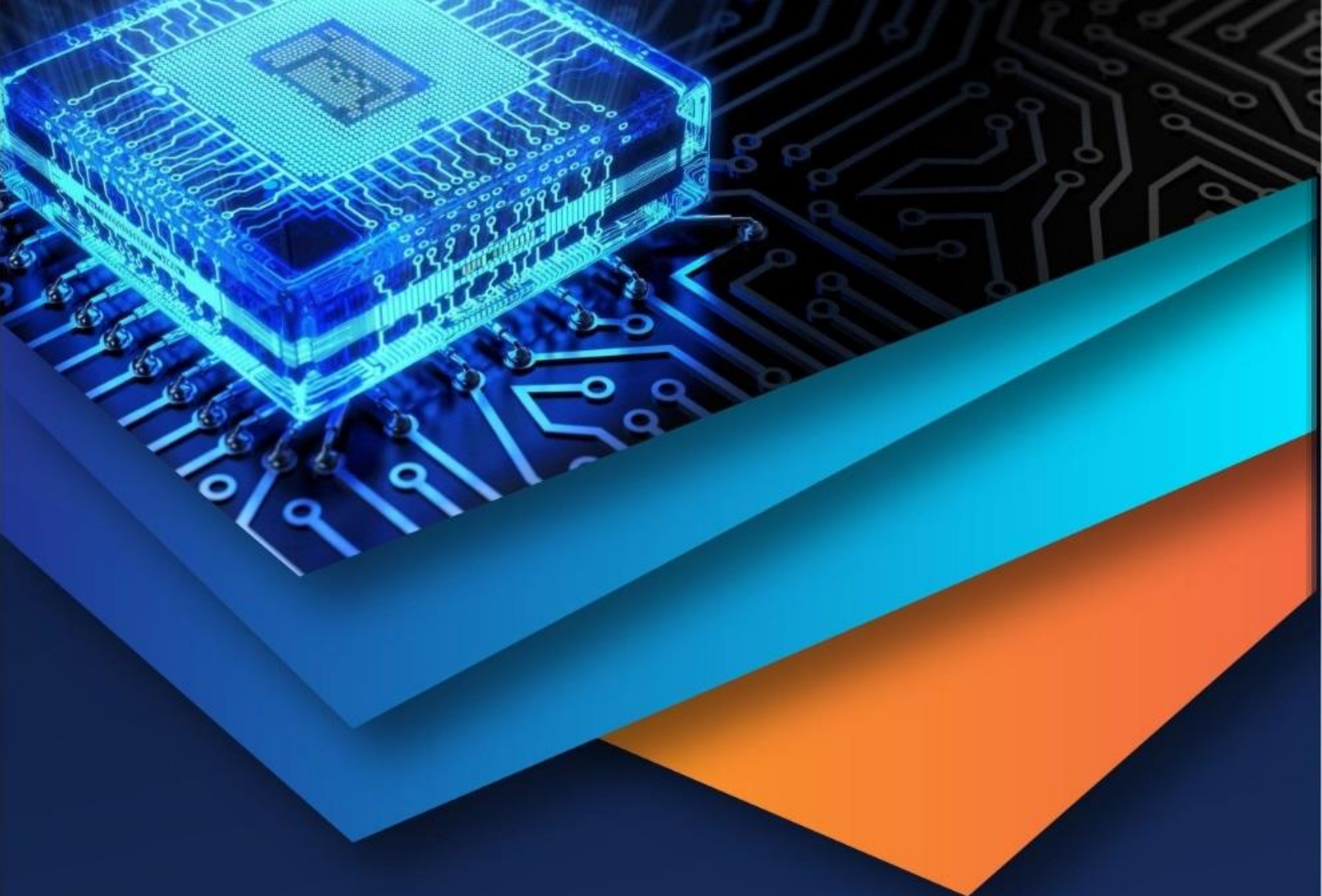
[3] P Vijayalakshmi, M Aarthi - Sign language to speech conversion (2016 International Conference on Recent Trends in Information Technology (ICRTIT)).

[4] Jie Huang, Wengang Zhou, Houqiang Li, Weiping Li - Sign Language Recognition using 3D convolutional neural networks (2015 IEEE International Conference on Multimedia and Expo (ICME)).

[5] Taner Arsan and O guz Ulgen - Sign Language Converter (International Journal of Computer Science and Engineering Survey (IJC-SES) Vol.6, No.4, August 2015).

[6] Qutaishat Munib, Moussa Habeeb, Bayan Takruri, Hiba Abed,Al-Malik - American sign language (ASL) recognition based on Hough transform and neural networks (Expert Systems with Applications Volume 32, Issue 1, January 2007, Pages 24-37)

.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY