Sign Language Detector And Translator

A PROJECT REPORT
Submitted by

SANJU RAJ PRASAD
SWAPNIL RATHOD
JAIDEEP RATHWA

In Partial fulfillment for the award of

the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Under the Guidance of
Prof. Yatin Shukla

Department of Computer Science & Engineering

Parul University, Vadodara

April, 2022

PARUL UNIVERSITY

CERTIFICATE

This is to Certify that Project-III -Subject code 203105450 of 8th Semester
entitled "**Sign Language Detector And Translator**" of Group No. 4

PUCSE_13 has been successfully completed by

SANJU RAJ PRASAD - 18030105259
SWAPNIL RATHOD - 180303105270
JAIDEEP RATHWA - 180303105271

under my guidance in partial fulfillment of the Bachelor of Technology (B.TECH) in
Computer Science and Engineering of Parul University in Academic Year 2021-

2022.

Project Guide   Project Coordinator
Prof. Yatin Shukla   Prof. Khushali Mistry
                                                       Prof. Yatin Shukla

Head of Department,   External Examiner
   CSE
Dr. Amit Barve

SIGN/GESTURE DETECTOR & TRANSLATOR

Acknowledgement

ABSTRACT

Object detection is a task in computer vision that involves identifying and localising objects in order to detect or classify them. Human–computer interaction, video surveillance, satellite images, transportation systems, and activity identification are some of the most common object detection applications.For visual imagery, a convolutional neural network (CNN) made up of a series of neural network layers is utilised in the larger family of deep learning architectures. Deep CNN architectures produce excellent results when it comes to object detection in digital images.

Sign language is a vital part of communication between human to human, human to machine and many more. Understanding sign language can be a tough task and moreover translating them to others so that they can understand what one is saying can be painful. According to world health organization aka WHO by 2050 nearly 2 billion people are projected to have some degree of hearing loss will require hearing rehabilitation. Our aim is to provide a better way of communication to these group of people by providing them a software which can talk for them i.e which can translate their thoughts and emotions and transfer it to others. Gesture detection, text to speech, neural networks image processing, perceptual analysis are some methods used to build this project and by studying the results obtained by them gave us the information about how people with deficiency think differently as their brain reacts different.

INTRODUCTION

The sign language is made by specifications of hand and facial idioms to express their views and thoughts of speech and hearing disabled persons with the normal people. Sign language is a way communication and not everybody is aware of it. Sign language and gestures are used at many places where verbal communication cannot takes place, for example during surgery doctor uses hand gestures, in military soldier also uses gesture for communicating, the people who having deaf disability uses these languages to communicate with others.

Objective of this project is to provide a better way of communication where sign language play a vital role, for example people with deaf disability faces many problem while having job interviews, online examinations, remote business deals etc.

Algorithms are a reliable way to handle computer vision problems. Object detection in computer vision is the process of finding instances of objects from a specific class in a digital picture or video. It is a task which involves classifying and locating items in order to detect them. It scales one or more items and determines where the object is presented in the image.

Face recognition system, emotion detection systems, video surveillance, vehicle tracking, and autonomous vehicle driving are all becoming increasingly prominent, therefore fast and precise object detection systems are in high demand. The term "object detection" refers to the process of locating and classifying items within a digital image. CNN-based object detectors are used in a number of applications as a result of the progressive results from deep CNN architectures. It has been classified as a single-stage or two-stage object detection model based on the methodology. R-CNN, Fast R-CNN, Masked R-CNN, SSD, and YOLO are among the CNN-based models covered in this work. Apart from that, it explains the various characteristics of the available datasets. It also goes into the specifics of previous studies that used object detection models in many fields of application.

LITERATURE REVIEW

2.1 Perceptual Encoding of Fingerspelled

The main objective of this journal is to measure brain activation during perception of fingerspelled letters, printed letters and abstract shapes in six congenitally, profoundly deaf signer and six normal people.

Normal people showed extrastraite cortex activation when exposed to fingerspelled letters and printed letters, on the other hand the people with deaf disability tend to activate the broader network in fingerspelled and printed letters.

The results suggested that on one hand different cerebral areas gets activated in case of people with deaf disability and on the other hand common cerebral areas gets activated when tested with normal people.

2.2 Convolution of Neural Networks for Image

Convolution network is used to describe structure applied to neural network to two dimensional array usually images.A convolution neural network image mapping is characterized by strong constraint of requiring that each neural connection implements the same transformation this increases the chances of generalization.
CNN takes raw data without the need for an initial separate pre-processing or feature extraction stage: In CNN feature extraction and pre-processing stages occur naturally in the framework.
CNN performs mapping temporally distributed arrays in arbitrary dimensions. They appear suitable for application to time series, images or videos.

CNN characteristic :-
Translation invariance.
Local connectivity.
Optional progressive decrease in spatial resolution.

2.3 Hand Gesture Recognition

The main purpose of building/ making a hand gesture is to symbolize something and is a way to communicate to others.Hand gestures can be used for telling a computer to do something, it is also used by commandos and other forces while doing drills and critical operations. Basically it takes the place where verbal communication cannot happen. These hand gesture also used by those people who have deaf disability.

2.4 Text To Speech Synthesis

Speech synthesis is field of computer science which deals with designing computer system that synthesis written text to audio format.
The basic idea of text to speech technology is to convert written inputs to spoken output by generating synthetic speech. There are several ways of doing it:-

1.Simple voice recording and playing on demand.
2.Splitting of speech into 30-50 units and reassembling them in a fluent speech.
3.The use of approx 400 diaphones (splitting of phrases at the center of units and not the transitions).
Continues speech is a set of complicated audio signals which makes producing them artificially difficult. Speech signal are usually considered as voiced and unvoiced but in some cases they are something in between.

2.5 Automated Language Translation

Languages embody many indigenous values and concepts and indigenous people histories and development. India as multi-lingual country faced with challenges such as local language revitalization, sharing knowledge and information and the main reason for that is India have hundreds of local languages.
The improvement in computer aided technology binded with cheap computer devices such as smart phones have reduced the cost of translation very significantly.
Technologies that offer speech to speech or text to text really filled the gapes of having multiple local languages.

PROJECT FLOW AND METHODOLOGY

The project flow to acquire information and the to process it is as follow :-
1. Gathering the image from digital device mainly smart phone cameras.
2. Using that image for processing the information gained by it.
3. The image then transferred to trained neural networks modal to extract the meaning of the symbol.
4. The output of the modal will be a text.
5. The text then transferred to text to speech converter and then the output will be   in voice format.

A basic diagram to show the work flow:-
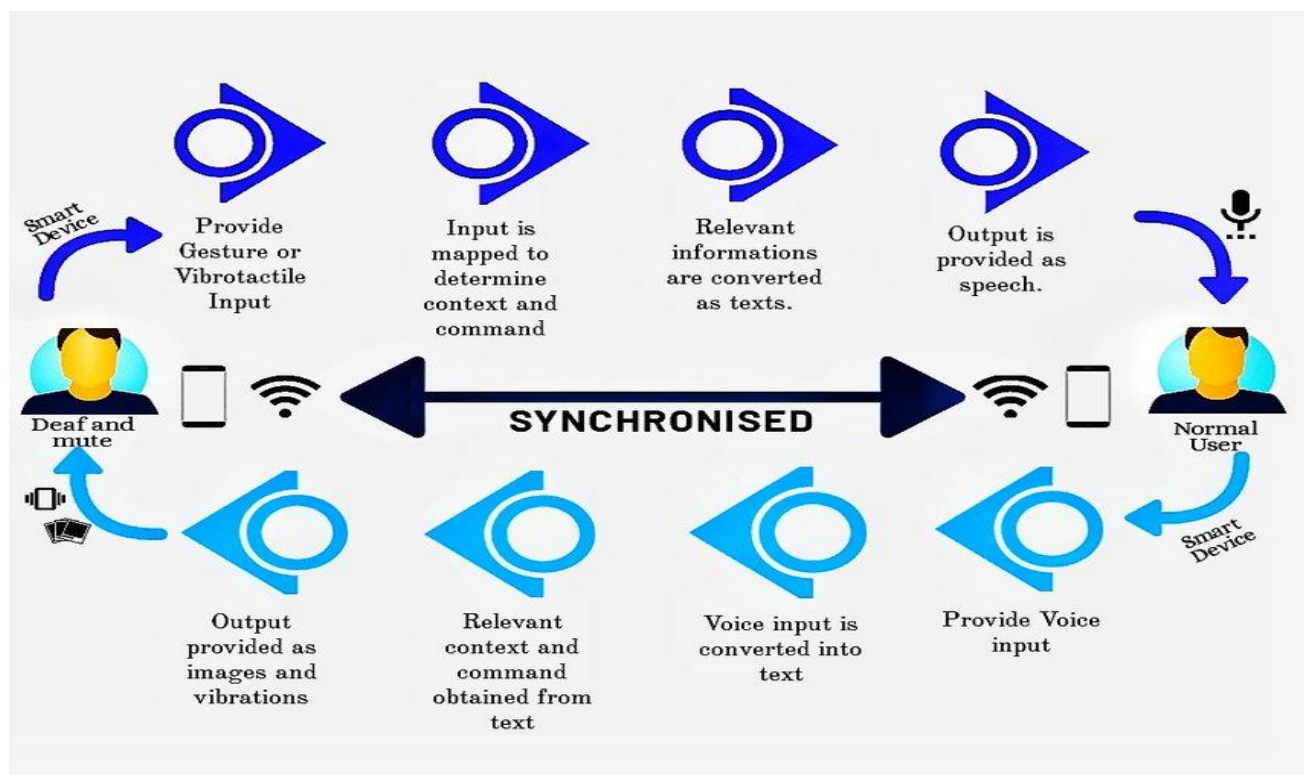
Perceptual Encoding Experiment

This research uses experimental methodology to figure out how a person's brain having deaf disability reacts to certain perceptual things like symbols, patterns versus how a normal person's brain reacts to theses symbols and patterns.

The way brain reacts when one more senses are not working properly makes this methodology relevant for sign and gesture detector software which eventually produces a translated sign language to casual human language which can be understand easily by normal people.

Convolution Neural Network for image processing

The CNN architecture involves a total of five layers i.e a single input and output map and three hidden layers. The filter size used in all cases are 5x5 . A representative data set of 30 - 50 frames from online recording session were manually classified for training and validation of the neural networks. A training data set will be generated using 20 images, sampling 100 crack and 200 non-crack pixel location, yielding a total of 6000 inputs.

Although all the pixel location had associated targets, not all pixels are used in training because of computational expense, low proportion of crack to clean.

References

1.**Wikipedia Object detection** (https://en.wikipedia.org/wiki/Object_detection)
2.**Wikipedia Region based detections**
(https://en.wikipedia.org/wiki/Region_Based_Convolutional_Neural_Networks)
3.**David Cocard MobileNetSSD** (https://medium.com/axinc-ai/mobilenetssd-a-machine-learning-model-for-fast-object-detection-37352ce6da7d).
4.**Wikipedia**, 2014. sign language [Online]
5.Available at: https://en.wikipedia.org/wiki/Sign_language
Renaud Brochard, 2004. Perceptual encoding of finger spelled
Available at:
https://www.academia.edu/45541208/PERCEPTUAL_ENCODING_OF_FINGERSPELLED_A _PRINTED_ALPHABET_BY_DEAF_SIGNERS_AN_f_MRI_STUDY_1_2

6.**Browne Matthew**, 2003. Convolution Neural Networks. 16ᵗʰ Australian Conference on AI,Perth, Australia