Virtual Assistant implementation with Sign Language

A seminar report
submitted in partial fulfillment of
the requirements for the award of the degree of
BACHELOR OF TECHNOLOGY

in

Computer Science & Engineering

from

APJ Abdul Kalam Kerala Technological University

Submitted By

IJAS SUBAIR K (MEA17CS022)

MEA Engineering College
Department of Computer Science and Engineering
Vengoor P.O, Perinthalmanna, Malappuram, Kerala-679325
NOVEMBER 2020

Department of Computer Science and Engineering

MEA ENGINEERING COLLEGE

PERINTHALMANNA-679325

Certificate

*This is to certify that the seminar report entitled "***Virtual Assistant Implementation With Sign Language***" is a bonafide record of the work done by* **IJAS SUBAIR K(MEA17CS022)** *under our supervision and guidance. The report has been submitted in partial fulfillment of the requirement for award of the Degree of* **Bachelor of Technology** *in* **Computer Science & Engineering** *from the APJ Abdul Kalam Kerala Technological University for the year 2020.*

Mr Muhammad Shameem P
Seminar Guide
Assistant Professor
Dept.of Computer Science & Engineering
MEA Engineering College Perinthalmanna

Dr. Raji C.G.
Head of The Department
Dept.of Computer Science and Engineering
MEA Engineering College Perinthalmanna

Acknowledgements

An endeavor over a long period may be successful only with advice and guidance of many well wishers. I take this opportunity to express my gratitude to all who encouraged me to complete this seminar. I would like to express my deep sense of gratitude to my respected **Principal Prof. Dr. G. Ramesh** for his inspiration and for creating an atmosphere in the college to do the seminar.

I would like to thank **Dr Raji C.G, Head of the department, Computer Science and Engineering** for providing permission and facilities to conduct the seminar in a systematic way. I am highly indebted to **Mr Muhammad Shameem P**, Asst. Professor in Computer Science and Engineering for guiding me and giving timely advises, suggestions and wholehearted moral support in the successful completion of this seminar.

My sincere thanks to seminar coordinator **Mr Ismail P. K.**, Asst. Professor in Computer Science and Engineering for his wholehearted moral support in completion of this seminar.

Last but not least, I would like to thank all the teaching and non-teaching staff and my friends who have helped me in every possible way in the completion of my seminar.

IJAS SUBAIR K(MEA17CS022) DATE:28/11/2020

Abstract

This technology is all about the interface developed, that allows deaf mutes to make use of various voice automated virtual assistants with help of Sign Language. Majority of Virtual Assistants work on basis of audio inputs and produces audio outputs which in turn makes it impossible to be used by people with hearing and speaking disabilities. The project makes various voice controlled virtual assistants respond to hand gestures and also produces results in form of text outputs. It makes use of concepts like Deep Learning, Convolutional Neural Network, Tensor Flow, Python Audio Modules.

A webcam first captures the hand gestures, then Convolutional Neural Network interprets the images produced and produces rational languages. These languages are then mapped to pre-defined datasets using Deep learning. For this purpose, Neural Networks are linked with Tensor flow library. The designed system will then produce audio input for the Digital Assistant, using one of the Python text to speech module. The final audio output of the Digital Assistant will be converted into text format using one of the Python speech to text module which will be displayed on the viewing screen.

Contents

Acknowledgements ii

Abstract iii

Contents iii

List of Figures v

1 INTRODUCTION 1
- 1.1 Sign Language Communication . . . 2
- 1.2 Overview Of Implementation . . . 3

2 LITERATURE REVIEW 4
- 2.1 Challenges Faced . . . 5

3 METHODOLOGY 6
- 3.1 Training Dataset . . . 8
- 3.2 Tensor Flow . . . 9
- 3.3 Machine Learning . . . 11
  - 3.3.1 Deep learning . . . 11
- 3.4 Convolutional Neural Network . . . 13
- 3.5 Computer Vision . . . 14
  - 3.5.1 OpenCV . . . 14
- 3.6 Python Text and Speech APIs . . . 16
  - 3.6.1 Text-to-Speech . . . 16
  - 3.6.2 Speech-to-Text . . . 17

4 RESULT 19

5 CONCLUSION 22

6 FUTURE SCOPE 23

7 REFERENCES 24

List of Figures

1.1 sign languages 2

3.1 Data Flow 7
3.2 Training Accuracy 9
3.3 Tensor flow 10
3.4 Deep Learning 12
3.5 OpenCV 15
3.6 Text to speech 17
3.7 Speech to Text 18

4.1 Hand Gesture for term Weather 19
4.2 Sequence frame 20
4.3 Interface 20
4.4 Classifier's Accuracy 21

CHAPTER 1

INTRODUCTION

Virtual Assistant devices have been an integral part of our lives nowadays, but most of them are Voice Automated. The most commonly used virtual assistants are Alexa, GoogleHome, Apple Siri, and Microsoft Cortana. These assistants listen to users' questions and respond accordingly making life easier, so they have been a very important part of home automation. Since these assistants are purely VoiceAutomated, the Deaf and Mute find it difficult to make use of this technology as noted in [8]. The agenda of the project is to develop an interface that will help the deaf and dumb to easily use these virtual assistants with ease. As of now, it might seem irrelevant to design such a system, but in the long run it could help deaf people enjoy their social and personal lives equally. Designing such an interface will allow them to find their freedom while using such technologies and could increase their confidence in the digital age.This article focuses on research that gives the idea of combining two modern technologies which are hand gesture recognition and virtual voice assistants in order to allow hearing / speaking people to interact with digital gadgets and communicate. even with the outside world. This research work implemented Alexa, an audio-based virtual assistant. The proposed system succeeded in replacing the speech recognition technique with the hand gesture recognition technique.

1.1 Sign Language Communication

Sign language(also known as signed languages) are languages that use the visualmanual mode to convey meaning. Sign languages are expressed through manual articulations in combination with non-manual elements. Sign languages are full-fledged natural languages with their own grammar and lexicon. Sign languages are not universal and are not mutually intelligible to each other, although there are also similarities between sign languages Linguists view spoken and signed communication as types of natural language, which means that both have emerged through an abstract and prolonged aging process and have evolved over time without meticulous planning. Sign language is not to be confused with body language, a type of non-verbal communication. Wherever communities of deaf people exist, sign languages have developed as a useful means of communication and form the core of local Deaf cultures. Although the signature is mainly used by the deaf and hard of hearing, it is also used by people with hearing, such as those who are unable to speak physically, those who have problems with spoken language due to a disability or condition (augmentative communication and alternative), or those with deaf family members, such as children of deaf adults. It is not clear how many sign languages currently exist in the world. Each country generally has its own native sign language, and some have more than one. The 2013 edition of Ethnologue lists 137 sign languages, some of which have obtained some form of legal recognition.

FIGURE 1.1: sign languages

1.2 Overview Of Implementation

The proposed system makes use of the following technologies: TensorFlow which is the most important library used to design and develop the model of these systems, Convolutional NeuralNetwork, is a Deep Learning algorithm which has been used to serve the purpose of Image Recognition, which helps to convert the images into matrix form which can be understood by the model and which makes it ready for Classifier, and finally OpenCV which will act as an eye of the system which will acquire and process hand gestures in real time and predictive results with the help by Classifier.

With increasing trends in technology, personal assistant devices are becoming more and more popular. But such devices are voice automated. They need audio inputs and provide audio outputs. So what if someone does not have their own voice or are not in a condition to speak properly, that's where this project comes into light. Such people can easily communicate with these devices using an interface that takes hand gestures as an input and provides audio as well as text output. This project has the capacity to bridge the gap between such impaired people and booming technology.

CHAPTER 2

LITERATURE REVIEW

Every existing virtual assistant today is voice automated, making it unusable for the deaf and people with certain disabilities. This leads to the need for a system that can help people with speech or hearing impairments to use such virtual personal assistants[8]. Artificial neural network is used in most of the cases where the static detection is performed[1]. However, there are few drawbacks with regard to the efficiency of recognizing distinguishing features from images, which can be improved by using the convolutional neural network. Compared to its predecessors, convolutional neural network recognizes important distinguishing features more efficiently and without human supervision. The ArtificialNeural Network uses a one-to-one mapping which increases the number of nodes required, reducing efficiency, while the Convolutional Neural Network uses too many, which keeps the number of nodes low and greatly improves efficiency[5].

Many systems designed with such goals typically use more physical hardware, such as the design observed in Cyber Glove, which requires the manufacture of such hardware gadgets and makes it mandatory for users to wear them, while accessing the virtual assistants [11]. Many systems are designed so that their application is limited to only certain sign languages or series of hand gestures [9], while the proposed system is designed so that we have the flexibility to switch to a standard sign language simply by changing the data set and the model to train for it.

2.1 Challenges Faced

Challenges may arise when researchers work with languages less studied in multilingual regions. It is often the case that at least one of these less-studied languages is acquired by children and used in contexts where they are a minority language, for example, in New York City, 50 percent of children use a language other than English at home, including That is Haitian Creole, Yiddish, African languages, Tagalog, Urdu, and Gujarati, and many other studies, while many studies have focused on the acquisition and use of two spoken languages, individuals who acquire more than one sign language and those who acquire both spoken and signature languages are also multilingual. Transcription and analysis of sign languages present specific challenges: they do not benefit from standard orthography, nor do they currently have a standard notation system. 10 The simultaneous use of different channels to produce speech - hands and face - complicates the accurate presentation of the different components of speech and may have specific effects on a particular method in the context of interactions.

Sign language, research shows that “deaf children in such a multilingual situation often produce classes in which both manual and audio channels are used simultaneously. The expressions expressed through each distinct channel may be separate or can be combined, in this case The two are transcribed independently of each other may not provide an imprecise meaning for the full presentation Ultimately, the data must be shared between researchers working in both spoken and sign languages.

CHAPTER 3

METHODOLOGY

The most basic explanation of the system workflow is as follows: a hand gesture is performed in front of the webcam as seen in [10]. This signal gesture is converted to text and the text output is converted to audio and served as input to the assistant. The wizard processes the question and answers in audio format. This audio format is converted to text output. The text output will be shown on the display screen.

Now to understand the technical workflow of the complete system, a short explanation is as follows: the first step is provide a training dataset and train the system with a variety of hand gestures named with their respective labels. This is the most time consuming. He passed. The better the underlying system software specifications, the less time required for training. Once the training is complete, the next phase is the prediction mode. Now it uses the input image from a webcam and runs it through the classifier to find its closest neighbors based on the training examples and labels provided in the previous step. If a certain prediction threshold is crossed, the label will be added in the frame, suggesting that the system recognized the next hand gesture accordingly. Then the Python Text-to-Speech module is used for speech synthesis to speak the detected tag to the digital assistant, after collecting a series of hand gestures and executing it in the presence of certain call commands that will be declared previously. If the spoken word is 'Alexa', it causes the nearby Echo to wake up and start listening to the message and then Alexa responds to the query in voice format. Meanwhile the system will start the python Voice to Text module which will listen to the answered query from Alexa and then convert it to text format and display it in the output frame on the display

screen. The entire process can be repeated several times. But hand gestures will only be recognized if they meet the standards of the trained dataset.

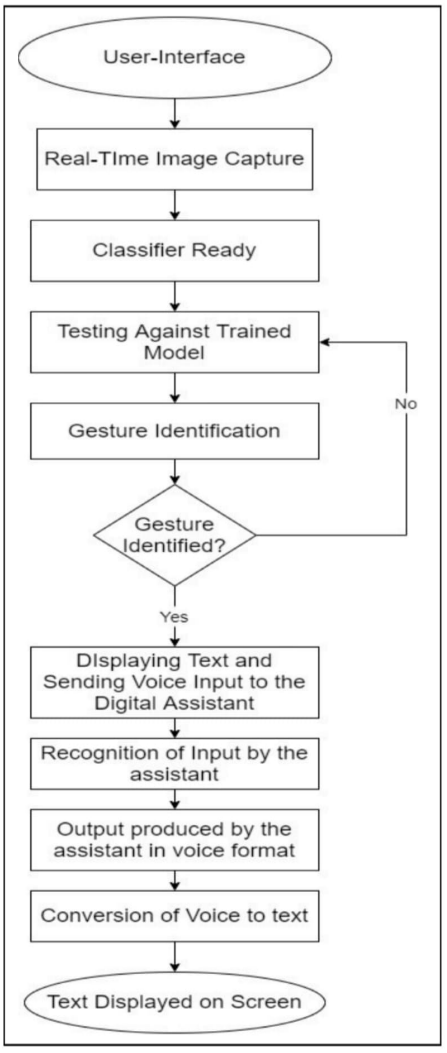

FIGURE 3.1: Data Flow

3.1 Training Dataset

Dataset is the most fundamental element of any Machine Learning Model. As it is a process of feeding into Machine's memory to help classify whatever it insights in future for the designed application. Since our system is an interface for Real-Time Classification of Hand Gestures our Dataset will purely consist of large number of Images in form of .jpeg, .jpg, these are the only two extensions that out model is accepting. The designed model makes use of a Labelled dataset method for training our system, thus assigning labels to folder names will simply use sub-files of images to be trained under assigned labels. Each label is being trained with about more than 2000 images captured at various possible angles in order to make system learn better and classify more accurately and quicker as observed in [2]. Once the model is completely trained for a set of particular labelled images it gets Classifier ready and can be used for testing the system's prediction rate. However it was noticed that retraining the same set of labels tends to give better results in terms of accuracy and speed of predicting the Hand gestures as observed in [3].

Basically the model will be trained more number of times for the same set of labels, higher is the success rate. But it is required to keep a note that any changes made with the labels folder before training will lead to the system that is being trained for the very first time. In simple terms if it is required to make any changes in the labels folder that is adding new labels or replacing the existing labels, the model will need to be trained again from the beginning. It was observed that for training of around 15 labels on an average configured system, it takes about 12-15 hours straight of model training for the first time. However retraining of same set of labels requires comparatively lesser amount of time.

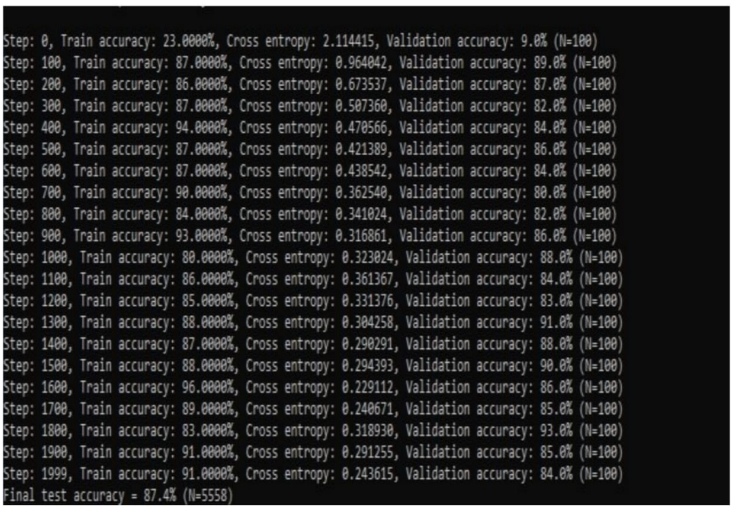

FIGURE 3.2: Training Accuracy

Fig. 3.2, is the part of training a dataset that shows the training accuracy obtained at respective number of step, along with the Cross Entropy value and Validation Accuracy for the same. Basically train accuracy is the value which let us understand how well the training is taking place while the Validation Accuracy let us know how the model will react while predicting the data it has not seen before.

3.2 Tensor Flow

The best part about using the TensorFlow library is that it is an open source library with many pre-designed models useful in machine learning and especially deep learning. of Tensor Flow is needed to understand the meaning of two terms, where theTensor here is considered an N-dimensional array and Flow refers to the operations graph. Every math calculation in TensorFlow is considered a graph of operations where the nodes in the graph are operations and the edges are just tensors.

Any math calculation is written as a data flow diagram in Python Frontend or C or Java, as in our case, Python is used. Then, TensorFlow Execution Engin comes into picture and makes it deployable on any hardware of the embedded system, be it CPU or Android or iOS. TensorFlow is a machine learning framework that includes using the dataset to train deep learning models and helps predict and improvise future outcomes.

The biggest advantage of using TensorFlow is it's feature of providing Abstraction, that is the developer does not need to work on every small aspects of designing the model as it is managed by the library itself, thus giving the developer freedom to focus on logic building.

TensorFlow in our system helps us in training the model using the provided dataset. TensorFlow object recognition algorithms helps us classify and identify different hand gestures when combined with use of OpenCV. By analysing thousands of photos, Tensorflow can help classifying and identifying real-time hand gestures. It makes possible to develop a model which can help identify 3D images and classify it on basis of 2D images from its feed dataset. TensorFlow is capable of processing more information and spot more patterns.

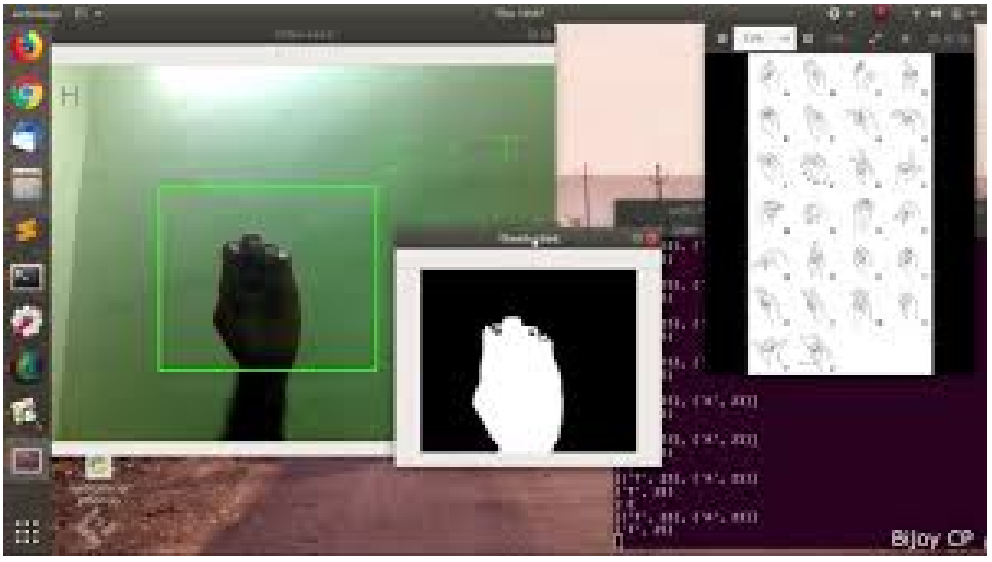

FIGURE 3.3: Tensor flow

3.3 Machine Learning

Machine learning is an application of artificial intelligence (AI) that enables systems to automatically learn from experience and improve without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it for themselves. The learning process begins with observations or data such as examples, direct experience or instructions to look for patterns in data and make better decisions in the future using the examples we provide. The main goal is to allow computers to learn automatically without human intervention or assistance and to adjust actions accordingly.

Deep Learning is basically a subset of Machine Learning model which consists of algorithms that make use of multilayer neural networks. Deep Learning makes use of Neural Network most of the times to implement its functioning. A Neural Network is a collection of layers that transforms the input in some way to produce output.

3.3.1 Deep learning

The field of artificial intelligence is essentially when machines can perform tasks that normally require human intelligence. It encompasses machine learning, where machines can learn by experience and acquire skills without human involvement. Deep learning is a subset of machine learning in which artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data. Similar to how we learn from experience, the deep learning algorithm would perform a task repeatedly, each time adjusting it a bit to improve the outcome. We refer to "deep learning" because neural networks have several (deep) layers that enable learning. Virtually any problem that requires "thinking" to solve is a problem that deep learning can learn to solve. The amount of data we generate every day is staggering - currently estimated at 2.6 quintillion bytes - and it is the resource that makes deep learning possible. Since deep learning algorithms require a ton of data to learn from, this increase in data creation is one of the reasons deep learning capabilities have grown in recent years. In addition to creating more data, deep learning algorithms benefit from the stronger computing power available today, as well as the proliferation of artificial intelligence (AI) as a service. Artificial intelligence as a service has given smaller organizations access to artificial

intelligence technology and specifically the artificial intelligence algorithms required for deep learning without a large upfront investment. -connected. The more deep learning algorithms learn, the better they work.

Image can be termed as matrix of pixel values so it may seem that classification can be an easier task simply based on matrix classification but that is not the case with complex matrix images or images with similar forms of matrix or a very huge dataset of images with minimal changes in the matrix. This may lead to clash in prediction scores and thereby affecting the accuracy and speed of classifier model. This is where Neural Network comes into picture and thus it is required to use deep learning over machine learning. Machine Learning works with lesser number of layers when compared with Deep Learning as observed from [12] and thus not preferred for technologies like Image Recognition which requires need of Convolutional Neural Networks.

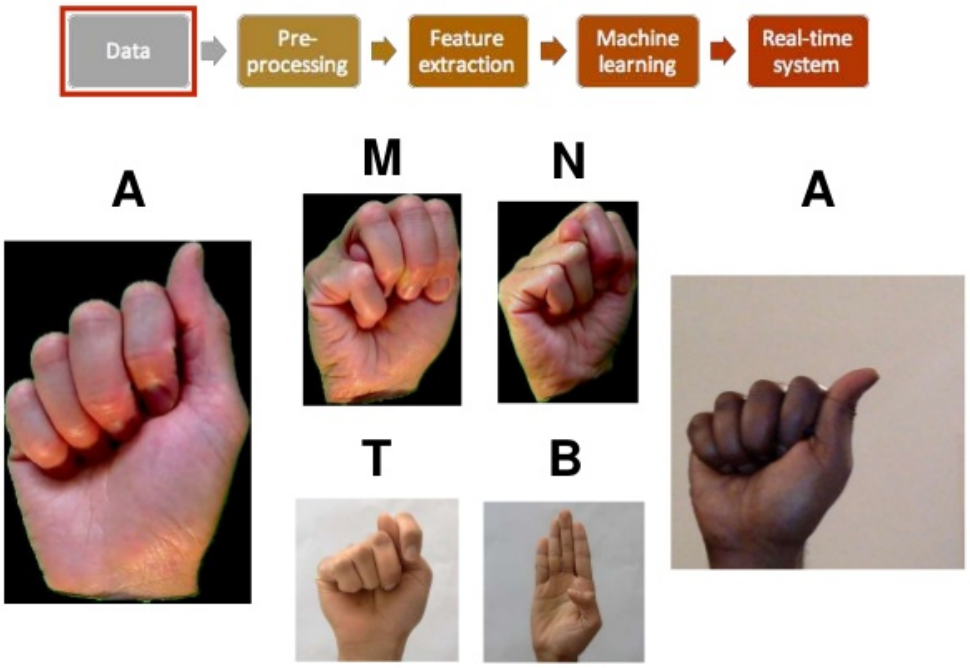

FIGURE 3.4: Deep Learning

3.4 Convolutional Neural Network

A convolutional neural network is nothing more than a DeepLearning algorithm capable of assigning biases and weights to different objects in an image and on the basis of that same image it can differentiate one image from another. It consists of processing different layers of image classification and it is designed with means to represent the functioning of neurons in the human brain[4].

Even if the image with minimum pixel is taken into account, it still needs 4x4 matrix and is required to look at the same image in different channels of color formats like RGB, Greyscale, HSV, etc., so it's very difficult to process thousands of For example, photos with high pixel rates. Pixels. Here comes the need for a convolutional neural network that winds each image into its basic, reduced form of the matrix that can be distinguished at the same time. This increases accuracy and speed and also reduces classifier model processing. The foundation layer is also supported by the assembly layer to reduce the need to manipulate the classifier model. It also changes the matrix based on the dominant traits. MAX Pool and AVGPooling[6].

The Inception-v3 convolutional neural network was implemented during the design of this system. Inception v3 is a 48-layer deep neural network. Inception Network is better than most convolutional neural networks because it simply does not always dig deeper into layers like other convolutional neural networks, instead it believes in working more broadly on the same layer before going deeper into the next layer. This is why bottlenecks are used during model training. A bottleneck in the neural network is just more likely to have fewer neurons than the layers above or below. The TensorFlow bottleneck is the last step of the preprocessing phase that begins before the actual training of the dataset begins.

3.5 Computer Vision

Computer vision is an interdisciplinary scientific field that deals with the question of how computers can gain a comprehensive understanding from digital images or videos. From a technical point of view, it seeks to understand and automate tasks that the human visual system can perform. The tasks of computer vision include methods of capturing, processing, analyzing and understanding digital images, as well as extracting high-dimensional data from the real world to generate numerical or symbolic information, z Understanding in this context means the conversion of visual images (the input of the retina) into Descriptions of the world that make sense for thought processes and can evoke appropriate action. This understanding of the image can be viewed as the disentanglement of symbolic information from image data using models constructed with the help of geometry, physics, statistics and learning theory.

The scientific discipline of computer vision deals with the theory behind artificial systems from which information is extracted and images. The image data can take many forms, e.g. B. video sequences, views from several cameras, multi-dimensional data from a 3D scanner or a medical scanning device. The technological discipline of computer vision seeks to apply its theories and models to the construction of computer vision systems.

3.5.1 OpenCV

OpenCV is an open source library for Computer Vision. Now since all training and sorting is ready to run when an eye to eye is needed, the system designed to capture real-time images of hand gestures that can then be submitted for sorting and identification. OpenCV adds intelligence to Deep Learning models for visualization image processing. Here the images are considered in 2 channels as: RGB Channel and Grayscale Channel, so once the image is captured by OpenCV, it is first converted to a gray channel and then undergoes morphological processing as shown. in [9]. OpenCV makes use of the Numpy library for the numerical calculation of images in the form of a pixel matrix.

A blue box of a particular size was designed with the help of OpenCV so as to take into account the hand gestures present inside this blue box. It then converts the image

over different channels, then converts the image to a reconstructed matrix shape so that the classifier model can compare it with previously learned labeled images. It will then predict a gesture suggestion based on the generated score. As OpenCV converts the hand gesture in real time, it will continually suggest predictions due to the slightest movement of the hand gesture in real time. The confirmed prediction with the highest score will enter the sequence until a COMAND CALL is executed. Then the whole sequence will go into the next step of the designed interface, that is, it will be converted to audio format which then will wake up the virtual voice assistant and become the input request.

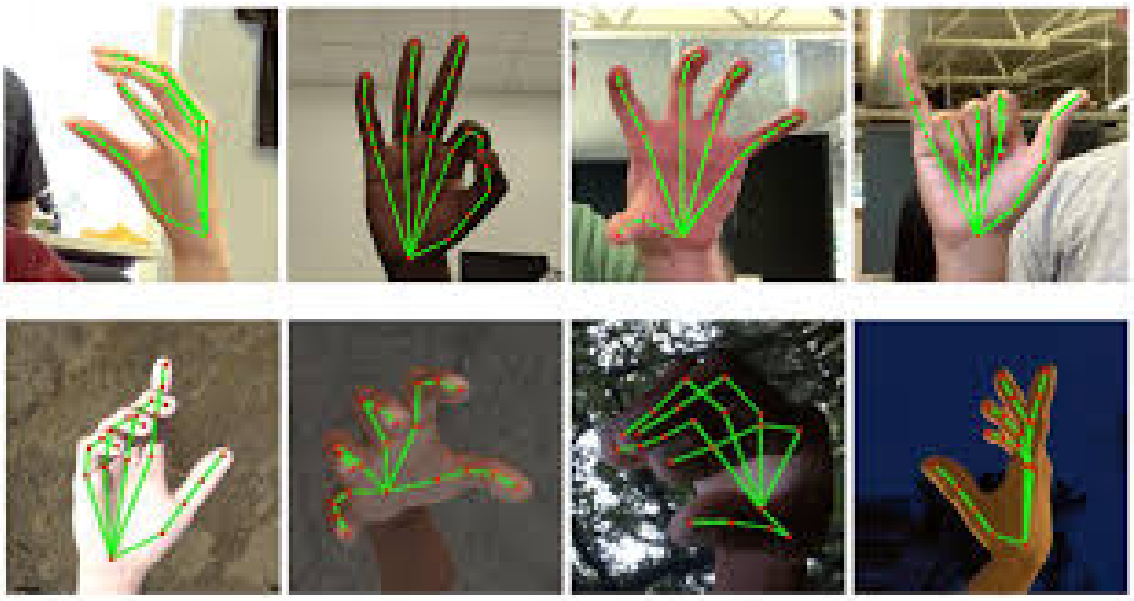

FIGURE 3.5: OpenCV

3.6 Python Text and Speech APIs

Several APIs are available for converting text to speech in Python. One such API is the Google Text-to-Speech API commonly known as the gTTS API. gTTS is a very easy to use tool that converts the entered text to audio which can be saved as an mp3 file. The gTTS API supports multiple languages including English, Hindi, Tamil, French, German and many more. The speech can be delivered in any of the two available audio speeds, fast or slow. However, starting from the last update, the voice of the generated audio cannot be changed.

3.6.1 Text-to-Speech

The Python text-to-speech library that used is very simple and easy to use. It makes use of modules like pyttsx3 and engine.io which let us change different properties like rate and intervals of text to speech conversion and outflow.pyttsx is a cross-platform text to speech library which is platform independent. The major advantage of using this library for text-to-speech conversion is that it works offline. However, pyttsx supports only Python 2.x. Hence, we will see pyttsx3 which is modified to work on both Python 2.x and Python 3.x with the same code.

Text-to-speech (TTS) technology reads digital text out loud. You can take words on computers, smartphones, tablets and convert them to audio. Additionally, all kinds of text files can be read aloud, including Word, page documents, and online web pages that can be read aloud. TTS can help children who have trouble reading. There are many tools and applications available to convert text to speech. Python comes with many useful and easily accessible libraries, and we will see how we can offer text-to-speech with Python in this article.

Different API's are available in Python in order to convert text to speech. One of Such API's is the Google Text to Speech commonly known as the gTTS API. It is very easy to use the library which converts the text entered, into an audio file which can be saved as a mp3 file. It supports several languages and the speech can be delivered in any one of the two available audio speeds, fast or slow.

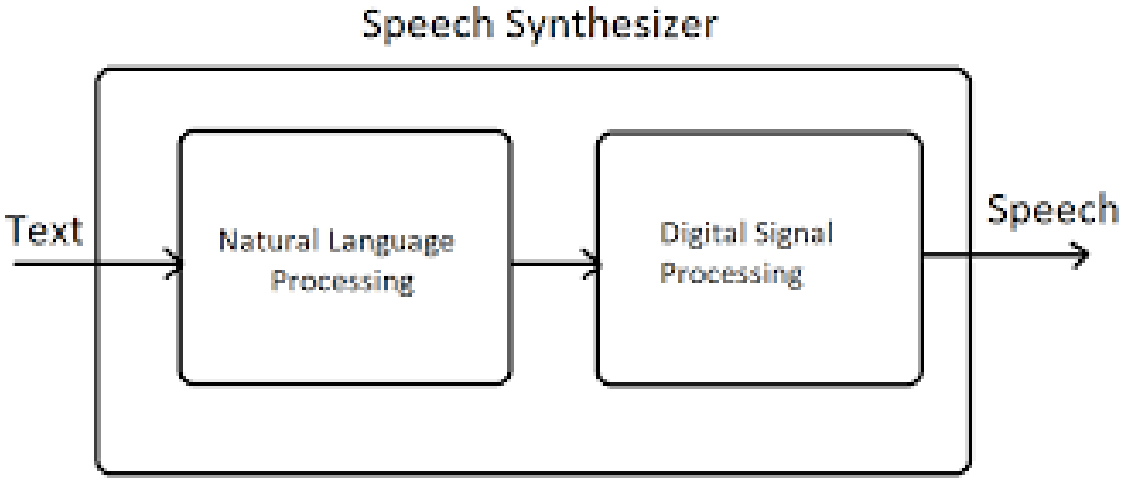

FIGURE 3.6: Text to speech

3.6.2 Speech-to-Text

The Python speech-to-text library by which practicing makes use of speech recognition module. It let us adjust the ambient noise and also helps in recording the audio in form of mp4 files.The first component of speech recognition is of course speech. Speech must be converted from physical sound to an electrical signal with a microphone and then converted to digital data with an analog-to-digital converter. Once digitized, multiple models can be used to convert the audio to text. Most modern speech recognition systems are based on a so-called Hidden Markov Model (HMM). This approach is based on the assumption that a speech signal, when viewed on a sufficiently short timescale (e.g. ten milliseconds), can reasonably be approximated as a stationary process, i.e. a process in which the statistical properties change over time Do not change over time In a typical HMM, the speech signal is divided into 10 millisecond fragments.

The power spectrum of each fragment, which is essentially a representation of the power of the signal as a function of frequency, is mapped onto a real number vector known as cepstral coefficients. The dimension of this vector is usually small - sometimes as low as 10, although more accurate systems can be 32 or more. The final output of the HMM is a consequence of these vectors. To decode speech into text, groups of vectors are matched to one or more phonemes - a basic unit of speech. This calculation requires training because the sound of a phoneme varies from speaker to speaker and even varies from one utterance to another by the same speaker. A special algorithm is then used

to determine the most likely word (or words) that will produce the given sequence of phonemes.

FIGURE 3.7: Speech to Text

CHAPTER 4

RESULT

A short demonstration of this project is given below with help of images. Here, it performed a Hand Gesture by representing a term "Weather" and once CALL COMMAND is received, a query asking weather will be sent to Virtual Voice Assistant and the real-time output will be converted into text and displayed on output frame.

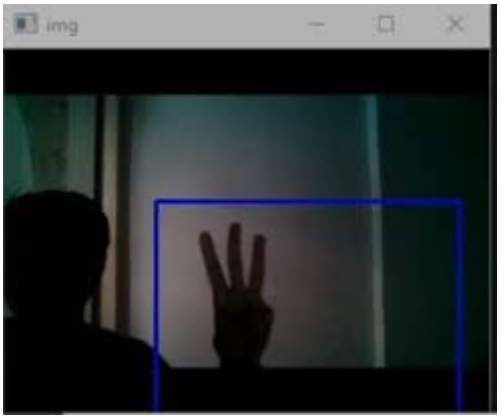

FIGURE 4.1: Hand Gesture for term Weather

Fig. 4.1, shows how the designed system will capture realtime hand gestures, only the ones inside blue box, using Web Camera. In the above figure person is performing a hand gesture representing word 'Weather'.

FIGURE 4.2: Sequence frame

Fig. 4.2, Represents a window frame which will be displaying system's suggestions for captured hand gesture.

The system capturing a real-time hand gesture for CALL Command, so that the previously captured words in the sequence frame can now be converted into Audio format as an input for Assistant.

The output frame which displays the real-time response generated by Voice Assistant, but converted in text format and displayed on the screen.

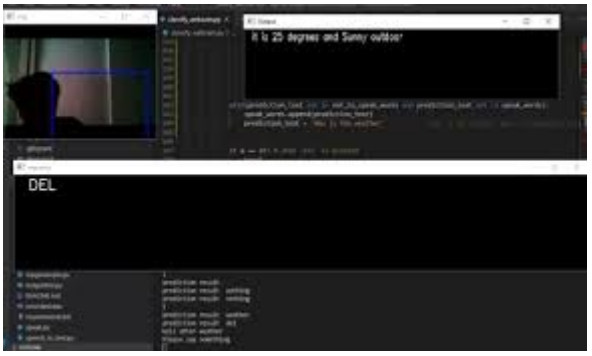

FIGURE 4.3: Interface

Fig. 4.3, is a view of real-time interface of the designed system with all the window frames tied up together, performing their expected roles. Now, again if no gesture is being performed the sequence window will print DEL, which is also another CALL Command.

FIGURE 4.4: Classifier's Accuracy

Fig. 4.4, shows us the score which indicates the accuracy by which the gesture is being classified by the model. It ranges between 0 to 1 as shown above.

CHAPTER 5

CONCLUSION

The designed system was successful in capturing HandGestures using the integrated web camera and processing them and converting them to text format and displaying them on the input frame, then converting them to audio format when receiving a CALLCOMMAND. The audio becomes a request for VirtualAssistant and again the audio output has been successfully converted to text format and displayed on the screen as shown above in the results. Most of the time, under the privileged conditions, the system was able to provide precise and best results. However, sometimes in poor lighting conditions and in the absence of a suitable background, the system has tried to produce correct and expected results.

During the development of this system, many difficulties are faced as in the latest updates these virtual assistants have stopped responding to digital voices of a certain frequency and bandwidth, in order not to reactivate these devices in case of commercial advertising (as in the case of ALEXA) so you need to select the digital voice, which is used in our Python speech synthesis library to avoid this problem.

Also as the designed system is trained completely using the chosen Dataset, thus selection of Dataset should be done on basis of the Standard Sign Language used by the deaf-mutes in targeted region as these languages tend to change locally and globally.

Basically the system can be considered as a boon to people with hearing disabilities or speaking disabilities or both at the same time. These system would not only bring technology into their Personal lives but also give rise of opportunities in their professional life.

CHAPTER 6

FUTURE SCOPE

The system currently being designed operates entirely on the underlying data set used to train the system, which limits its use to specific groups of people who are communicating with the similar sign language. However, it has been found that there are different forms of sign languages around the world. Therefore, the dataset used must be changed according to the standard sign languages used in that nation or region. Since the graphical user interface of the current system is simple but not visually comforting, it aims to design and create a better visual interface that makes it even more vibrant and eye-catching, making it easier and more interesting for our target audience to use.

The existing virtual voice assistants are basically a form of smart speakers that are voice automated. Our designed user interface is an add-on to make it accessible to the deaf and mute using a web camera and laptop. If this system gets a good response from the users, it can even reach a greater scope by being able to fully integrate our system with these Virtual Assistant speakers in a better form, the speakers themselves being made up of a web camera that acts as an eye and respond to HandGestures when combined with processing skills.

The input goes in at one end, and then it flows through this system of multiple operations and comes out the other end as output.This is why it is called TensorFlow because the tensor goes in it flows through a list of operations, and then it comes out the other side.

CHAPTER 7

REFERENCES

[1] Yusnita, L., Rosalina, R., Roestam, R. and Wahyu, R., 2017. Implementation of Real-Time Static Hand Gesture Recognition Using Artificial Neural Network. CommIT (Communication and Information Technology) Journal, 11(2), p.85.

[2] Rathi, P., Kuwar Gupta, R., Agarwal, S. and Shukla, A., 2020. Sign Language Recognition Using ResNet50 Deep Neural Network Architecture. SSRN Electronic Journal.

[3] V. Adithya, P. R. Vinod and U. Gopalakrishnan, "Artificial neural network based method for Indian sign language recognition," 2013 IEEE Conference on Information Communication Technologies, Thuckalay, Tamil Nadu, India, 2013, pp. 1080-1085.

[4] Guru99.com. 2020. Tensorflow Image Classification: CNN(Convolutional Neural Network). [online] Available at: ¡https://www.guru99.com/convnet-tensorflow-imageclassification.html¿

[5] Guo, T., Dong, J., Li, H. and Gao, Y., 2017. Simple Convolutional Neural Network on Image Classification. IEEE 2nd International Conference on Big Data Analytics, pp.1-2.

[6] Medium. 2020. A Comprehensive Guide To Convolutional Neural Networks—The ELI5 Way. [online] Available at: ¡https://towardsdatascience.com/a-comprehensive-guide-toconvolutional-neural-networks-the-eli5-way-3bd2b1164a53¿.

[7] Medium. 2020. Deep Learning With Tensorflow: Part 1 — Theory And Setup. [online] Available at: ¡https://towardsdatascience.com/deep-learning-with-tensorflow-part1-b19ce7803428¿.

[8] Issac, R. and Narayanan, A., 2018. Virtual Personal Assistant. Journal of Network Communications and Emerging Technologies (JNCET), Volume 8(Issue 10, October (2018).

[9] Lai, H. and Lai, H., 2014. Real-Time Dynamic Hand Gesture Recognition. International Symposium on Computer, Consumer and Control, pp.658-661.

[10] Pankajakshan, P. and Thilagavathi B, 2015. Sign language recognition system. 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).

[11] K. A. Bhaskaran, A. G. Nair, K. D. Ram, K. Ananthanarayanan and H. R. Nandi Vardhan, "Smart gloves for hand gesture recognition: Sign language to speech conversion system," 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA), Kollam, 2016, pp. 1-6, doi: 10.1109/RAHA.2016.7931887.

[12] Ertham, F. and Aydin, G., 2017. Data Classification with Deep Learning using Tensorflow. IEEE 2nd International Conference on Computer Science and Engineering, pp.757-7.