LANGUAGE PRONUNCIATION FOR BETTER UNDERSTANDING

Jayashubha J [1]
Asst.Professor
Department of Computer Science
KSSEM
Bangalore, India
jayashubha@kssem.edu.in

Karthik Gowda D [2]
Student
Department of Computer Science
KSSEM
Bangalore, India
gowdakarthik206@gmail.com

Ram Reddy [3]
Student
Department of Computer Science
KSSEM
Bangalore, India
Ramramareddy1209@gmail.com

Shankar K S [4]
Student
Department of Computer Science
KSSEM
Bangalore, India
shankarks311@gmail.com

Hamsaveena S [5]
Student
Department of Computer Science
KSSEM
Bangalore, India
hamsaveena2309@gmail.com

***Abstract* - The aim of our project to automate the interface to overcome from the language barrier among countries and also states within the country, the above mentioned application will perform the various features in the interface. It includes 4 modules voice recognition, translation and speech synthesis and image translation and gives audio of the translated language. Also the application accepts text boards as a picture and converts it into the language needed. Interface is able to recognize the text present in the image which stored in system or captured using camera and translate the text into the language needed and display the translation result back on to the screen of system.**

Keywords—Speech Recognition, OCR technology, Language Translator, image extraction, text to speech

I. INTRODUCTION

Machine replication of human functions like reading is an ancient dream. However, over the last five decades, machine reading has grown from a dream to reality. Visually impaired people report numerous difficulties with accessing printed text using existing technology, including problems with alignment, focus, accuracy, mobility and efficiency. We present a smart device that assists the visually impaired and traveler's which effectively and efficiently reads paper-printed text. The proposed project uses the methodology of a camera based assistive device that can be used by people to read Text document. The framework is on implementing image capturing technique in an embedded system based on Raspberry Pi board. The design is motivated by preliminary studies with visually impaired people, and it is small-scale and mobile, which enables a more manageable operation with little setup. In this project we have proposed a text read out system for the traveler's and visually challenged. The proposed fully integrated system has a camera as an input device to feed the printed text document for digitization. Speech is probably the most efficient medium for communication between humans.

1.1 problem statement

Use of mobile devices has increased a lot. Many Text to speech , multilingual translation, text extraction interfaces are developed for mobile users. But similar type of interface is not developed for desktop users. We are creating an interface that consist all of the above interfaces in one single interface.

1.2 Objectives

Our main aim is to combine all different tasks such as text translation, text synthesis and text extraction from image all embedded in one so that we get a user friendly interface.

1.3. *Scope*

We develop this interface for desktop application. Here we are integrating the text to speech, image extraction and language translator in one system so user doesn't have to download for the different application.

II. RELATED WORK

Asha G. Hagargundet et al carried out a work and they concluded that the basic framework is an embedded system that captures an image, extracts only the region of interest (i.e. region of the image that contains text) and converts that text to speech. It is implemented using a Raspberry Pi and a Raspberry Pi camera. The captured image undergoes a series of image pre-processing steps to locate only that part of the image that contains the text and removes the background. Two tools are used convert the new image (which contains only the text) to speech. They are OCR (Optical Character Recognition) software and TTS (Text-to-Speech) engines. The audio output is heard through the raspberry pi's audio jack using speakers or earphones.

Aaron James S et al carried out a work and they concluded that Optical character recognition (OCR) is the identification of printed characters using photoelectric devices and computer software. It coverts images of typed, handwritten or printed text into machine encoded text from scanned document or from subtitle text superimposed on an image. In this research these images are converted into audio output. OCR is used in machine process such as cognitive computing, machine translation, text to speech, key data and text mining. It is mainly used in the field of research in Character recognition, Artificial intelligence and computer vision.

III. SNAPSHOTS

Fig. 1. Text to audio converter

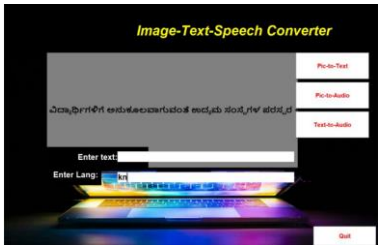

Fig. 1. Image converted to text

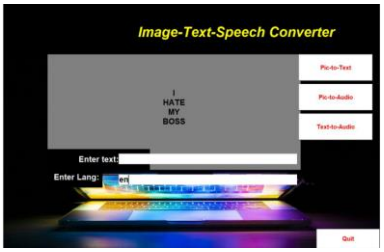

Fig. 1. Pic to Text converter

IV. INCORPORATED PACKAGES

Speech adaptation

Customize speech recognition to transcribe domain-specific terms and rare words by providing hints and boost your transcription accuracy of specific words or phrases. Automatically convert spoken numbers into addresses, years, currencies, and more using classes.

Domain-specific models

Choose from a selection of trained models for voice control and phone call and video transcription optimized for domain-specific quality requirements. For example, our enhanced phone call model is tuned for audio originated from telephony, such as phone calls recorded at an 8khz sampling rate.

Easily compare quality

Experiment on your speech audio with our easy-to-use user interface. Try different configurations to optimize quality and accuracy.

Speech-to-Text On-Prem

Have full control over your infrastructure and protected speech data while leveraging Google's speech recognition technology on-premises, right in your own private data centers.

V. THE PROPOSED METHOD

The aim of the proposed system is to develop a system that has capability to perform Translation, Converting text to speech, Text Extraction. The system proposed here will be developed for a small domain of words.

A. Image to Text

- Insert an image in Jpeg format.
- Extract the text in the image and give an output in text format.

Fig. 1 Image to text conversion

B. Speech to Text

- System listens to the users speech or audio(mp3) and converts it into text format.

Fig. 2 Speech to text conversion

C. Text to Speech

- It convert the text format into the speech format.

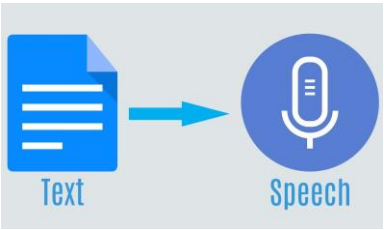

Fig. 3 Text to speech conversion

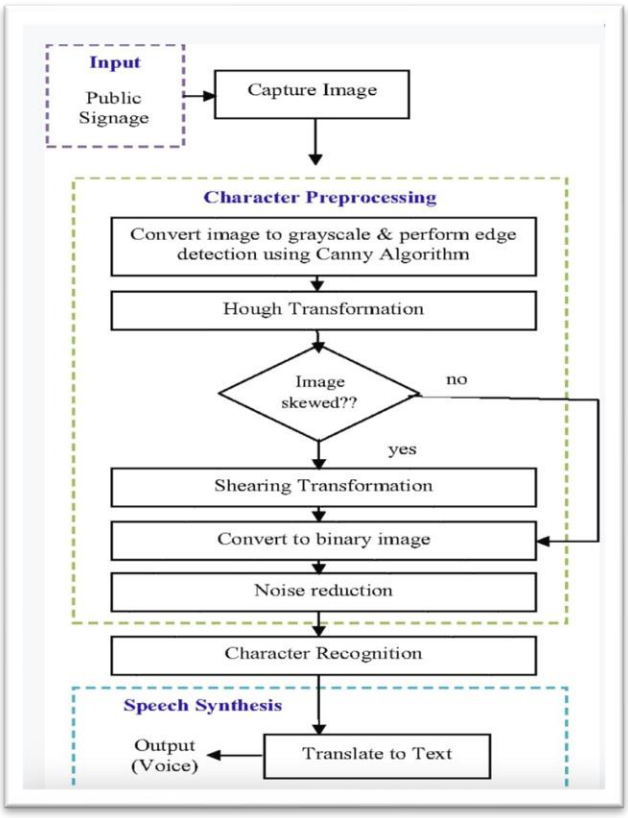

Fig. 4 Flow chart of the interface

D.Image Reshaping

The input during relegation of an image is a three-dimensional tensor, where each channel has aprominent unique pixel. All the images must have identically tantamount size corresponding to 3D feature tensor. How- ever, neither images are customarily coextensive nor their corresponding feature tensors. Most CNNs can only accept fine-tuned images. This engenders several problems throughout data collection and implementation of model. However, reconfiguring the input images before augmenting them into the network can help to surmount this constraint.

The images are normalized to converge the pixel range between 0 and 1. Then they are converted to 4 dimensional arrays using *data=np. reshape (data, (data. shape [0],img size, img size,1)*) where 1 indicates the Grayscale image. As, the final layer of the neural network has 2 outputs – with mask and without mask i.e., it has categorical representation, the data is converted to categorical labels.

become ascendant in miscellaneous computer vision tasks. The current method makes use of Sequential CNN.

The First Convolution layer is followed by Rectified LinearUnit (ReLU) and Max Pooling layers. The Convolution layer learns from 200 filters. Kernel size is set to 3 x 3 which specifies the height and width of the 2D convolution window.As the model should be aware of the shape of the inputexpected, the first layer in the model needs to be provided with information about input shape. Following layers can perform instinctive shape reckoning. In this case, *in-put shape* is specified as *data. shape [1:]* which returns the dimensions of the data array from index 1. Default padding is "valid" where the spatial dimensions are sanctioned totruncate and the input volume is non-zero padded. The activation parameter to the Conv2D class is set as "relu".It represents an approximately linear function that possesses all the have default value (1,1).

VI. Result And Analysis

People with disabilities like physical impairments often find it difficult in sending a text message or typing a statement in general. The system supports them by allowing them to speak their text through microphone and performing the conversion from audio to text.This system also helps people connect with any languages especially Indian Language from their own language.The system basically uses two techniques Speech Recognition and Translation.

VII. Conclusions

In this proposed system, we implemented the system for user who phasing problems of language barrier and also it user interface is also user friendly so that user can easily interact with this interfaceso because of this system don't have to use dictionary for understanding the meaning of word, so it automatically reduce the user task for understanding the languages for communication.

References

[1] Keerthana S, Meghana H, Priyanka K, Sahana V. Rao, Ashwini B "Smart Home Using Internet of Things", proceedings of Perspectives in Communication, Embedded -systems and signal processing, 2017

[2] Sutar Shekhar, P. Sameer, Kamad Neha, Prof. Devkate Laxman, "An Intelligent Voice Assistant Using Android Platform", IJARCSMS, ISSN: 232-7782, 2017

[3] Rishabh Shah, Siddhant Lahoti, Prof. Lavanya. K, "An Intelligent Chatbot using Natural Language Processing", International Journal of Engineering Research ,Vol.6 , pp.281-286, 2017

[4] Ravivanshikumar, Sangpal, Tanvee, Gawand, Sahil Vaykar, "JARVIS: An interpretation of AIML with integration of gTTS and Python", proceedings of the 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kanpur, 2019.

[5] Dhiraj Pratap Singh, Deepika Sherawat, Sonia, "Voice activated desktop assistant using Python", proceedings of High Technology Letters, ISSN: 1006-6748, 2020.